Psychological Review
1988, Vol. 95, No. 3, 385–408

0033-295X/88/$00.75

A Retrieval Theory of Priming in Memory

Roger Ratcliff and Gail McKoon
Northwestern University

We present a theory of priming that is designed to account for phenomena usually attributed to the action of a spreading activation process. The theory assumes that a prime and target are combined at retrieval into a compound cue that is used to access memory. If the representations of the prime and target are associated in memory, the match is greater than if they are not associated, and this greater match facilitates the response to the target. The compound cue mechanism can be implemented within the framework of several memory models; descriptions of these implementations are presented. We summarize empirical results that have been taken as evidence for a spreading activation process and show that the retrieval theory can also account for these phenomena and that, in some cases, the retrieval theory provides predictions that are more constrained than those provided by spreading activation theories. Also, two experiments are reported that address predictions about the range of priming (in terms of number of connected concepts) and the decay rate of priming (in terms of intervening items). In both cases, the retrieval theory provides a better account of the data than spreading activation. Finally, contrasts between the compound cue theory and long-term priming phenomena are presented.

Because the amount of information stored in human memory is so large, a successful model of memory requires mechanisms that allow fast and efficient search and access. When several cues for retrieval are presented to the processing system, the cues must be used to focus access onto some subset of the information in memory. The information in memory is often assumed to be a semantic network of concepts with each concept directly connected to related concepts. One mechanism that has been proposed to accomplish focusing within a semantic network model is spreading activation. It is assumed that processing a concept temporarily activates that concept and closely related concepts as activation spreads from link to link through the network. Activation will maximally activate the closest set of concepts, and this set is then available to selection or decision processes. A primary source of evidence for this activation process is the priming phenomenon, in which presentation of one item will speed responses to a related item.

In this article, we propose an alternative theory that can be described generally as a retrieval/decision theory or, more specifically, a compound cue theory. The mechanism by which search is focused according to this theory involves no temporary activation of the long-term memory system. Instead, items presented to the system join together to form a compound cue, and the familiarity of this compound cue is determined by the strengths of connections between the cue and items in memory. The familiarity of the compound cue is assessed by direct access or by parallel comparisons to all items in memory (depending on the way the theory is implemented), and it is assumed that the greater the familiarity, the faster the response time (specific models for latency assumptions are described below). In this article, the retrieval theory is applied to priming phenomena and is shown to be capable of explaining the same empirical findings as spreading activation theories. Two new experiments are also presented in which data are successfully predicted by the retrieval model but that require modification of current models of spreading activation. In the latter part of this article, implementation of the compound cue mechanism for priming within several different models is evaluated.

In semantic network models, memory is assumed to consist of a set of interconnected nodes, with each node representing a concept. Nodes are connected if they are related (dog–cat) or if they have been studied together ("baby" and "concrete" would be connected if the sentence "the baby dropped to the concrete" was studied). When an item is presented to the system, the activation of the concept representing the item is increased, and activation spreads through the network, increasing the activation level of other nearby concepts. The amount of activation given to connected concepts is assumed to be a function of distance; the closer some concept is in memory to the input concept, the more it will be activated. If a retrieval query is presented to the system, then activation will spread from the concepts in the query to connected concepts. For example, in the sentence above, if "baby" is presented as a cue for recall, activation will spread from "baby" to "dropped" and "concrete" and to the node representing the whole sentence. For a recognition test of a single word, the amount of activation in the node representing the word determines whether the decision will be positive or negative. If a whole sentence, previously studied, is pre-

This research was supported by National Science Foundation Grant BNS 8510361 to Roger Ratcliff, National Institutes of Health Grant HD18812 to Gail McKoon, and National Science Foundation Grant BNS 8516350 to Gail McKoon.

We would like to thank John Anderson, Steve Grossberg, Doug Hintzman, George Mandler, Mike Masson, Mary Jo Nissen, Greg Stone, two anonymous reviewers, and especially Gordon Logan for their comments on the article.

Correspondence concerning this article should be addressed to Roger Ratcliff, Psychology Department, Northwestern University, Evanston, Illinois 60208.

sented for recognition, then either the activation at the node representing the sentence is used as the basis for the recognition decision, or the connections representing the sentence become available for evaluation.

The process of spreading activation has been proposed as a general retrieval mechanism for semantic network models, but by and large there has been little direct evidence for the activation process. The most direct evidence has been considered to be priming phenomena, and so the theory developed here and the contrast with spreading activation theory not only concern the domain of priming phenomena, but have important implications for retrieval theories in general. In this context, priming is usually defined as the facilitation given by the presentation of one item (the prime) to a response to an immediately following test item. Spreading activation theories account for such facilitation by assuming that activation spreads from the prime to the target, so that when the target is presented, its activation level has already been raised. Thus, a faster and/or more accurate response can be made to the target, because less additional activation is needed to reach a response criterion. For example, if "baby" is presented as a prime, the concept "concrete" in memory will be activated. Then when "concrete" is presented as a target, the response will be facilitated.

The retrieval, or compound cue, theory presented in this article contrasts with spreading activation theories in its assumptions about processing. In the retrieval theory, the prime does not facilitate a response to the target by affecting the target in long-term memory. Instead, the prime and target together are matched against long-term memory. Response to the target will be facilitated to the extent that the prime and target are associated in memory. By providing a competitor to spreading activation, we remove priming phenomena as unequivocal support for spreading activation theories and so undermine a major line of empirical justification for the theories.

The assumption that the prime and target join together in a compound cue can be implemented within several theoretical frameworks. The key to the implementation is a nonlinear component of goodness-of-match between test items and items in memory (see Grossberg, 1980). The nonlinear component is required in order to provide a boost to goodness-of-match when the two test items in the compound are associated in memory (for priming) and more generally in some cases so the model can learn (e.g., Murdock, 1986). In the case of priming, if A–B and C–D were studied and encoded in memory, then in a linear system with no special associative component, the match with memory (i.e., A–B plus C–D) for A–B as a test item would be just as great as the match for A–D. Nonlinearity is introduced into the models either with a nonlinear transformation of the "raw" goodness-of-match between the compound cue and memory (e.g., product of strengths in Gillund & Shiffrin, 1984, and cubing familiarity in Hintzman, 1986a) or with an additional associative component (see discussion of Murdock, 1982, below). The implications of these different implementations vary somewhat, but the main effect is the same across the models.

Initially, we present the compound cue model for priming within the framework of the Gillund and Shiffrin (1984) model because the scheme for priming is already implicit within the framework of that model. In the Gillund and Shiffrin (1984) model, memory is assumed to be a matrix of the strengths between possible cues to the system and concepts (or "images") stored in the system. When an item (cue) is presented, its total familiarity is assessed by summing the strengths between the cue and all images in memory. Recognition decisions are based on this calculated familiarity. Recall involves a searchlike process (Raaijmakers & Shiffrin, 1981). On each recall cycle, the relative strength from a cue to an image (compared with other strengths from that cue to other images) determines which image should be sampled. Once an image has been sampled, the probability that the image will be retrieved is determined from its strength. Thus, both recall and recognition can be modeled using different retrieval processes operating on one memory structure.

For priming, our retrieval model asserts that the prime and target form a compound cue, implemented as in Gillund and Shiffrin (1984). The familiarity of this cue is calculated by summing over all images in memory the strength between the prime and an image *multiplied* by the strength between the target and that image. To the extent that the prime and target are directly connected to each other in memory or are directly connected to one or more common images, then the overall familiarity of the compound will be higher than would be the familiarity of the target alone. Reaction time and accuracy are assumed to be related to familiarity, so that higher familiarity leads to facilitation of responses: Thus, according to the retrieval model, presentation of the prime does not affect cue-to-image strengths in long-term memory; specifically, the prime does not affect strength from the prime cue to a related target's image, before the target is processed, the way it would in a spreading activation account (unless prime presentation time is long, and learning or strategic retrieval takes place). Instead, the prime is combined with the target to form a compound cue and so the prime simply helps determine the familiarity value that determines response time and accuracy at the time the target is presented. Thus, in contrast to the active process of spreading activation, the retrieval theory assumes that the processes producing priming are passive, a result of combining the prime and target for retrieval.

Spreading Activation

The concept of activation has become very popular in cognitive psychology. There have been a number of theories that use activation as a basic process, and now the term "activation" is becoming part of the language, so that it is not uncommon to see the terms "priming" and "activation" used interchangeably. The acceptance of the activation process has occurred partly because activation serves as a theoretical expression of the empirical phenomenon priming, and partly because there are no successful competing theoretical explanations of priming. The aims of the theory presented in this article are to provide a framework within which to explain priming phenomena and a competitor to spreading activation that will produce serious evaluation of that popular yet relatively untested process.

Spreading activation appears as a component of a wide variety of theories for a number of different empirical tasks and results. In this article, discussion centers on priming; first on priming in item recognition between single words or sentences, and second on priming in lexical decision, naming latency, and word identification. In each of these areas, spreading activation is a major component of theoretical interpretations of data.

Probably the earliest spreading activation model to influence recent cognitive psychology was the semantic network model that Quillian (1966; Collins & Loftus, 1975; Collins & Quillian, 1969) applied to semantic verification in order to account for retrieval of semantic facts. The data the model explained were obtained from a semantic verification task in which subjects were required to respond true or false to questions such as "Is a robin a bird?" The results of interest were increases in reaction time as a function of increasing distance between the concepts in the proposed network representation of the concepts. The increases in reaction time were modeled by assuming that activation took time to spread from node to node through the network. In the area of word recognition, Meyer and Schvaneveldt (1971) were among the first to consider spreading activation as the mechanism responsible for priming in lexical decision. Later, Anderson (1976, 1983) incorporated spreading activation into his cognitive theory to provide the mechanism for activating information (loosely equivalent to short-term memory) for processing. Other models closely related to those of Anderson (1983) and Collins and Loftus (1975) use spreading activation as the major process for accessing memory. Among these are Dell's (1986) model for word production, Dosher's (1982) model for sentence matching, Levin's (1976) model for problem solving, and McClelland and Rumelhart's (1981) model for word perception. There are also many models that include spreading activation as a more secondary component (e.g., Kieras, 1981; Miller, 1981).

Spreading activation is also a major processing component of "connectionist" models. Connectionist models have evolved in psychology, theoretical neuroscience, and artificial intelligence, and include some of the models already noted (Dell, 1986; McClelland & Rumelhart, 1981). These theories use spreading activation as the process that translates input to output via transformations built into the connectionist network. Input adds activation to the network, and activation of the output nodes is determined by transmission of activation through the network. Although these connectionist models require spreading activation as a basic mechanism, they typically do not require that priming between concepts be explained as activation spreading from prime to target in the manner of the Collins and Quillian (1969) and Anderson (1976, 1983) models. In fact, these models have not been specified in sufficient detail to provide any kind of comprehensive account of priming; this point is elaborated later in this article.

This brief review illustrates the profound importance that spreading activation has had in psychological theory and the monopoly that spreading activation has in theoretical accounts of priming. Although there are some notable exceptions (Morton, 1969; Smith, Shoben, & Rips, 1974), spreading activation has been the dominant view of access to and retrieval from organized memory. The retrieval theory proposed in this article challenges spreading activation in that it can replace spreading activation as an explanation of priming in such areas as item recognition and lexical decision. In replacing spreading activation as an explanation of priming, the retrieval theory also calls into question spreading activation accounts of other phenomena. The specific models that can implement the retrieval theory must then be called on to provide alternative mechanisms for retrieval and search processes. For example, in the Gillund and Shiffrin (1984) model, the search mechanism developed for recall serves to access and make available information from memory for evaluation.

In the section that follows, the retrieval model is described in detail. Then the claim that it is a competitor for spreading activation is validated in three ways: (a) by showing that the retrieval theory can account for major aspects of the data obtained in priming experiments, (b) by showing in two new experiments that the retrieval model can predict data that current spreading activation models can account for only with major alterations, and (c) by showing that the retrieval model gives a better overall picture of priming phenomena than do other current models that use spreading activation.

The Compound Cue Retrieval/Decision Theory

The compound cue model is introduced within the framework of the recognition model proposed by Gillund and Shiffrin (1984). We use the Gillund and Shiffrin (1984) model because it has a cue combination mechanism already in place. However, it should be stressed that the compound cue mechanism is not tightly tied to the success of the Gillund and Shiffrin model. We will show how the compound cue mechanism can be implemented in other models that introduce the necessary nonlinearity via the same multiplicative rule as the Gillund and Shiffrin model (Grossberg & Stone, 1986; Medin & Schaffer, 1978), by raising the goodness-of-match value to a power (Hintzman, 1986a), or by an additional associative component of goodness-of-match (Murdock, 1982). Implementation of the compound cue view in these other models is presented later in this article.

A second reason we use the Gillund and Shiffrin model is that text representation can be integrated into this framework whereas other (vector) models have no natural way of representing text. Below, we qualitatively extend the Gillund and Shiffrin model beyond memory for single words to encompass processes for encoding, storing, and retrieving textual information. These processes are based on proposals by Kintsch and Vipond (1979) and Anderson (1983).

To provide a complete quantitative explanation of priming, we also need to extend the Gillund and Shiffrin model to map the familiarity values given by the model into reaction time and accuracy. This is done by assuming that familiarity drives a diffusion process that maps familiarity to reaction time and accuracy. These aspects of the retrieval theory are specified in the following sections.

Quantitative Formulation in the Gillund and Shiffrin (1984) Framework

Responses to items presented for recognition are based on the assessed familiarity of those items, as in the Gillund and Shiffrin (1984) model. According to this model, long-term memory is composed of images. At encoding, items enter a short-term buffer, and cue to target strengths between each item in the buffer and each other item in the buffer (including self-strength for each item and context to item strength for each item) are increased as a function of time. As a result of encoding, each cue item is related to each image in memory by some strength. This strength has some residual value if the cue and image are not directly connected to each other (i.e., were not rehearsed together) and some larger value if they are connected. The model also uses contextual cues and images to represent strength between items and contextual information, where context includes the experimental situation, the particular list of materials just studied, and so on. For a single cue, its familiarity is calculated as the strength between the cue and an image in memory multiplied by the strength between the context and that image in memory, and then summing this product over all images in memory. When more than one cue is presented, the cues are assumed to be assembled in short-term memory (the same buffer as used for encoding) and the familiarity of this compound cue is assessed. For priming, when a prime is presented preceding a target item, then the two items are assumed to form the compound cue (along with context). The familiarity of this compound is the sum over all images in memory of the strength of the prime to an image multiplied by the strength of the target to that same image multiplied by the strength of the context to that same image, where the strength of the prime is given less weight than the strength of the target. The familiarity of a compound cue is given by:

$$F(\mathrm{i}, \mathrm{j}) = \sum_{\mathrm{k}} S_{\mathrm{ck}} S_{\mathrm{ik}}{}^{W_{\mathrm{p}}} S_{\mathrm{jk}}{}^{(1 - W_{\mathrm{p}})}, \tag{1}$$

where i is the prime item, j is the target item, k are all the images in memory, c is the context cue, and W_p is the weighting on the prime (varying between 0 and 1).

The prime is given less weight in the calculation of familiarity because the response is made to the target, not the prime. For example, in recognition, if the prime and target were weighted equally, then a previously studied prime (for which the correct response would be "old") and a new target (for which the correct response would be "new") would have about chance accuracy. Weighting the prime less gives higher accuracy for the target.

To illustrate the model, a numerical example is given in Table 1. In the retrieval structure shown in Table 1, the self-strength of a cue to its own image in memory is set to 1.0. Also, the strength between a cue and an image to which it is directly connected is set to 1.0. Other cue-to-image strengths are set to be residual, 0.2. In the first calculations of familiarity of a compound cue, the prime and target are weighted equally with weights set to 1.0 to make the example simple. The calculations show that familiarity is greater for primes and targets that share associates than for primes and targets that do not share associates. In the second set of calculations, the prime and target are weighted differently; the resulting familiarity values differ in magnitude but show the same pattern.

The examples in Table 1 show important features of the retrieval model. These are mentioned here and are discussed in detail later in the article. First, the largest value of familiarity will be obtained when a prime and target are identical. (Calculating this value in Table 1, without weights, $F[2, 2] = 3.12$.) In practice, however, such repetition priming is difficult to compare with associative priming because of the extra effects of repeating identical encoding operations (cf., Ratcliff, Hockley, & McKoon, 1985). When the prime and target are not identical, then the greater the number of common associates they share (e.g., S_{23} in Table 1), the greater the value of familiarity. Thus, the more connected two concepts are in memory, the greater is the familiarity and so the faster the reaction time on the target. This calculation of familiarity introduces a nonlinear component because strengths are multiplied. To see this, consider the case in which items A and B share associates in memory. Then when A–B is tested, the sum of products involves two large numbers multiplied together. If the test were A–C where A and C were not associated, large strengths associated with cue A (e.g., A–K where A and K are associated) would be multiplied with small strengths associated with cue C (C–K where C and K are not associated) and vice versa, and this would lead to a smaller value of familiarity.

The second important feature of the model is that the range of priming is limited. Familiarity is increased over residual only when the prime and target are directly connected to each other or when they are each directly connected to the same intervening item or items. The more such directly connected intervening items, the greater the familiarity. Thus the Gillund and Shiffrin implementation of the compound cue theory of priming predicts both one- and two-step priming with two-step priming much smaller in magnitude than one-step priming but with a maximum of two-step priming. It should be noted, however, that not all implementations of the compound cue notion will predict two-step priming (e.g., Hintzman, 1986b, and Murdock, 1982, predict only one-step priming). A third important feature is that one of the components of the familiarity calculation represents a backward association. For example, with item 2 as prime and item 3 as target, the terms S_{23} and S_{32} both enter the equation, the former a forward association and the latter a backward association. Empirical findings with respect to both range effects and backward priming effects are discussed later.

Mapping From Familiarity to Reaction Time

The mapping from familiarity to reaction time is made using the assumption that extreme values of familiarity (high or low) lead to fast and accurate responses (yes and no, respectively) while intermediate values lead to slower and less accurate responses (Norman & Wickelgren, 1969; Ratcliff, 1978).

To explicitly model this assumption, the familiarity value given by the computation of the strength of the compound cue is used as the drift rate in a diffusion (random walk) decision process (Ratcliff, 1978, 1981, 1985, 1988). The diffusion process is used because it provides an excellent account of reaction time, accuracy, reaction time distributions, and the growth of accuracy as a function of time, as well as the relations among these measures, in a number of different kinds of tasks (e.g., item recognition of sub- and supraspan lists, item recognition in discrete and continuous list paradigms, letter matching, and response signal and deadline recognition procedures).

In the diffusion process, there are two response boundaries, one for positive responses and one for negative responses. The larger the drift rate, the faster the diffusion process approaches the positive boundary; the smaller the drift rate, the faster the process approaches the negative boundary. There are two sources of variance in the diffusion model: one is variability

Table 1
Sample Calculation of Familiarity: Numerical Example

Cue	Target 1	2	3	4 . . .	Cue	Target 1	2	3	4
1	S_{11}	S_{12}	S_{13}	S_{14}	1	1.0	1.0	0.2	0.2
2	S_{21}	S_{22}	S_{23}	S_{24}	2	1.0	1.0	1.0	0.2
3	S_{31}	S_{32}	S_{33}	S_{34}	3	0.2	1.0	1.0	1.0
4	S_{41}	S_{42}	S_{43}	S_{44}	4	0.2	0.2	1.0	1.0
5	S_{51}	S_{52}	S_{53}	S_{54}	5	0.2	0.2	0.2	1.0
.					.				
.					.				
.					.				

As a first illustration of the familiarity calculation, the strengths on the prime and target are not weighted differentially, and the context cue is set to 1.0 (so it can be ignored). Then

$$F(\mathrm{i},\mathrm{j}) = \sum_{\mathrm{k}} (S_{\mathrm{ik}} S_{\mathrm{jk}}).$$

$$\begin{aligned}\text{So } F(2,3) &= S_{21}S_{31} + S_{22}S_{32} + S_{23}S_{33} + S_{24}S_{34} + S_{25}S_{35} + S_{26}S_{36} + \cdots \\ &= 1.0 \times 0.2 + 1.0 \times 1.0 + 1.0 \times 1.0 + 0.2 \times 1.0 + 0.2 \times 0.2 + 0.2 \times 0.2 + \cdots \\ &= 2.48 + \cdots \\ \text{and } F(2,5) &= S_{21}S_{51} + S_{22}S_{52} + S_{23}S_{53} + S_{24}S_{54} + S_{25}S_{55} + S_{26}S_{56} + \cdots \\ &= 1.0 \times 0.2 + 1.0 \times 0.2 + 1.0 \times 0.2 + 0.2 \times 1.0 + 0.2 \times 1.0 + 0.2 \times 1.0 + \cdots \\ &= 1.20 + \cdots\end{aligned}$$

Items 2 and 3 share associates (2 → 3 and 3 → 2) so that $F(2, 3)$ is greater than $F(2, 5)$.
If weights are introduced so that $F(\mathrm{i},\mathrm{j}) = \sum_{\mathrm{k}} (S_{\mathrm{ik}}^{.1} S_{\mathrm{jk}}^{.9})$, then

$$F(2,3) = 3.49 + \cdots$$

and

$$F(2,5) = 3.26 + \cdots$$

around the drift rate in the comparison process (given in the diffusion model), and the other is variability across items in their representation in memory (provided by the Gillund and Shiffrin model). A complete description of the diffusion model is given by Ratcliff (1978).

In order to make the mapping between the retrieval model and the diffusion process complete, a transformation will probably be required to convert familiarity to drift rate. In the simplest case, this would be a linear transformation, but selection of a transformation has not yet been made because the detailed parametric tests of the model that would be necessary to determine this transformation have not yet been carried out.

It is interesting to note that the notion of a compound cue might be able to account for sequential effects in which a response to a test item is faster if the same response is required to the immediately preceding test items, as compared with mixed positive and negative test items. The account would be that the test probe is a compound consisting of the test item and one or more of the preceding test items (with the most recent weighted most heavily) and that familiarity of this compound will be higher if all items are positive than if they are mixed and will be lower if all items are negative than if they are mixed. Thus responses will be faster and more accurate for consistent than mixed sequences (see Falmagne, 1965; Theios, 1973).

Extension to Textual Information

The model proposed by Gillund and Shiffrin (1984) forms the basis of the retrieval model described here, but considerable enhancement is required to account for memory for text. First, the Gillund and Shiffrin model assumes that encoding is accomplished by a simple buffer system that holds only four words at a time. This is not adequate for text because, for example, McKoon and Ratcliff (1980a; Ratcliff & McKoon, 1981a) have shown that priming can be obtained between concepts that are much more than four words apart. Second, the model has no mechanism for storing labeled relations and in retrieval computes the familiarity only of individual items. Although this is sufficient to account for most experimental results with item recognition for single words, recognition of sentences requires access to information about order and grammatical relations. For example, if a subject studied "John hit Bill," then in the Gillund and Shiffrin model the probes "John hit Bill" and "Bill hit John" would have equal familiarity values, even though "Bill hit John" does not correctly represent the meaning of the studied sentence.

To encompass the encoding of sentences, the retrieval model assumes the buffer model for text encoding proposed by Kintsch and Vipond (1979; Kintsch, 1974). In this model, propositions are the units in short-term memory. They are processed in cycles, with some small number of propositions (four to eight) processed on each cycle. Important propositions (e.g., topics) are maintained in short-term memory from one cycle to the next, so that connected structures can be built to represent relations in the text.

The propositions processed on each cycle are encoded into long-term memory so that single-word concepts are connected to each other and propositions are connected to each other. Strengths among concepts in the same proposition are assumed to be greater than strengths among concepts in different propositions, and in general, the relative strengths between concepts and between propositions are determined by the extent to which the items are processed together in the same cycle. The strengths among the single-word concepts determine the familiarity of compounds made up of single word prime-target cues. Similarly, when propositions are presented as test items (in the form of simple sentences), the familiarity of a compound made up of a prime proposition and a target proposition is computed as in Equation 1, using proposition connection strengths (cf McKoon & Ratcliff, 1980a, Experiment 2; Seifert, McKoon, Abelson, & Ratcliff, 1986).

The retrieval model is intended to apply not only to situations in which items are presented for test in standard memory paradigms but also to tests of on-line processing such as those done in reading, and perhaps even as a component process during on-line processing in reading. For example, during reading, the contents of short-term (or working) memory are the words explicitly stated in the text plus the propositions (meanings) formed out of those words. These kinds of information combine to form the compound that determines the familiarity of the information and this compound is assumed to be used in the process of on-line retrieval (of information needed to comprehend the text) that takes place during reading. If a test item is presented immediately after reading, as is often done in experiments designed to investigate on-line processing (e.g., Dell, McKoon, & Ratcliff, 1983; Swinney, 1979), then that test item is assumed to enter a compound cue with the compound that already represents the text just read. The result determines familiarity and thus the response for the test item. Foss (1982) has presented data that suggest that there is no (or at least very slow) decay in words studied in sentences. From this he argues that in discourse, semantic priming results from concepts remaining in an active state or in short-term memory (e.g., Kintsch & Vipond, 1979) because of processing requirements in discourse. For the retrieval theory, this would require the assumption of larger or more complex compound cues during reading than in single-item retrieval situations.

In the memory representation of a text, labeled relations (e.g., grammatical or semantic) are assumed to be represented in a different component of the representation than simple strength associations. When a sentence is presented for recognition, the strength component is assumed to lead to an assessment of familiarity relatively quickly in processing, while the relational component is assumed to give information only later in processing. The delay may reflect different information available at a later time in a unitary retrieval process, or it may reflect implicit retrieval of the kind used for recall in the Raaijmakers and Shiffrin (1981) model. Evidence for this two-component retrieval of sentences has been presented by Ratcliff and McKoon (in press), and is reviewed below.

*Relation Between the Retrieval Theory and ACT**

At first sight, the matrix of connection strengths proposed as the representation of text by the retrieval model (taken from Gillund and Shiffrin) seems quite different from the network representation proposed by ACT* (Anderson, 1983). In fact, it is easy to translate from one theory to the other. The ACT* model assumes that concepts are represented in a network, with links serving as labeled associations between the concepts (relation, object, etc.). In calculating asymptotic activation of the network, Anderson and Pirolli (1984) converted from the network representation to a matrix representation to facilitate computations. They assumed that entries in the matrix consist of strengths between concepts, with the strengths being positive for connected concepts and zero for concepts not connected. This is like the Gillund and Shiffrin representation, with one main difference: When items are not connected in the Gillund and Shiffrin model, the strengths are assumed to be set at some residual value rather than zero. Thus the representational assumptions for the two models are surprisingly compatible.

The two models also are similar in processing assumptions in that both allow for more than one component of retrieval processing, that is, for different kinds of information to become available at different points in time. ACT* explicitly allows for activation of nodes followed by retrieval of pathways of connections for evaluation. So, although the models currently differ in specific processing assumptions, the more global similarities suggest that ACT* could be altered to substitute the kind of processing proposed by the retrieval model for priming.

Summary

According to the retrieval model, predictions for priming effects can be deduced from the contents of short-term memory with the more recent items weighted more. The items in short-term memory are used to form a compound cue, and the familiarity of this cue determines accuracy and response time. The value of familiarity for any compound depends, in turn, on the retrieval structure in memory. Only when the items in the compound are directly connected to each other or each directly connected to some other common image will familiarity be relatively large and priming effects obtained. (For other than the Gillund and Shiffrin, 1984, implementation, only items that are directly connected will produce priming effects.) This retrieval model is applied to textual information through integration with the text structure models of Anderson (1983) and Kintsch (1974) and the text encoding model of Kintsch and Vipond (1979). In addition, the diffusion retrieval model of Ratcliff (1978) is a good candidate to translate from familiarity of the compound cue to reaction time and accuracy.

Comparisons Between the Retrieval Model and Spreading Activation

Spreading activation has been used to account for a large number of empirical phenomena, many of which are listed in Table 2. The retrieval model is a viable alternative to spreading activation because it can also account for these phenomena, as discussed in the sections that follow. These sections illustrate applications of the retrieval model, compare retrieval model explanations of specific effects with spreading activation explanations, and show several cases in which the retrieval model provides a more natural explanation than spreading activation.

Table 2
Comparison of the Spreading Activation and Retrieval Theories for Priming

Effect	Spreading activation	Retrieval
Range (see Experiment 1)	Parameter of the model.	Direct and singly mediated.
Decay	Parameter of the model.	Number of items in the compound cue (three items in the cue = two-item decay).
Onset	In older versions, a function of the number of links. In recent versions, constant onset.	A function of the time to encode the prime into the compound cue with the target.
Automatic processes versus strategic process	Target is faster to match if previously activated. If anticipation is wrong, slower because have to invoke other processes (Anderson, 1983).	Build compound cue (prime + target). Implicit recall/retrieval with retrieved information used to form compound cue. May require a criterion shift to produce inhibition in response time.
Neutral priming condition	Produces no activation.	Previous target replaces a neutral prime in the compound cue.
Forward versus backward priming	Backward is a problem because priming can only occur in forward direction.	Compound cue has both forward and backward components.
Priming of ambiguous words	Both meanings are activated, then later one meaning is selected by context.	Compound cue has higher familiarity for both meanings as long as the word is in the compound cue; later the compound cue contains words associated with one meaning.
Organized material	First stage: activation. Second stage: evaluation of activated pathways. Cannot occur in parallel.	Familiarity process provides some evidence and recall process provides other evidence. These can operate independently and possibly in parallel.
Continuous versus discrete processing	Continuous growth of activation.	Building the compound cue is not all or none; features from the prime are gradually added until the target is presented.
Priming and memory structure	More activation implies more connections and larger priming effects	Greater familiarity implies greater overlap and larger priming effects.

Decay of Priming

As other test items intervene between prime and target, the amount of facilitation on the target is reduced. According to the retrieval model, the decay must be rapid because the effect of an earlier prime must be small and must get smaller as more items intervene. Empirically, data show that decay of priming is rapid. For associative priming (e.g., between baby and concrete when subjects studied "the baby hit the concrete"), Ratcliff et al. (1985) showed that facilitation had decayed to one third of its initial value with only one item intervening between prime and target. This decay of priming is one factor that differentiates theoretically between the spreading activation and retrieval models.

As an aside, it is useful to define quantitatively what is meant by decay. Mathematically, decay is often represented by an exponential function. Given an exponential function, $y = y_0\exp(-t/\tau)$, then the mean of the exponential is τ. The measure of decay most often used for the exponential is the time constant and mean τ, so that when $t = \tau$, the height of the exponential is y_0/e where $e = 2.718$, that is, about one third of the starting height. Thus we talk about decay rate in terms of the time constant τ or, equivalently, the number of intervening items before the function has fallen to one third of its initial value.

For the retrieval model, decay is predicted from the number of items that make up the compound cue in short-term memory. If the compound is made up of only two items (Gillund & Shiffrin, 1984), as in Equation 1, then the target must immediately follow the prime in the test list in order for the compound cue to contain them both and produce facilitation. With even one intervening item, the prime would not be part of the compound cue and there would be no facilitation. However, it is quite possible for the compound cue to contain more than two items, with the earlier items weighted less than the prime and the prime weighted less than the target. But even if this were done, priming effects would still be expected to decay at least within one or two intervening items because the earlier primes with low weights would contribute almost nothing to the familiarity of the compound cue. The reason that most of the weight must be on the target is that the response has to be made to the target; for example, if the weight was equally distributed over the last three items, then if items i-1 and i-2 were "old" items and item i was "new," the response would be "old" even though the target was "new." We know this does not happen, so this constrains the weights so that most weight is on the target. We also know that sequential effects in reaction time usually extend only a short range back in the sequence of prior responses. Both of these suggest the model has to assume low weights on earlier items, and weights larger than zero only on one or at most two items prior to the target.

In contrast, spreading activation theory could allow any value of decay; this is a free parameter, constrained only by fits to data. For example, Anderson (1976) assumed a periodic dampening of the network, and Anderson (1983) and some connectionist models assume decay as a function of time (e.g., McClelland & Rumelhart, 1981). However, these specific assumptions have not been chosen via a comprehensive examination of parameter spaces, so it is not clear how much latitude the models have in their assumptions about decay of priming. Current spreading activation accounts of decay of priming were determined in a post hoc fashion by assumptions that depend directly on the decay data being fitted. In contrast, the retrieval theory makes a prediction constrained by the weights that must be assumed on the last few items in memory (see the argument in the prior paragraph).

It is important to note, however, that the specific mechanisms and parameter values adopted in ACT* are consistent with rapid decay. When a node is activated from an external source, it remains activated after the source is removed for a period that can be as short as 400 ms, and then activation decays rapidly (in tens of milliseconds). Activation in a node can also be maintained by an internal source, but again, once the source is removed, activation decays rapidly. With these mechanisms already in place, ACT* can account for one-item decay functions. This issue is discussed further at the end of Experiment 2.

Automatic and Strategic Effects in Priming

A basic distinction in spreading activation theories is between automatic and strategic processes (Logan, 1980; McKoon & Ratcliff, 1979, 1986b; Neely, 1977; Posner, 1978; Posner & Snyder, 1975; Ratcliff & McKoon, 1981b). Automatic processes have been assumed to arise from a rapid spread of activation, whereas strategic processes have been assumed to reflect slower evaluation of activated pathways.

In the retrieval model, automatic processes reflect the assembly of the compound cue in short-term memory and the assessment of its familiarity. Slower strategic processes are explained in terms of another process involving explicit retrieval or recall. In the Gillund and Shiffrin (1984) framework, this would take the form of the recall process proposed by Raaijmakers and Shiffrin (1981). Within the ACT framework, there are other processes that would serve the same purpose as the explicit recall process in the Gillund and Shiffrin model.

Onset of priming. One critical aspect of automatic processes is their speed of onset. In experiments that manipulate stimulus onset asynchrony (SOA) between prime and target, facilitation from automatic processes rises above zero by about 100 ms and asymptotes by about 150 ms. Priming with this speed of onset has been found for preexperimental associations (e.g., dog–cat) and for newly learned associations (den Heyer, Briand, & Dannenbring, 1983; McKoon & Ratcliff, 1986b; McKoon, Ratcliff, & Dell, 1985; Neely, 1977), for example, between words in the same studied sentence (Ratcliff & McKoon, 1981b). For strategic processes, on the other hand, the onset of facilitation is delayed by several hundred milliseconds (Neely, 1977; Posner & Snyder, 1975; Ratcliff & McKoon, 1981b).

In early spreading activation theories, it was assumed that activation took time to spread through a network, about 100 ms per link (e.g., Anderson, 1976; Collins & Quillian, 1969). So the time of onset for priming from one concept to another would be a function of the number of links between them. Ratcliff and McKoon (1981a) tested this prediction directly by examining the time course for onset of priming, both for concepts near to each other in newly learned texts and for concepts far apart. Results showed that although there was greater asymptotic priming for concepts near to each other, priming for the near concepts did *not* have an earlier onset than priming for the far apart concepts. Contrary to the then current models of spreading activation (Anderson, 1976; Collins & Quillian, 1969; Collins & Loftus, 1975), time of onset was not a function of number of intervening links.

In response to this and related results, Anderson (1983) revised the ACT spreading activation model so that activation spreads quickly between concepts (see also Wickelgren, 1976) and the onset of the growth of activation is very rapid (e.g., 5–10 ms per link). Reaction time differences among conditions are now attributed to different rates of growth of activation at nodes, where the rate is determined by the strength of connections.

In the retrieval model, the onset time for priming is predicted to be the same no matter how strong the connection in memory between prime and target. The onset time is simply a function of the time to form the compound cue. Thus, the results of Ratcliff and McKoon (1981a) are a direct prediction of the retrieval model.

In some situations, there has been a failure to obtain semantic priming at short SOAs (den Heyer, 1986; Neely & Durgunoglu, 1985; Smith, 1979). Smith (1979) found that if subjects were required to perform a letter search task on the prime word, then semantic priming was not obtained. Although the factors governing the failure to obtain semantic priming are not yet fully understood, it is clear that automatic semantic priming is not automatic in the sense that it occurs whenever prime and target are semantically related, under any and all circumstances. It seems more reasonable to suppose that when the subject is set up to perform in such a way as to process semantic relations, then the process runs off quickly and satisfies the criteria for an automatic process proposed by Posner and Snyder (1975). The compound cue account of this is simple but post hoc. When semantic priming is not obtained, the prime is not encoded into the compound cue; when semantic priming is obtained, the prime is part of the compound cue. To put this slightly differently, a compound cue will carry information relevant to the task at hand; if semantic relationships are not relevant (from the subject's point of view, as in a letter search task), then they will not be part of the information in the compound cue. However, before any significant theoretical progress can be made on this issue, it will be necessary to obtain a clear understanding of what experimental factors govern the failure to obtain semantic priming.

Pre- and postlexical priming effects. Current views of word recognition often distinguish between pre- and postlexical processes (Forster, 1981; Onifer & Swinney, 1981; McKoon & Ratcliff, 1987; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Seidenberg, Waters, Sanders, & Langer, 1984). Given a prime followed by a target presented for lexical decision, prelexical processes occur before the target is presented and give facilitation on the target by speeding lexical access. Postlexical processes do not occur until after the target is presented and give facilitation by speeding lexical selection and decision processes. Prelexical effects can be either automatic or strategic. When automatic, they would usually be attributed to spreading activation, and when strategic, to the development of subjects' expectations that primes will be followed by associated words as targets. According to the retrieval theory, automatic facilitation between high associates would be due not to spreading activation but to the formation of a compound cue made up of the prime and target. Thus, the automatic component of facilitation would be neither pre- nor postlexical, as those terms are usually used, but a product of the joint association of prime and target.

The same spreading activation process that gives rise to automatic facilitation in lexical decision is sometimes assumed to also give rise to automatic facilitation in naming latency. It is often argued that the two tasks involve mainly the same processes up to the response stage, at which point lexical decision is subject to decision strategies and postlexical processes that are not part of naming. However, recently West and Stanovitch (1986) have shown effects of syntax, which would be labeled postlexical, on naming latencies. This result could be taken to suggest that lexical decision and naming processes may be more similar than previously believed. Instead, we believe that although the two tasks could be affected by some variables in similar ways, their processes are fundamentally different. In lexical decision, subjects can provide a response based on some notion of familiarity such as "I have seen something like that before," and the response is a binary decision. In contrast, the naming task requires the subject to retrieve one of tens of thousands of names from the mental lexicon. Although both these tasks require some kind of access to the mental lexicon, we believe the processes underlying these tasks probably differ in significant ways so that simple subtractive logic is not applicable.

Effects of Neutral Primes

In experiments that examine priming between single words, a neutral condition is used to give baseline response times against which facilitation versus inhibition can be measured. The neutral prime is often a row of *X*s, a row of random letters, or some word such as "ready." Which kind of neutral is used does not appear to matter in item recognition; we have used all three and found no significant differences. In lexical decision, it does seem that there are different effects of different neutral primes; for example, de Groot, Thomassen, and Hudson (1982) found differences between a neutral word and a row of *X*s.

In terms of spreading activation, neutral primes provide baselines because they activate no targets in memory and so produce no facilitation. This would be true for any of the neutral primes. The only possible differences among neutral primes might be in encoding; a row of *X*s might, for example, require less attention than a string of random letters, and so leave more processing capacity or time for the target. A word such as "ready" however, may require processing as a meaningful letter string before it can be determined that it is a neutral prime.

The retrieval theory allows several possible interpretations of the effect of neutral primes, all different from the spreading activation interpretation. The theory could assume that the neutral prime was related to targets by the average value of residual strength. This solution would produce the correct behavior for neutral priming conditions, but is not particularly satisfying. The theory could not assume that strength was set to zero because then the familiarity of the compound would be zero. Another possibility would be to assume that subjects change the relative weights of prime and target when the prime is neutral. However, this assumption would not fit the data; putting more weight on the target would lead to faster reaction times, and the neutral priming condition does not usually have faster reaction times. The interpretation we favor is that neutral primes are not included in the compound cue (after some practice with them); instead the compound cue is made up of the previous target and the current target. This compound cue would have the same familiarity as if the previous target were actually the prime. Support for this interpretation is provided in Experiment 2 below.

Forward Versus Backward Priming

The forward or backward direction of priming refers to direction either in time or in memory structure. For direction in memory structure, spreading activation theories predict that the amount of priming is determined by the strength of association from the prime to the target in a forward direction. However, the retrieval theory makes a different prediction. In the Gillund and Shiffrin (1984) implementation (Equation 1), one of the terms involves backward association (S_{kj}), so that strength of association from target to prime is an important determiner of the size of the priming effect.

The empirical results that are available support the retrieval theory. Koriat (1981) and Seidenberg et al. (1984) have shown backward priming in lexical decision, between prime and target pairs in which the association from prime to target was nonexistent (or very weak) and the association from target to prime was strong. Although both sets of experiments used SOAs that were long enough to allow strategic processes to affect priming (650 ms and 500 ms, respectively), it seems unlikely that strategic processes would be responsible for priming when there is no association (or only a weak association) from prime to target. Both forward and backward associations are also demonstrated in memory for sentences. It might be thought that associations between words were stronger in the forward direction (left to right) than in the backward direction (cf. Kolers & Roediger, 1984), but Ratcliff and McKoon (1978) and McKoon and Ratcliff (1980a) have shown that priming between concept words is independent of the presentation order within the study sentences.

The second phenomena that has been labeled forward versus backward priming concerns the direction of priming in time. Kiger and Glass (1983; see also Schustack, 1981) presented the prime for a target *after* the target, and still obtained priming on the target response times. For the retrieval model, this finding presents no problem; the compound cue would be formed in the same way as if the prime were presented first, and there would be an SOA function that would map out the drop in influence of the prime as the target was processed by itself (i.e., at long delays, processing on the target might be completed before the prime was presented). Spreading activation would account for these results in a similar fashion (e.g., Logan, 1980).

Priming of Multiple Meanings of Ambiguous Words

There is considerable data that both senses of an ambiguous word are primed directly after the word is presented, but that later in processing (after other intervening words), only the sense appropriate to the context is primed (cf. Onifer & Swinney, 1981; Swinney, 1979). A similar pattern is found when information possibly relevant to an anaphor shows facilitation immediately after presentation of the anaphor, but not later in processing (Dell et al., 1983; McKoon & Ratcliff, 1980b). Spreading activation explains these effects by postulating initial activation followed by decay due to the processing of other material. The sense appropriate to the context is maintained because it receives activation from other compatible sources.

The retrieval model explains these effects in terms of the contents of the compound cue at the time the target is presented. A compound cue that includes an ambiguous word (e.g., "pot," the word in the text) and a target test word representing one of its senses (e.g., "marijuana") forms a compound cue that gives a high value of familiarity. Once the prime word is no longer part of the compound, as will happen after one or two intervening words, then the compound cue will be made up of one or more of those intervening words, plus any concepts held in the short-term memory representation of the text, plus the target. This compound cue is more likely to be related to a contextually relevant target, and so familiarity will be higher for a target representing the sense of the ambiguous word relevant to the context. Thus in the retrieval model, it is not that the contextually inappropriate meaning is suppressed. Rather, with delayed test, the ambiguous word is no longer available to support the contextually inappropriate meaning.

Priming from ambiguous words often shows context sensitivity. Even when both meanings of an ambiguous word are primed, the contextually relevant meaning may be primed more than the contextually inappropriate meaning (Simpson, 1984). One instance in which only the contextually relevant meaning is primed has been described by Seidenberg et al. (1982), who used sentences that contained (prior to the ambiguous word) other words highly associatively related to one of the ambiguous word's meanings. According to the retrieval model, the related words would determine the meaning of the ambiguous word that was retrieved from memory, and the related words and appropriate meaning would enter the compound cue with the test word. In such a context, greater priming of the contextually relevant meaning would be expected.

The explanation of the results of Dell et al. (1983) is similar. Presentation of an anaphor produces retrieval of propositions related to that anaphor, and these enter short-term memory (Kintsch & Vipond, 1979). Immediately after presentation of the anaphor, target concepts connected to the concepts brought into short-term memory will have high familiarity. After intervening words, only those concepts still in working memory will give higher familiarity to targets.

Retrieval of Organized Material

Theories based on spreading activation deal with retrieval of organized material by assuming two component processes: the first component, spreading activation, makes pathways among concepts available for the second component, which is evaluation in some theories (Collins & Loftus, 1975) or the operation of productions in other theories (Anderson, 1983). Both these two-component theories and the retrieval theory predict two phases in retrieval. In the compound cue/retrieval theory in the Gillund and Shiffrin framework, the first component of processing would be the measurement of familiarity and the second would be the recall component of the Raaijmakers and Shiffrin (1981) model.

The two phases of retrieval have been demonstrated in a recent experiment by Ratcliff and McKoon (in press). Subjects studied lists of sentences, some active and some passive. Then sentences were presented for recognition; subjects were required to decide whether a test sentence matched a studied sentence in meaning. For example, if "John hit Bill" was studied, the correct response for the sentence "Bill was hit by John" would be yes; correct responses for the sentences "Bill hit John" and "John was hit by Bill" would be no. There were also new test sentences (correct response: no) that contained no words from studied sentences. A response signal procedure was used, so that the time course of retrieval could be examined. Results showed that early in processing (by about 400 ms), subjects could discriminate the new sentences from sentences which contained studied words, but that positive and negative versions of the studied sentences could not be discriminated from each other until about 750 ms of processing time had elapsed. In a second experiment, irreversible sentences were included. In these sentences, reversing the positions of the subject and object resulted in an anomalous meaning (e.g., "the secretary wrapped the package"). Early in processing (until about 500 ms), the probability of responding positively increased at about the same rate for both the correct and the anomalous versions of these sentences. After 500 ms, negative responses to the anomalous versions increased until, by 2,000 ms, subjects were highly accurate. Recently, Gronlund and Ratcliff (1988) have obtained similar results using paired associates and a decision task that required subjects to decide whether a test pair was intact (the two items were in the same study pair) or rearranged (the two items were from different study pairs). The intact/rearranged information was available only after 600 ms of processing time, whereas information that discriminated old items from completely new items was available by about 300 ms of processing time.

These results support the two-component retrieval hypotheses of both spreading activation theories and the retrieval theory. The first component involves familiarity or amount of overall activation and not relational information. Only the second component involves relational information, and so only the second component can discriminate between correct and incorrect relations.

Continuous Versus Discrete Processes

Recently, Yantis and Meyer (1988; see also Meyer, Yantis, Osman, & Smith, 1985) presented an application of a new method for examining discrete versus continuous processes in priming in lexical decision. Subjects performed a lexical decision task in which a prime preceded the target letter string (no response was required to the prime). The SOA between the prime and target was manipulated to be short (with little priming), long (priming had asymptoted), or medium (in between the other two conditions). The SOA was adjusted for each subject individually so that performance in the medium condition would produce a priming effect about half the size of the priming in the long condition. The data of interest were the reaction time distributions in the medium SOA condition. Yantis and Meyer described two possible results. First, the distribution could be a probability mixture of processes, with slower processes from unprimed responses and faster processes from primed responses. This result would imply that processing was all or none such that the target was either primed or not primed but never partially primed. The second possible result was that the distribution would be narrower than what would be predicted by a probability mixture of the primed and unprimed distributions, leading to the conclusion that the process was continuous. The data supported this second prediction.

For spreading activation, the explanation of these results is natural because activation is assumed to grow continuously with time. However, superficially, continuous processing may seem at odds with the notion of a compound cue that might be thought the result of an all-or-none process (either the prime is part of the compound or it is not). However, we simply assume that the process of building a compound cue evolves over time (e.g., as features are encoded into the compound cue), and that familiarity is increased as information is added to the compound cue. In a procedure in which the SOA between prime and target is kept short, we assume that part of the prime is encoded into the compound and when the target is presented, encoding of the target into the compound cue interrupts further encoding of the prime. With these assumptions, and the assumption that familiarity is continuously available for the decision process, the retrieval theory is consistent with continuous processing. One might think from the earlier description of the theory that simultaneous presentation of prime and target would be optimal for formation of a compound cue, but we believe instead that fairly rapid sequential presentation will be optimal because there is no spatial division of attention when the prime and target are sequentially presented as there is with simultaneous presentation. Also, with rapid sequential presentation, the amount of processing on the prime could be made optimal (e.g., 150 to 250 ms) for forming the compound cue.

Summary

The retrieval model, by its structure, makes strong predictions about the decay of priming, the onset of priming, and forward versus backward priming. Data are, in these cases, consistent with the predictions of the retrieval model. Spreading activation, on the other hand, accounts for these effects in a more equivocal or post hoc fashion. The retrieval theory also makes strong predictions with respect to the range of priming effects through memory structures and with respect to the effects of items intervening between prime and target and changing the contents of the compound cue. These predictions are tested in the experiments presented in the next sections, and support is obtained for the retrieval model over spreading activation models.

Experiment 1: Range of Priming

The range of spreading activation through memory is one of the main features of the spreading activation model, yet one that has not been clearly evaluated empirically. Spreading activation predicts that the amount of priming between two concepts is a function of the distance between them. In contrast, the retrieval model predicts that there will be priming between two concepts only if they are directly connected to each other or if they are separated by no more than one associate (in the Gillund and Shiffrin, 1984, implementation of the model). These contrasting predictions are important in distinguishing between the two models, but previously there has been no clear empirical test.

The earliest work that provided evidence about the range of activation was done by Collins and Quillian (1969). They found that distance between concepts affected verification time: Responding yes to "Is a robin a bird?" was faster than responding yes to "Is a robin an animal?" where animal was assumed to be farther from robin than bird in a hierarchy of concepts. However, Smith, Shoben, and Rips (1974) showed that distance in the hierarchy was confounded with similarity and that similarity rather than distance was the main predictor of performance. Since then, it has become clear that the semantic verification paradigm is not likely to be able to address questions about distance relations in semantic memory because there are so many other variables that are confounded with distance in the semantic network (see Glass & Holyoak, 1975; McCloskey, 1980; also Ratcliff & McKoon, 1982).

To avoid such confounds in the study of distance effects, McKoon and Ratcliff (1981a) used newly learned materials, short paragraphs in which distances between concepts could be tightly controlled, and a priming procedure in which distances could be directly examined. For example, in the paragraph that begins "The youth stole a car. The car sideswiped a pole. The pole hit a hydrant," "pole" should be relatively closely connected to "hydrant" while "youth" is relatively far from "hydrant." The effect of one test item (a prime) on the immediately following test item (the target) was investigated in item recognition ("Was this word in one of the studied paragraphs?"). It was found that when the prime and target were from the same paragraph, the closer the prime and target were in the paragraph, the faster were response times on the target. Primes facilitated target response times even when as many as four concepts (or three propositions) separated them in the paragraph (see also generalizations to visually presented materials in McKoon, 1981; McNamara, Ratcliff & McKoon, 1984).

At first thought, this would seem to be strong evidence against the retrieval model of priming. However, there are two reasons that such long-distance priming might be expected from the retrieval model. First, facilitation is assessed against a condition in which the prime is from a different (unrelated) paragraph from the target. This may not be the best baseline condition from the point of view of the Gillund and Shiffrin (1984) implementation, because the residual strength between nouns from different paragraphs may be lower than the residual strength between nouns from the same paragraph, and this would lead to a lower value of familiarity when the prime and target come from different paragraphs than when they come from the same paragraph. A better baseline may instead be the condition in which prime and target are far apart in the same paragraph. Relative to this condition, there should be no priming except for the very nearest concepts.

The second way in which the retrieval model could show long-distance priming stems from the way text is assumed to be encoded. In the Kintsch and Vipond (1979) model, as many as seven propositions can be held in short-term memory at once. This means that all the propositions of the paragraphs used by McKoon and Ratcliff (1981a) could have been in memory at the same time. Furthermore, some propositions are held over from one processing cycle to the next, so that long-range connections can be established. In sum, the data from McKoon and Ratcliff (1981a) can be made compatible with the retrieval theory.

Another set of experiments designed to test distance effects has been performed by de Groot (1983) and Balota and Lorch (1986), who attempted to measure distance effects in semantic (or lexical) memory. These experiments used two concepts that are either connected by one mediating item (beach–box) or directly connected (sand–box). Spreading activation must predict priming in the mediated case, because beach is highly associated to sand and sand to box, so that the distance from beach to box in memory should be relatively short. However, obtaining mediated priming is difficult; neither de Groot nor Balota and Lorch obtained mediated priming in lexical decision, and when Balota and Lorch did obtain mediated priming in naming latency, the effect was small (and may be due to different mechanisms than those operating in lexical decision, as noted above). Thus it seems unlikely that priming would be found for distances greater than one mediator.

The conclusion is that one of the main tenets of spreading activation, that activation spreads through a network of connected associates in memory, has little empirical support. Even when two concepts are separated by only one intervening associate, priming between the two concepts is very small and difficult to detect (Balota & Lorch, 1986; de Groot, 1983). The one study that does show longer range effects in priming (McKoon & Ratcliff, 1981a) can be interpreted in terms of the retrieval model as well as the spreading activation model. Thus, a major and critical assumption of spreading activation has minimal empirical support.

Experiments 1A and 1B were designed to examine the range of priming with respect to the spreading activation and retrieval theories. Stories were written that were similar to those used by McKoon and Ratcliff (1981a, Experiment 1), except that they were long enough that all the propositions could not possibly fit in short-term memory at the same time (about 630 words). The stories also had a linear structure such that any character or event was connected to the characters and events immediately preceding and following it, but not to any others. The stories were written to have no long-range connections among propositions, as shown in the example in Table 3 (we call them "rambling" stories).

For each story, there were two target test sentences from the last third of the story. According to the retrieval theory, these targets can be primed only by information close to them in the representation of the story in memory. Because of the linear structure of the stories, this would have to be information from sentences immediately preceding or following the targets in the story. In contrast, by spreading activation theories, amount of priming should be a function of distance between the priming and target sentences. Sentences at medium distances from the target should prime the target more than sentences far from the target (so long as the medium distance is not beyond the range of spreading activation), and sentences at medium distances should prime less than sentences close to the target. In Experiment 1A, the primes were near to the target or far from it in the same story, or from a different story. Experiment 1B used these three conditions plus a fourth condition in which the prime was a medium distance from the target (see Table 3 for examples).

Table 3
Example of a Story Used in Experiments 1A and 1B

Harvey had been unemployed for 6 months when he read an ad in the paper for a traveling salesman job. The interviews were to be held today only, from 1–3 P.M. Harvey looked at his watch and realized that he didn't have a lot of time. He dressed quickly and jumped into his beat-up sedan. Harvey got on the freeway, but there had been an accident and traffic was jammed for miles, so he got off at the nearest exit and took a shortcut through a suburb.

As Harvey was rounding a curve at 50 mph, a cat ran out in front of him and he swerved to avoid hitting it. Unfortunately for Harvey, his steering picked that very same moment to go out and he headed for a nearby telephone pole. He was going so fast that the front bumper sheared off the top of the telephone pole, leaving Harvey shaken, but otherwise unharmed.

The telephone pole continued to sail through the air another 500 feet down the block before it crashed into a fire hydrant conveniently located in Mrs. Bambeck's front yard. The telephone pole had picked up speed on its way down the block, so when it hit the hydrant the force of the blow was enough to "knock its block off," releasing a torrent of water. The water gushed out of the hydrant, running down the street in both directions. However, most of the water headed for the decorative boulders Mrs. Bambeck had arranged in her front yard in the shape of a rabbit. The boulders acted as a dam and served to route much of the water toward Mrs. Bambeck's prize tulip and pansy flower beds.

At the moment the first gush of water hit her precious ground cover, Mrs. Bambeck was in the kitchen grinding some mocha java coffee beans in the little electric grinder her nephew had given her last Christmas. She let out a little scream, not too loud so as not to wake the cat, Bartholemew, and ran out the door without bothering to take out her hot rollers (which were quite cold by now) to try to save her flowers.

As Mrs. Bambeck was out in the garden, sloshing around and trying to keep her prize pansies from floating away, the coffee grinder in the kitchen continued to grind and grind and grind. It finally became so hot that the cord began to burn, and then the polyester curtains which were near the cord, and then the cabinets. Soon the whole kitchen was on fire. Little Jimmy Werner was walking to his friend's house when he passed the Bambeck house and saw smoke seeping out from under the door. He ran home as fast as his little legs would carry him and told his mother, who then called the fire department.

The big, new engine was rolled out for the first time. The chief was anxious to see how his new toy would perform. When they got to the scene of the fire, the chief was less than delighted to discover that the hoses were too short to reach from the fire hydrant to the house. One of the firefighters pointed to a rather large pond behind the house and suggested that it be used to combat the fire. The chief hated the thought that muddy water would have to be pumped through his brand new truck, but there wasn't much choice. The neighbors were already beginning to look annoyed with the lack of action displayed by the fire department. The chief gave the command and the new fire engine began to slurp up the muddy water. In a few seconds, the house was being doused with some very dirty water.

First target: The new fire engine began to slurp up the muddy water.
Near prime: The chief was anxious to see his new toy perform.
Middle prime: The telephone pole hit a fire hydrant.
Far prime: Harvey took a shortcut through a suburb.
Second target: The cord of the coffee grinder began to burn.
Near prime: Mrs. Bambeck was in the garden trying to save her pansies.
Middle prime: The boulders were arranged in the shape of a rabbit.
Far prime: Harvey dressed quickly and jumped into his beat-up sedan.

Spreading activation theories would predict that a far-condition prime would facilitate the response on a target more than a prime from a different story, if the far prime were still within the range of spreading activation. At first, it might seem that the retrieval theory should predict no facilitation in the far condition. However, the far prime may have a greater residual strength to the target than a prime from a different story. This is because there may be some small amount of shared information between the far prime and the target, such as whether the information was studied in the first or second story. So, a small amount of facilitation in the far prime condition relative to the other-story condition would be consistent with the retrieval theory.

Table 4
Results from Experiments 1A and 1B

	Experiment 1A		Experiment 1B	
Priming condition	Response time (in ms)	Error rate %	Response time (in ms)	Error rate %
Near	1,506	7	1,543	11
Middle			1,672	10
Far	1,617	8	1,647	9
Other story	1,680	10	1,726	17
Filler true sentences	1,695	14	1,734	14
Filler false sentences	1,807	17	1,869	27

Method

Subjects. In Experiment 1A, there were 21 subjects and in Experiment 1B there were 32. All subjects participated as part of a requirement for an introductory psychology course or for payment of $6.00.

Materials. Twelve stories were written, with the number of lines per story (as displayed on a CRT screen) varying between 54 and 60, and the number of words per story varying between 627 and 650. Each story had exactly six paragraphs. The stories were written so that each episode in the story led to the next episode, but so that no episode was connected to any but the immediately preceding and immediately following episodes. The episodes corresponded roughly to the paragraphs, as in the example in Table 3. (All 12 stories are available from the authors.)

For each story, 14 test sentences were written. Two of these were target test sentences, taken from the last third of the story. For each of these targets, there were three primes: one far from the target in the surface structure of the story (the far condition), one near (within one or two sentences) to the target (the near condition), and one between the near and far primes (the middle condition). For all of these test sentences, the correct response was "true" and the wording was exactly the same as in the story, except for deletions and substitutions of referents for pronouns. The number of words in the targets varied from 7 to 11, and the number of words in the primes varied from 7 to 11. The distances from the primes to the targets in numbers of words were: near, 3.5; middle, 11.9; and far, 37.3. In a propositional representation of the stories, the numbers of concepts separating the primes and targets were 1.7, 4.0, and 6.1, respectively. These were calculated as the shortest possible path from prime to target. For example, for the second target in Table 3, the path from far prime to target is Harvey, car, pole, hydrant, Mrs. Bambeck, and coffee grinder, a distance of 5 concepts.

The other six test sentences for each story included one true test sentence and five false test sentences. The false sentences expressed clear negations of some fact explicitly stated in the story.

Procedure. All experimental materials were presented on the CRT screen of a terminal connected to a Radio Shack Color Computer, which controlled the real-time aspects of the experiment. Responses were made by pressing keys on the CRT's keyboard: "?/" for "true" and "Z" for "false."

For the two experiments (A and B), the procedure was the same. The experiments began with 50 strings of letters presented for lexical decision, to give the subjects practice at responding using the CRT keyboard. Then, for the experiment proper, there were 2 practice trials and 12 experimental trials. On each trial, two stories were presented for study. One of these was one of the 12 stories described above, and the other was a story used for another, unrelated experiment (the length of the two stories was about the same). Presentation of the two study stories began with a prompt to the subject to press the space bar on the keyboard. When the space bar was pressed, the first paragraph of the first story was displayed. Subjects were instructed to read a paragraph carefully, and then press the space bar again when they had finished reading. Then the screen was cleared and the next paragraph was presented. When all of the paragraphs of a story had been presented, there was a 3-s pause, and then the second story was presented in the same way, one paragraph at a time. The order of presentation of the two stories was randomized.

After the second story had been presented, there was a 3-s pause, and then a row of asterisks was presented for 500 ms to signal the beginning of the test list, which contained 24 test sentences. The test sentences were presented one at a time and remained on the CRT screen until the subject pressed one of the response keys. Then the screen was cleared, and if the response was correct, the next test sentence was presented after a 100-ms pause. If the response was not correct, then the word "ERROR!" was presented for 4200 ms. Also, if a response was slower than 3 s, the message "TOO SLOW!!" was presented for 2 s before the next test sentence. After the last test sentence, the message to press the space bar to start the next trial was presented.

Of the 24 test sentences given on one trial, half on average were from each of the two studied stories. For 14, the correct response was "true" and for 10, "false."

Design. In Experiment 1A, there were three conditions: the target sentence was primed by the near prime, the far prime, or a sentence from the other story that was studied. These three conditions were combined with three sets of stories (four per group) and three groups of subjects (7 per group) in a Latin square design. In Experiment 1B, there were four conditions: the prime was near, far, middle, or a sentence from the other story. These four conditions were combined in a Latin square with four sets of stories (three per set) and four groups of subjects (8 per group). In both experiments, the two target sentences for a given story were always in the same priming condition. A different random order of presentation of stories and test sentences was used for every second subject. The only restrictions on the random order were that a target test sentence could not appear in the first two positions in a test list and that the position in the test list immediately preceding a prime could not be filled by a sentence from the same story as the prime's target.

Results

Means were calculated for each subject in each condition, and means of these means are shown in Table 4. The means include only correct responses on targets preceded by correct responses on primes. Not shown are the average reading times per paragraph, 16.8 s in Experiment 1A and 17.7 s in Experiment 1B.

According to both the retrieval theory and the spreading activation theories, targets in the near priming condition should have the fastest response times; this facilitation is clearly shown in the data.

According to spreading activation, targets in the far priming condition may have faster response times than targets in the other-story priming condition, if the distance is not beyond the range of spreading activation. According to the retrieval theory, targets in the far condition would show either no facilitation at all because they were not encoded together, or only a small amount of facilitation due to residual strength. The data are consistent with both theories: There is a nonsignificant amount of facilitation in the far condition relative to the other-story condition.

The theories differ with respect to the middle priming condition. The retrieval theory predicts that target response times in this condition will be no different than in the far condition; neither prime could have been encoded with the target, and so the only possible facilitation would be due to residual strength. In contrast, spreading activation predicts greater facilitation in the middle than the far conditions because the middle prime is closer in the memory representation of the story to the target (and, we think, should not be beyond the reach of spreading activation). The data support the retrieval theory.

Analyses of variance demonstrate the results just stated. In Experiment 1A, analysis of variance showed the differences in response times between the three priming conditions to be significant: $F(2, 40) = 6.6$ with subjects as the random variable and $F(2, 22) = 10.3$ with stories as the random variable. Planned comparisons showed the difference between the far and other-story priming conditions not significant (Fs < 2.2). The MS_e was 37.9 ms. There were no significant differences in error rates. In Experiment 1B, overall differences among conditions were significant: $F(3, 93) = 3.4$ with subjects as the random variable and $F(3, 33) = 5.9$ with stories as the random variable. The near and middle priming conditions were significantly different, $F(1, 93) = 5.2$ and $F(1, 33) = 7.4$, but the far and other-story conditions were not (Fs < 2.7). The MS_e was 40.1 ms. There were no significant differences in error rates.

Conclusion

The data from Experiments 1A and 1B demonstrate that the range of priming is not large; priming information must be very close to target information to produce facilitation. Of course, it is not clear exactly how far activation should spread in spreading activation models. It may be that even the distance between the middle primes and the targets in our experiments was too far. But, if this distance, four concepts on average, is too far, then the function of the spreading part of activation theories is considerably reduced: A spreading activation theory in which activation spreads to only one or two nodes is a different theory from those currently proposed. On the other hand, the data from the experiments fit the predictions of the retrieval theory exactly. Only when priming information is close enough to target information to be encoded simultaneously in working memory does the prime give facilitation to the target in a later test.

Experiment 2

The basic difference between the retrieval model and spreading activation is that spreading activation assumes that a prime temporarily affects (activates) associated concepts in long-term memory, whereas the retrieval model assumes that the prime has no effect on related concepts in long-term memory. In spreading activation theories, if the target that follows the prime is a concept related to the prime, the target will already have been activated to some extent when it is presented, so that less time will be required for activation to reach a response threshold. Thus, activation spreading from the prime speeds responses on the target as a result of temporary modifications to the state of activation of concepts in long-term memory.

The retrieval theory accounts for priming without assuming any modifications to long-term memory. The prime affects responses to the target by entering a compound cue with the target in short-term memory. Thus, accounts of priming are given in terms of the structure of the cue set in short-term memory at the time of test.

In the experiment presented below, the spreading activation and retrieval explanations of priming are tested directly against each other. For spreading activation, priming effects are determined by the concepts that are activated in long-term memory, whereas for the retrieval theory, they are determined by the information (weighted by recency) in short-term memory. In the experiment, activation levels in long-term memory are held constant while the content of short-term memory is varied using an SOA manipulation. At very short SOAs (50 ms) between prime and target, it is assumed that the prime is not registered sufficiently in short-term memory to enter the compound cue, so that the compound cue is made up of the previous target plus the current target. At a longer SOA (300 ms), the prime does enter the compound cue, bumping out the previous target (cf. Grossberg & Stone, 1986). This SOA manipulation should not affect concepts in long-term memory that were activated by the previous target so long as the time difference is not too great and so long as the prime is not related to either the current or previous target; the concepts activated by the previous target should still be activated when the current target is presented, no matter what the SOA (50 or 300 ms) for the intervening prime.

On each trial of the experiment, subjects studied five short sentences. Then prime-target test pairs were presented, with the prime displayed either for 50 or 300 ms, and the target displayed until the subject responded yes if the target word had appeared in one of the studied sentences or no if it had not. The conditions of the experiment are shown in Table 5.

For a given target, the previous target could be a word from either the same or a different sentence, and the prime could be a word from the same sentence, a different sentence, or the neutral word "ready." According to spreading activation models, the activation given by the previous target should not be affected by an intervening prime. If the preceding target is from the same sentence, then responses on the current target should be speeded. This should be true even if the prime is from a different sentence and even if the prime is displayed for 300 ms. In contrast, the retrieval model predicts that at the 300 ms SOA, there will be time for the prime to enter the compound cue with the target, and "bump out" the previous target, so that the previous target will not affect response times. For the short SOA, priming should be a function only of the prior target and not the prime.

Method

Subjects. Twenty-seven subjects participated in one 1-hr session to fulfill a course requirement.

Table 5
Results of Experiment 2

	Previous target from same sentence			Previous target from different sentence			Previous target not controlled		
Prime	*N*	RT in ms	% error rate	*N*	RT in ms	% error rate	*N*	RT in ms	% error rate
				50-ms stimulus onset asynchrony					
Same sentence				362	634	8	351	647	6
Neutral	196	584	6	192	622	5			
Different sentence	195	591	3	180	639	8	340	667	10
				300-ms stimulus onset asynchrony					
Same sentence				377	599	5	379	618	3
Neutral	190	590	4	186	631	6			
Different sentence	177	636	7	178	642	10	329	662	11

Note. RT = response time. Empty cells in the table represent condition that were not included in the experimental design.

Materials. A set of 160 sentences was chosen from materials previously used by Ratcliff and McKoon (1978). Each sentence was of the form [THE or A] NOUN1 VERBED [THE or A] NOUN2, where the articles were optional. Nouns and verbs from these sentences were used as test items. For negative test times, there was a list of nouns and verbs from another set of sentences from the same source.

Procedure. All of the experimental materials were displayed on a CRT terminal, one terminal for each subject, with one to four subjects tested in each session. Real-time aspects of the experiment were controlled by Radio Shack Color Computers, driven by an Apple computer.

Subjects were given two kinds of practice. First, 50 strings of letters were presented for lexical decision; the subjects were told to practice speed and accuracy. Second, they were given practice on two trials identical to the experimental trials.

Following the two practice trials, there were 32 experimental trials. Each of these trials began with an instruction to the subject to press the space bar on the CRT keyboard to initiate the trial. Then five sentences were presented for study, one at a time, with each sentence displayed for 5 s and followed by a 500-ms blank interval. After the last sentence, a row of asterisks was presented for 1 s to indicate the beginning of the list of 13 test items. Each test item was made up of a warning signal (a row of periods), a prime, and a target. These were presented in the following way: the warning signal for 200 ms, then erased; the prime presented in the same location as the warning signal for either 50 or 300 ms, then erased; the target on the line below where the prime had been, displayed until the subject responded, then erased. The subject responded by pressing the "?/" key on the CRT keyboard if the target word had appeared in one of the five studied sentences, and pressing the "Z" key if it had not. If the response was correct, the warning signal for the next trial was presented after a 50-ms blank interval. If the response was not correct, the word "ERROR!" was presented for 3,600 ms before the next warning signal. After the last test item, the instruction to press the space bar to begin the next trial was presented.

Design. There were 16 different conditions in which targets could appear in the experiment. Ten of those conditions are shown in Table 5. For these 10 conditions, the previous target was as indicated in Table 5 and the prime for the previous target was the neutral prime. For the other six conditions, the previous target was not controlled (it could be either from the same or a different sentence, or it could be a word that did not appear in any studied sentence). These six conditions were made up of the following three conditions at each of the two SOAs: neutral prime and studied target, neutral prime and nonstudied target, and studied prime and nonstudied target.

Sentences were assigned to conditions randomly, and order of presentation of study and test materials was random (different for each second subject) except that the 10 conditions in Table 5 were not assigned to the first position in a test list.

Over all trials of the experiment, the probabilities of the different kinds of trials were as follows: prime and target from same studied sentence, 2/13; prime and target from different studied sentences, 2/13; neutral prime and target from studied sentence, 4/13; prime from studied sentence and target not from any studied sentence, 5/26; neutral prime and target not from any studied sentence, 5/26.

Results

Means were calculated for each subject in each condition, and means of these means are shown in Table 5. For the conditions not shown in Table 5 (with the previous target not controlled): at the 50-ms SOA, neutral prime and target from a studied sentence, 632 ms and 7% errors (1,075 observations); neutral prime and word not from a studied sentence, 717 ms and 11% errors (796 observations); word from a studied sentence prime and target not from a studied sentence, 727 ms and 14% errors (777 observations). At the 300-ms SOA, neutral prime and target from a studied sentence, 616 ms and 7% errors (1,110 observations); neutral prime and word not from a studied sentence, 713 ms and 14% errors (800 observations); word from a studied sentence prime and word not from a studied sentence target, 725 ms and 14% errors (758 observations).

According to the retrieval model, priming at the 50-ms SOA should depend not on the prime for the target but on the previous target. This is what the data show, faster response times when the previous target was from the same sentence, slower response times when it was not. On the other hand, at the 300-ms SOA, priming depends on both the prime and the previous target. If the prime is a word from a different sentence, then the previous target will be bumped from the compound cue and there will be no priming. If the prime is neutral, then the previous target will still be part of the compound cue, and there will still be priming. Again, this is what the data show.

The means were analyzed by analysis of variance—two SOA conditions crossed with the seven conditions at each SOA, as shown in Table 5. For reaction times, differences between conditions were significant, $F(6, 156) = 19.9$, and the interaction between SOA and conditions was significant, $F(6, 156) = 7.1$. The MS_e was 7.2 ms. For error rates, conditions were significantly different, $F(6, 156) = 4.00$, and the interaction was not, $F(6, 156) = 1.8$. The MS_e was 0.02. In neither analysis was the difference in SOA significant ($Fs < 1.0$).

Dunn's test was used to evaluate predictions of the retrieval model. At the 50-ms SOA, conditions 3, 4, and 5 were predicted to be slower than conditions 1 and 2. At the 300-ms SOA, condition 1 was predicted to be faster than condition 2, and condition 3 was predicted to be faster than conditions 4 and 5. Also, condition 6 was predicted to be faster than condition 7. In these latter two conditions, the previous target was not controlled, so they should replicate the priming effect usually found in experiments of this type. By Dunn's test, these four predicted differences have to be larger than 26.7 to be significant at the 0.05 level. All four are larger, and all the other differences are smaller. Thus, the pattern of data fits the prediction of the retrieval model exactly.

The comparison of importance for the spreading activation model is between the condition where the previous target is from the same sentence and the condition where it is from a different sentence, at the 300-ms SOA. When the previous target is from the same sentence, it should activate the current target, leading to priming, no matter what the intervening prime (as is shown in the 50-ms conditions). But there is no priming; when the intervening prime is from a different sentence, response times when the previous target is from the same sentence are almost identical to response times when the previous target is from a different sentence.

Conclusion

In Experiment 2, the predictions of the retrieval model were completely satisfied. When the contents of short-term memory included two nouns from the same sentence, priming was obtained; when the two nouns were from different sentences, priming was not obtained. Earlier versions of spreading activation are not consistent with these results (e.g., Anderson, 1976; Collins & Loftus, 1975; Collins & Quillian, 1969). They would have to add an inhibition process in which primes suppress activation in unrelated words, but such a suppression process would be counter to the passive, unlimited capacity nature of the models. However, Anderson's ACT* (1983) has been modified to provide rapid changes in activation as a function of time and so should be considered separately. It turns out that with relatively few additional assumptions, the compound cue explanation of priming can be incorporated into ACT*.

In ACT*, an external source causes activation of a node in memory. If the source is removed, then the node will remain active for an additional period of time, from 400 ms (Swinney, 1979) to 4 s (Meyer, Schvaneveldt, & Ruddy, 1975), and then activation will decay rapidly (i.e., in tens of milliseconds). This rapid decay can be postponed if activation is maintained by an internal source, but only so long as the internal source continues to provide activation. It is also possible for a special goal element to maintain activation for one internal source so that the system can maintain focus on a current goal of computation.

To apply ACT* to the results of Experiment 2, the two conditions of importance are the conditions in which the target and prior target are from the same sentence and the prime is presented for 300 ms. When the prime is neutral, the goal element could be assumed to maintain attention on the prior target, keeping it active so that it would facilitate the response on the next target. When the prime is a word from another sentence, it would be assumed that there was no source to maintain activation of the prior target, and so it would remain active for only 400 ms (less than the time interval between targets). The critical assumptions are that the neutral prime does not require that activation of the prior target node be turned off, but the unrelated word prime does require it to be turned off. Because activation decays so quickly within this model, it is only when activation of the prior target node is maintained by an internal source and activation of the current target node is maintained by external sources that priming is obtained. This is quite similar to our notion that the prime and target have to be part of a compound cue. (However, it should still be noted that ACT* predicts distance effects in priming that are not obtained experimentally; c.f. Experiment 1.)

Our Experiment 2 is similar to experiments performed by Meyer and his colleagues (Meyer, Schvaneveldt, & Ruddy, 1972; Schvaneveldt & Meyer, 1973) in which subjects made lexical decisions for triples of letter strings. In one experiment (Schvaneveldt & Meyer, 1973) in which subjects were required to decide whether three letter strings were all words, associations between the first and third words of a triple facilitated responses just as much as associations between the second and third words. This is counter to the results of Experiment 2, where an interposed prime eliminated facilitation (see also Davelaar & Coltheart, 1975). But in another experiment (Meyer et al., 1972), associations between the first and third words produced facilitation that was about half that from adjacent words when a word intervened between the first and third words, while there was no priming when a nonword intervened. Also, Masson (private communication, November 1986) and O'Seaghdha (1986) have obtained results consistent with Experiment 2; Masson used a rapidly paced, successive lexical decision task (similar to that of Meyer et al., 1972), and O'Seaghdha interposed syntactically incorrect function words between primes and lexical decision targets. As in Experiment 2, interposing items between prime and target eliminated facilitation.

One way to view these conflicting results among the data of Experiment 2, Meyer et al.'s different experiments, Masson's experiment, and O'Seaghdha's experiment is to consider the kinds of strategies subjects could use. In the Meyer et al. experiments, the test items were grouped as triples, so that subjects could easily notice the associative relations and take advantage of them to include more than two items in the compound cue. In contrast, in Experiment 2 and in rapidly paced lexical decision, the subjects may not focus on relations between the current target and the previous target. From this point of view, all the results could be consistent with the retrieval theory.

Schvaneveldt and Meyer (1973) suggested several models for the lexical decision task, and one of them, location shifting, makes similar predictions to the retrieval model. According to the location shifting idea, memory locations are searched serially, and time to shift from one location to another increases as a function of distance between locations. This model makes similar prediction to the retrieval model when the compound cue consists of only two items (the prime and target). The difference between the location shifting model and the retrieval model is that the location shifting model accounts for priming effects by an active processor moving through memory while the compound cue account is in terms of the relation of the compound retrieval cue to the memory representation. (Also, the models make different predictions about other effects such as range of priming.)

Introduction to Comparisons Among Models

In this section, we consider several different classes of models and describe how they relate to our view of priming. Some of these models might appear at first thought to be incompatible with the retrieval theory. Either their representational assumptions or their processing assumptions or both are quite different to those proposed in the retrieval theory. However, we show that the essence of the retrieval theory—the formation of a compound cue through the interaction of prime and target—can be reformulated in some of these models.

Recent Memory Models

We are fortunate in memory research at present to have a group of interesting models of memory phenomena. Most of these models have some commitment to parallel processing. Some of them assume that information about items is kept separate in memory, and then at retrieval, for recognition, contributions to the strength of response from all items are combined (Gillund & Shiffrin, 1984; Hintzman, 1986a). Other models assume that information about items is combined at input (Anderson, 1973; Eich, 1982, 1985; Murdock, 1982, 1983, 1985; Pike, 1984), so that information about studied items is distributed across a common memory system and there is no representation of the individual item. These models, except Gillund and Shiffrin's (1984), are vector or matrix models; that is, they assume the representation of an item is a set of (sometimes binary) features.

In this section, we describe how to implement our compound cue account of priming in these models. The implementations may not produce models that are identical to the Gillund and Shiffrin (1984) implementation presented above; for example, the Gillund and Shiffrin model predicts two steps for the range of priming, while the Murdock (1982) and Hintzman (1986a) implementations both predict one-step priming. However, our aim is to show that similar approaches can be developed within the frameworks of these models. The specific models to be discussed are those of Murdock (1982) and Hintzman (1986a); both contrast with the Gillund and Shiffrin (one node to a concept) class of models in that information about an individual item is distributed rather than localized. It is important to note that while spreading activation can be implemented in one node to a concept models, it would be difficult to implement in the framework of distributed models, especially when items are combined into a common memory vector or matrix at encoding (there are no separate entities for activation to spread from and to).

One of the features of current distributed models that is an advance over earlier distributed models is the recognition that a nonlinear component in either learning, storage, or retrieval is required. The most obvious example of the need for a nonlinear component is in Murdock's (1982) recognition model (the same as that of Anderson, 1973). In that model, longer presentation time for an item or multiple presentations of the item produce no better discriminability between old and new items in recognition than conditions with shorter presentation time or single presentations. Although the signal strength increases, the variance of the noise distribution also increases so as to keep the signal to noise ratio constant (equal performance). Different models add nonlinearity in different ways. Murdock (1986) has examined a scheme in which probabilistic encoding of features is used, but as yet the results are not conclusive. In Hintzman's model (1986a), the response of each item is cubed before summing over items, and this introduces the kind of nonlinearity needed for the system to increase recognition discriminability. Nonlinearity in the Gillund and Shiffrin model arises from the multiplication of cue strengths to provide focusing of search (simple addition would produce no advantage for cues that are related in memory over cues that are not related). This nonlinearity is essential for our view of priming because it results in items that are associatively related in memory having larger values of retrieval strength than items that are not associated.

Hintzman (1986a). Both the model of Gillund and Shiffrin (1984) and that of Hintzman (1986a) assume that each item as well as each presentation of an item is stored separately in memory and that information is combined across items at retrieval to produce a unidimensional variable (familiarity or echo intensity, respectively) upon which a yes/no recognition decision is made. The theories differ in that Hintzman assumes a vector of features for the representation of a concept, whereas Gillund and Shiffrin assume the concept itself is stored. However, despite this difference, implementation of the compound cue scheme to account for priming phenomena is straightforward in Hintzman's model.

The model assumes that paired associates are stored within a single vector, the first member of the pair in the first part of the vector and the second member of the pair in the second part of the vector. At retrieval, the test vector is correlated with each memory vector and this value (called similarity) is raised to the third power (this is called activation) and summed over all memory items to give the echo intensity (the analog of familiarity in Gillund and Shiffrin). For a prime–target pair, if the prime and target are contained in the same vector in memory, then similarity will be high and raising it to the third power will give an even higher value of echo intensity. If the prime and target are not contained in the same vector in memory, then similarity and echo intensity will both be much lower.

Table 6 shows a sample calculation for echo intensity in Hintzman's (1986a) model. Two cases are considered: First, when an intact pair is used as the test probe, the similarity of the probe A–B to A–B in memory is given by the number of nonzero elements (10), and the similarity of A–B to C–D is the

Table 6
Sample Calculation of a Compound Cue Implementation in Hintzman's (1986) Model

Assume two vectors, one for the pair A–B and the other for the pair C–D:

$$\text{A–B} = (1, 1, -1, -1, 0, 1, \hat{}\,, -1, -1, 0, 1, 1, 1)$$

$$\text{C–D} = (-1, 0, -1, -1, 1, 1, \hat{}\,, 0, -1, 1, 1, -1, 1),$$

where the symbol "ˆ" shows the break between the A and B members of the pair (and is used here only for clarity).
For the probe A–B when A–B and C–D were stored,
Activation Ac $= (10/10)^3 + ((6-2)/12)^3) = 1.037$.
For the probe A–D when A–B and C–D were stored.
Activation Ac $= ((8-1)/11)^3 + ((8-1)/11)^3 = 0.515$.

number of elements that match minus the number of elements that mismatch (4). Each of these is divided by the number of nonzero elements and the result cubed to give the activation value. These values are then summed to give echo intensity. The second case is when a test probe contains members of two different pairs (A–D). In this case, the similarity of the test probe to each memory vector is the same (7). When these similarities are cubed and summed, the result is a lower value of echo intensity than for the probe A–B. Because similarities are cubed, the A–B probe, with one larger and one smaller value of similarity (10 and 4), provides a larger value of echo intensity than does the A–D probe with two intermediate values (7). Thus the nonlinear transformation leads to larger values of echo intensity when the members of the test probe are connected in memory, and so produces priming effects.

Murdock (1982). This model assumes that the information about a single item is distributed across memory locations. For a paired associate A–B, it is assumed that the memory trace contains the individual vectors A and B and the convolution A*B. For cued recall, the correlation operation is used to retrieve one of the members of the pair given the other member as a cue.

To implement the compound cue scheme, it is assumed that at encoding, the prime and target, A–B, are encoded as the sum of A, B, and A*B. At retrieval, the test probe will be a combination of the target (B) and the convolution of the prime and target (A*B). The match between the convolution of the prime and target and the memory vector will be greater if the prime and target were associatively related in memory (i.e., the convolution of A*B is stored as the result of previous encoding). This same scheme can be implemented in the matrix model of Pike (1984; see also Eich, 1982). Thus, these models can account for the increased strength of related primes and targets by use of an additional associative component.

To conclude, the memory models with distributed representations can account for priming effects with a compound cue scheme. The prime and target are combined into a compound cue and this cue provides a greater value of familiarity (or strength or similarity) if the prime and target are associated in memory. It is difficult to imagine how a spreading activation process could be implemented in these vector models because there is no network of concepts among which activation can spread.

Relation of the Retrieval Theory to Connectionist Models

Our aim in this section is to describe how the retrieval theory of priming can be incorporated into connectionist frameworks, particularly with respect to assumptions about memory representation and prime–target combination. The vector memory models described above can be considered part of the general class of parallel distributed models that includes connectionist models, but we have separated the memory models and the connectionist models for reasons of organization. It is important to discuss the relation between these connectionist models and the retrieval theory because the retrieval theory denies the existence of the spreading activation process, whereas the connectionist models have spreading activation at the heart of their processing mechanisms. We shall attempt to show how the retrieval theory, based on both empirical and theoretical psychological considerations, can map into connectionist models that use parallel processing mechanisms.

There are several flavors of connectionist models organized along two basic dimensions. The first is local versus distributed: Some connectionist models assume that one node represents one concept (Dell, 1986; Feldman, 1981; McClelland & Rumelhart, 1981), while others suppose that a concept is distributed across a group of nodes (Ackley, Hinton, & Sejnowski, 1985; Anderson, Silverstein, Ritz, & Jones, 1977; Grossberg & Stone, 1986). Both local and global models assume a spreading activation process. The second dimension concerns the connectivity and organization of the network. The three main classes of models are single layer (e.g., the vector and matrix models, Eich, 1982; Hintzman, 1986a, 1986b; Murdock, 1982; Pike, 1984), single layer with autoassociative connections (Anderson et al., 1977; Hinton, 1981; Knapp & Anderson, 1984; Kohonen, 1978), and multilayer with connections from layer to layer (Ackley et al., 1985; Dell, 1986; Feldman, 1981; Grossberg & Stone, 1986; McClelland & Rumelhart, 1981). Most of the distributed memory models assume a single layer of nodes or features, whereas most models of perception or production assume multiple layers or autoassociative connections. We considered single layer models in the previous section; in this section the focus is on the multilayer connectionist models.

In a typical multilayer connectionist network, there are three layers of nodes. One represents the input, another the output, and the third an intermediate layer (connected to the input and output layers) called the *hidden unit layer.* When an item is presented for learning, activation spreads from the nodes activated at the input layer, through the hidden units, to the output layer, according to the weights (strengths) on the connections between the nodes at different layers. The pattern of activation at the output layer is compared with the correct pattern, and then the weights are adjusted (by back propagation or some other algorithm) to more nearly correspond to the correct pattern. In this manner, the system will learn to associate a given input to the desired pattern of activations in the output layer. Spreading activation is essential to this process as a mechanism by which activation is transferred from input nodes through the hidden nodes to the output nodes.

As noted in the introduction, superficially there appears to be a conflict between connectionist models and the compound cue account of priming because we argue that spreading activation is not necessary to explain priming phenomena (the major class of empirical phenomena used to support activation based theories). However, as in the case of the single layer models, it is possible to develop an account of priming that uses a compound cue. The activation process used to learn patterns need not be the process responsible for priming effects, that is, priming effects need not be explained by assuming that a priming item leaves the network in an altered state and that processing of the target is aided by this altered level of activation. Instead, priming can be viewed as resulting from a cue at retrieval that is a compound of the prime and target (implemented in Hinton's, 1981, model, for example). However, evaluation of this claim can only come when a comprehensive connectionist model of recognition or lexical decision is developed and the two submodels for priming, activation of long-term memory versus combined cue interactions, can be compared. A critical factor in this evaluation will be the rate of decay of activation as a function of time. If decay of activation of nodes or collections of nodes representing concepts is relatively slow, then priming could be explained in the spreading activation sense, but if decay is quite fast, as in Anderson's ACT* model, then the combined cue model will provide a more natural account of priming.

Grossberg and Stone (1986). Grossberg and Stone (1986) presented a model that shares the approach of the connectionist theorists. A signal causes activation at an input level, and the activation is transmitted up to the next level of nodes in the network. The activation at the input layer is multiplicatively gated by long-term memory traces to form the activation pattern for the next level. Multiplicative gating means that the signals are multiplied so that large input signals with no corresponding long-term memory trace and small input signals with large long-term memory traces are suppressed, and only input signals with long-term memory traces are passed on. In the same way, the higher level feeds back activation by gating top down to the lower layer. Thus, after some time, the input layer is tuned to the combination of the long-term memory traces and the input signal. Learning of new patterns is accomplished by an attentional gain control mechanism that allows short-term memory to stabilize and modify the network. Grossberg and colleagues (1980, 1984, 1986; Carpenter & Grossberg, 1987) have critiqued the notion of spreading activation usually presented in semantic network models (i.e., slow node-to-node serial spreading). The design principles he and his associates have developed center on issues formulated within the framework of a parallel neural network system. While the principles have been well worked out (e.g., proofs that the systems are stable, the systems can learn in the presence of time varying inputs, etc), in many cases, they have not been fitted directly to data, and additional mechanisms to produce quantitative predictions have not been specified in enough detail to allow explicit fitting. When such applications to data are possible, they should further theoretical development.

Grossberg and Stone (1986) presented an explanation of priming phenomena that at a cursory glance seems similar in spirit to a spreading activation account. The explanation is that a prime item sets the input level to receive possible target inputs, by feedback from the higher layer. However, the time course of processing is very rapid and so the approach is related more closely to our compound cue theory. It would also be possible to directly implement the compound cue mechanism proposed in this article within their framework. Patterns of activation from the prime and target would sum at the input layer to form a compound activation pattern and the activation would be transmitted up to the next layer. Long-term memory traces would gate this signal so that the more connected the two items in long-term memory (i.e., the more the two patterns of activity overlap), the larger the compound signal at the upper layer. This would lead to priming effects. Until the details of these two approaches have been worked out, we feel they both remain as possible ways of implementing priming phenomena within the Grossberg and Stone framework. Clearly, at a qualitative level, the approach provided by Grossberg and Stone (1986) is closely related to our own and comparisons between the possible different implementations of the priming schemes should lead to theoretical advances.

Is Priming a Unitary Process?

The main focus of this article is to use the retrieval theory to account for priming effects. However, up to this point, the only priming effects that have been considered are those in which priming is from one item to another and decays quickly, within a second or two or within one to two interpolated items. In this section, we address those priming phenomena that have much longer decay times—minutes, hours, or days—including priming in perceptual identification and word fragment completion (Jacoby, 1983a, 1983b; Jacoby & Dallas, 1981; Tulving, Schacter, & Stark, 1982), priming effects in stem completion and free association (Graf, Shimamura, & Squire, 1985; Shimamura, 1986), repetition effects in lexical decision (Monsell, 1983; Scarborough, Cortese, & Scarborough, 1977), and facilitation in cognitive skills (Cohen, 1984; Logan, in press). Short-term priming effects, those discussed in the first sections of this paper with reference to the retrieval theory, have been most closely associated with the concept of spreading activation (Anderson, 1976, 1983; Meyer & Schvaneveldt, 1971) because the processes proposed in spreading activation models require that activation decay rapidly once one process is completed and a new one begins. Long-term priming effects have been argued to represent a different memory system (Tulving, 1983; but see McKoon, Ratcliff, & Dell, 1986) or different, possibly procedural, processes (Cohen, 1984; Jacoby & Witherspoon, 1982).

The classification of priming phenomena according to decay rates requires extreme caution. All long-term effects may not be the result of the same mechanism, and any one effect may reflect more than one mechanism. For example, Shimamura (1986) has noted that some effects seem to decay within hours (stem completion) while others last for days (perceptual identification and fragment completion). Another example of this complication comes from a proposal by Cohen (1984) that long-term priming effects in lexical decision and fragment completion are mediated by the same processing system as cognitive and motor skills. The basis of this proposal was the finding that these three sets of phenomena are long lasting in normal populations and preserved in amnesics. However, Nissen, Knopman, and Schacter (in press) have presented results that suggest that such disparate tasks *cannot* be lumped together purely on the basis of decay rates. Nissen et al. compared the long-term effects of priming in a cognitive skills task (manual responses to digits) and in word fragment completion when subjects received the drug scopolomine (which induces amnesia) in the retention period. After half an hour, the priming effect in fragment completion was eliminated while priming in the skill task was preserved. Thus, although both of these tasks usually show long-term effects, administration of scopolomine produced a dissociation, and an indication that skills and long-range priming effects may not be mediated by the same processing system. Within a theoretically motivated framework, such dissociations may be able to be accommodated within a single processing system that incorporates principled reasons for predicting such differences in decay rates.

Another benchmark for separating memory systems is the dependence or independence of a subject's performance on different tasks. For example, Tulving et al. (1982) found that performance on recognition and priming effects in word fragment completion were independent, and Jacoby and Witherspoon (1982) found independence of recognition and priming in perceptual identification for words but not pseudowords. From these and similar results, Tulving (1983) argued that priming effects are mediated by the operations of a system other than "episodic memory." However, the use of statistical independence as a benchmark for separating memory systems is problematic. Witherspoon and Moscovitch (1987) have shown that performance on word fragment completion and perceptual identification is statistically independent, but one would not want to argue that these were mediated by separate memory systems (see also Jacoby & Witherspoon, 1982, and Mandler, Graf, & Kraft, 1986, for similar arguments). The more reasonable way of viewing these results is to suggest that the degree of statistical dependence reflects the degree of use of common processes or information (Jacoby & Witherspoon, 1982).

We have proposed a speculative account of long-term facilitation effects within the framework of a model that assumes that the priming effects are based on changes in criteria. The model was developed to account for data obtained by Ratcliff, McKoon, and Verwoerd (in press) from a task showing facilitation in perceptual identification. Subjects were asked to identify test words presented tachistoscopically for brief exposure durations (about 10 ms, followed by a mask). Some of the test words had been presented previously in sentences the subjects were asked to study. If subjects had seen a word before in the experiment, then their ability to identify the word was improved, but they were also likely to misreport visually similar words. For example, if they had seen "died" earlier, they were more likely to report "died" when it was tested *and* more likely to misreport "lied" as "died." When given a forced choice between "died" and "lied," there was no improvement in the average proportion correct or in d' as a function of having seen one of the choices earlier. These results indicate that the term "perceptual fluency" is a misnomer for the processing advantage obtained from a repetition. Rather, we should view the processing advantage as a bias toward earlier processed items (Jacoby, 1983a, p. 36). It is important to note that the concept of bias used here is not a simple response bias generated by conscious strategies on the part of the subjects. Instead, it should be viewed as a perceptual bias that is the result of alterations to criteria used in the processing system (so in Jacoby and Witherspoon's terms, perceptual bias leads to the impression that an item "jumps out" from the display screen).

This bias effect can be explained by a simple criterion model. When stimulus information is presented, one or more candidates are selected for evaluation as possible matches to the stimulus. This selection process can be viewed as an interactive parallel process where by the quality of the stimulus item determines the order of selection. If a particular candidate has been processed recently (i.e., earlier in the experiment), its criterion for selection will be lowered and then that candidate will be selected with a greater probability. (The selection of a candidate will also be affected by other factors, such as mode of presentation [Morton, 1979; see Jacoby, 1983b].) In perceptual identification, a recently processed candidate will be more likely to be reported than other candidates (either correctly or as an intrusion error), and in forced choice, it will be more likely to be chosen over the other alternative if the information that discriminates the two choices has not been adequately processed. This criterion model can be extended to other tasks such as word fragment completion (Tulving et al., 1982) and the perceptual identification task of Feustel, Shiffrin, and Salasoo (1983). In both these cases, long-term advantages would result from previously presented items being available earlier in processing (or available per se) as candidate solutions or responses. Also the model allows (in a post hoc way) for independence of performance on tasks such as fragment completion and perceptual identification by assuming that the two tasks share relatively few processes or relatively little of the information that gives rise to long-term priming effects.

The model just outlined is close to Morton's logogen model (Morton, 1969, 1970, 1979). The logogen system is a device for converting sensory information into semantic (or phonological) information. It is not designed to be a semantic net, and associative effects would be expected to be mediated by a separate semantic system. Presentation of a word lowers a criterion for recognition of the word, and this criterion rises slowly back to its original value. Morton (1970) argued that this gradual decay is necessary for the system to remain stable. The bias model we propose for long-term facilitation is closely related in spirit to the logogen model by the sharing of ideas such as parallel access, slowly decaying thresholds, and possibly multiple levels of representation (see also McClelland & Rumelhart's, 1981, discussion of the relationship of the logogen model to their model). Although Jacoby and Witherspoon (1982) have criticized the strict version of the logogen model, a model like we have described that allows for context effects is in line with Jacoby and Witherspoon's theoretical account of "perceptual fluency" and is consistent with the views described by Jacoby (1983b, pp. 501–502).

These results do not mean, however, that all the implicit tasks that have been studied in this context show priming effects resulting from perceptual bias. For example, Graf et al. (1985) have shown that cross-modal repetition priming can be obtained in word completion (in contrast to little or no cross-modal priming in perceptual identification; Jacoby & Witherspoon, 1982). In addition, they found long-term semantic priming effects in free association that were spared (and lasted for minutes) under amnesia. These results (see also the review of other findings in Shimamura, 1986) demonstrate that both semantic and perceptual representations or processes can be primed, but the nature of the priming is specific to the information involved in the task. If the task is perceptual identification, perceptual information is involved, whereas if the task is free association, semantic information can be involved. We argue that our bias interpretation can serve as a useful preliminary hypothesis to explain the mechanisms underlying this whole class of priming effects. For example, in free association, some of the component processes use semantic information and earlier presentation of a word will serve to produce bias in these representations or processes.

It is also possible to relate our criterion model to the two-phase model of Graf and Mandler (1984). They account for the separation of performance on fragment completion tasks on the one hand, and recognition and recall on the other hand, in terms of activation and elaboration processes. The activation process can be identified with what we have labeled as a perceptual bias mechanism. Graf and Mandler (1984) argued that if a cue is salient, it will activate the schema that represents the word and the word will simply "come to mind." If we reinterpret this as the lowering of a threshold, then the two accounts are quite similar. (The problem with a strict activation account of perceptual identification, for example, is that it would predict that overall performance would improve, in terms of d', whereas the criterion explanation requires no such improvement.)

The data provided by Ratcliff et al. (in press) and the simple criterion model give useful constraints for understanding differences between tasks. On the one hand, tasks like perceptual identification and word fragment completion require the subject to produce alternative candidates to match to impoverished stimuli. Prior presentation of a word biases the candidate production process toward earlier evaluation of that word, and this bias is long lasting. On the other hand, tasks like recognition require specific knowledge of previous presentation; this knowledge is reflected not in bias but in d', and decay is relatively rapid. Although this distinction provides some insights, it is important to realize that it does not explain why the bias effect is long term in these tasks (minutes, hours, or days) and why decay of the bias effect is different for the different tasks. We interpret differences in decay rates as reflecting different but possibly overlapping kinds of information for different tasks. For example, in perceptual identification, letters (and features) are impoverished at test so that partial information about features and letters is used to match candidates for words, while in lexical decision, the whole word is available and can be used to match lexical representations. These different kinds of information could contact different pieces of information in the lexicon, leading to different behaviors in the tasks.

The discussion in this section makes it clear that every phenomenon that has been labeled as "priming" *cannot* be classified within the framework of a unitary process theory. Although we have distinguished between two kinds of priming, one to be explained as the interaction of a compound cue and the other as a change in perceptual bias, there may be still other kinds of priming (cf. Wyer & Srull, 1986). It is only within the framework of specific models that it is possible to evaluate claims about the similarities or differences between priming effects in different paradigms. In addition, it must be kept in mind that more than one kind of priming may be operating in any experiment, so that relations between different kinds of priming must be addressed. For example, Monsell (1983) and Ratcliff et al. (1985) showed that the repetition effect in lexical decision (between two instances of the same word) involves both a short-term component of the kind found in priming between different words (in both lexical decision and recognition) and a long-term component that lasts considerably longer than the short-term effect. Also, Whitlow (1986) has shown that there is a component of priming in a semantic priming paradigm related to physical identity. In conclusion, we explain priming associated with rapid decay by the combined cue model and priming associated with long-term effects by the criterion model, but argue that the relations among different kinds of priming in any experimental situation are complex.

Summary and Conclusions

In this article, we have presented a retrieval theory that is designed to account for associative or semantic priming phenomena, priming phenomena that have a short range and rapid decay function. The theory assumes that the prime and target form a compound cue and that this compound interacts with memory to produce a value of resonance, goodness of match, or familiarity that is determined by associations in long-term memory between the prime and target. If the prime and target are directly associated in memory (or mediated by only a single other item in the Gillund & Shiffrin, 1984, formulation), then the familiarity value will be larger than if they are not associated. We assume that the value of familiarity drives a random walk or diffusion comparison process in which information is accumulated over time toward two criteria, one for positive and one for negative responses, and increased familiarity leads to a priming effect in reaction time. In this article, the retrieval theory has been discussed with respect to five main points:

1. The theory was shown to account for many of the findings that have been cited as evidence for spreading activation. These include range effects, the speed of onset of priming, priming of multiple meanings of ambiguous words, and the effects of neutral priming conditions.

2. The retrieval theory makes strong predictions about the decay rate of priming, backward priming, and the range of priming, all of which are supported by data. In contrast, accounts of these effects by spreading activation theories are less predictive.

3. Two new experiments presented in this article gave data more consistent with the compound cue view of priming than spreading activation views. In Experiment 1, the range of priming was shown to be smaller than might have been predicted from spreading activation theories, and in Experiment 2, priming was shown to depend on the contents of short-term memory (as predicted by the retrieval theory) rather than on activation changes in long-term memory.

4. It was shown that the compound cue theory could be implemented in a range of theoretical frameworks such as distributed memory models and connectionist models. The compound cue explanation of priming is not tied to any particular model and can serve as a general, model-independent explana-

tion of priming (though specific quantitative predictions may vary between implementations).

5. The compound cue model was compared with a model for long-term priming effects. The scopes of these two models were described, and the implications of classifying priming effects according to these two models were outlined.

The retrieval theory for fast-decaying priming phenomena is designed to account for data from a range of priming procedures and, at the same time, to provide a challenge to the extremely popular and influential spreading activation theory. This challenge is intended to force examination of spreading activation in the role it now serves as a widely accepted and widely applicable theory. Such an examination is long overdue, and should lead to new evidence that will increase our understanding of priming phenomena in particular and memory theories in general.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9,* 147–169.

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80,* 417–438.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R., & Pirolli, P. L. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 791–798.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 336–345.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing, 37,* 54–115.

Cohen, N. J. (1984). Preserved learning capacity in amnesia: Evidence for multiple memory systems. In N. Butters & L. Squire (Eds.), *The neuropsychology of memory* (pp. 83–103). New York: Guilford Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240–248.

de Groot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22,* 417–436.

de Groot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. W. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory and Cognition, 10,* 358–370.

Davelaar, E., & Coltheart, M. (1975). Effects of interpolated items on the association effect in lexical decision tasks. *Bulletin of the Psychonomic Society, 6,* 269–272.

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

Dell, G., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 121–132.

den Heyer, K. (1986). Manipulating attention-induced priming in a lexical decision task by means of repeated prime target presentations. *Journal of Memory and Language, 25,* 19–42.

den Heyer, K., Briand, K., & Dannenbring, G. L. (1983). Strategic factors in a lexical-decision task: Evidence for automatic and attention-driven processes. *Memory & Cognition, 11,* 374–381.

Dosher, B. (1982). Effect of sentence size and network distance on retrieval speed. *Journal of Experimental Psychology: Learning Memory, and Cognition, 8,* 173–207.

Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Eich, J. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92,* 1–38.

Falmagne, J. C. (1965). Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology, 12,* 77–124.

Feldman, J. (1981). A connectionist model of visual memory. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 49–82). Hillsdale, NJ: Erlbaum.

Feustel, T., Shiffrin, R., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112,* 309–346.

Forster, K. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology, 33,* 465–495.

Foss, D. J. (1982). A discourse on semantic priming. *Cognitive Psychology, 14,* 590–607.

Gillund, G., & Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychological Review, 19,* 1–65.

Glass, A., & Holyoak, K. (1975). Alternative conceptions of semantic memory. *Cognition, 3,* 313–339.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23,* 553–568.

Graf, P., Shimamura, A. P., & Squire, L. R. (1985). Priming across modalities and priming across category levels: Extending the domain of preserved function in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 386–396.

Gronlund, S. D. & Ratcliff, R. (1988). *The time course of item and associative information: Implications for global memory models.* Manuscript submitted for publication.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87,* 1–51.

Grossberg, S. (1984). Unitization, automaticity, temporal order, and word recognition. *Cognition and Brain Theory, 7,* 263–283.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machine, Vol. 1: Speech perception* (pp. 187–296). New York: Academic Press.

Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93,* 46–74.

Hinton, G. (1981). Implementing semantic networks in parallel hardware. In G. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 161–188). Hillsdale, NJ: Erlbaum.

Hintzman, D. (1986a). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. (1986b). Judgments of frequency and recognition memory in a multiple-trace memory model. University of Oregon Technical Report No. 86-11.

Jacoby, L. L. (1983a). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology, 9,* 21–38.

Jacoby, L. (1983b). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508.

Jacoby, L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3,* 306–340.

Jacoby, L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36,* 300–324.

Kieras, D. (1981). Component processes in the comprehension of simple prose. *Journal of Verbal Learning and Verbal Behavior, 20,* 1–23.

Kiger, J. I., & Glass, A. L. (1983). The facilitation of lexical decisions by a prime occurring after the target. *Memory and Cognition, 11,* 356–365.

Kintsch, W. (1974). *The representation of meaning in memory.* Hillsdale, NJ: Erlbaum.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. Nilsson (Ed.), *Perspectives on memory research* (pp. 329–365). Hillsdale, NJ: Erlbaum.

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 616–637.

Kohonen, T. (1978). *Associative memory: A system-theoretical approach.* Berlin: Springer-Verlag.

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

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory and Cognition, 9,* 587–598.

Levin, J. A. (1976). *Proteus: An activation framework for cognitive process models (ISI/WP-2).* Marina del Rey, CA: Information Sciences Institute.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12,* 523–553.

Logan, G. D. (In press). Toward an instance theory of automatization. *Psychological Review.*

Mandler, G., Graf, P. & Kraft, D. (1986). Activation and elaboration effects in recognition and word priming. *The Quarterly Journal of Experimental Psychology, 38,* 645–662.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 8,* 375–407.

McCloskey, M. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior, 19,* 485–502.

McKoon, G. (1981). The representation of pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 216–221.

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18,* 463–480.

McKoon, G., & Ratcliff, R. (1980a). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior, 19,* 369–386.

McKoon, G., & Ratcliff, R. (1980b). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668–682.

McKoon, G., & Ratcliff, R. (1986a). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 82–91.

McKoon, G., & Ratcliff, R. (1986b). The automatic activation of episodic information in a semantic memory task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 108–115.

McKoon, G., & Ratcliff, R. (1987). *The use of on-line lexical decision probes to investigate inference processes.* Unpublished manuscript.

McKoon, G., Ratcliff, R., & Dell, G. S. (1985). Semantic facilitation in episodic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 742–751.

McKoon, G., Ratcliff, R., & Dell, G. S. (1986). A critical evaluation of the semantic/episodic distinction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 295–306.

McNamara, T. P., Ratcliff, R., & McKoon, G. (1984). The mental representation of knowledge acquired from maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 723–732.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1972, November). *Activation of lexical memory.* Paper presented at the meeting of the Psychonomic Society, St. Louis, MO.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98–118). London: Academic Press.

Meyer, D. E., Yantis, S., Osman, A., & Smith, J. E. K. (1985). Temporal properties of human information processing: Tests of discrete versus continuous models. *Cognitive Psychology, 17,* 445–518.

Miller, J. R. (1981). Constructive processing of sentences: A simulation model of encoding and retrieval. *Journal of Verbal Learning and Verbal Behavior, 20,* 24–45.

Monsell, S. (1983). *Automaticity and persistence of semantic activation: Two indices.* Paper presented at the Experimental Psychology Meeting, St. Andrews, Scotland.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. E. Wrolstal, & H. Bouma (Eds.), *Processing visible language 1* (pp. 259–268). New York: Plenum Press.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review, 90,* 316–338.

Murdock, B. B. (1985). Theoretical note. Convolutions and matrix systems: A reply to Pike. *Psychological Review, 92,* 130–132.

Murdock, B. B. (1986, August). *Probabilistic encoding and learning in a distributed memory model.* Paper presented at the Mathematical Psychology meeting, Cambridge, MA.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity intention. *Journal of Experimental Psychology: General, 106,* 226–254.

Neely, J. H., & Durgunoglu, A. (1985). Dissociative episodic and semantic priming effects in episodic recognition and lexical decision tasks. *Journal of Memory and Language, 24,* 466–489.

Nissen, M. J., Knopman, D., & Schacter, D. L. (In press). Neurochemical dissociation of memory systems. *Neurology.*

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in short-term memory. *Journal of Mathematical Psychology, 6,* 192–208.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition, 9,* 225–236.

O'Seaghdha, P. G. (1986). The dependence of lexical relatedness effects on syntactic connectedness. Unpublished doctoral dissertation, University of Toronto.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91,* 281–294.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium.* Hillsdale, NJ: Erlbaum.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Quillian, M. R. (1969). The teachable language comprehender. *Communications of the ACM, 12,* 459–476.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88,* 552–572.

Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review, 92,* 212–225.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review, 95,* 238–255.

Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General, 114,* 435–450.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17,* 403–417.

Ratcliff, R., & McKoon, G. (1981a). Does activation really spread? *Psychological Review, 88,* 454–462.

Ratcliff, R., & McKoon, G. (1981b). Automatic and strategic priming in recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 204–215.

Ratcliff, R., & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Human Learning and, Memory, 8,* 16–36.

Ratcliff, R., & McKoon, G. (in press). Item familiarity versus relational information: The time course of retrieval of sentences. *Cognitive Psychology.*

Ratcliff, R., McKoon, G., & Verwoerd, M. J. (In press). A bias interpretation of facilitation in perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1–17.

Schustack, M. (1981). *Word-meaning comprehension: Syntactic and associative effects of sentential context.* Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh, PA.

Schvaneveldt, R., & Meyer, D. (1973). Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 395–409). New York: Academic Press.

Seidenberg, M., Tanenhaus, M., Leiman, J., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14,* 489–537.

Seidenberg, M., Waters, G., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition, 12,* 315–328.

Seifert, C., McKoon, G., Abelson, R., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 82–91.

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for dissociable memory function. *Quarterly Journal of Experimental Psychology, 38,* 619–644.

Simpson, G. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin, 96,* 316–340.

Smith, E. E., Shoben, E., & Rips, L. (1974). Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review, 81,* 214–241.

Smith, M. J. (1979). Contextual facilitation in a letter search task depends on how the prime is processed. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 239–251.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659.

Theios, J. (1973). Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. VII, pp. 43–85). New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tulving, E., Schacter, D., & Stark, H. (1982). Priming effects in word-fragment completion are independent of recognition effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8,* 336–341.

West, R. F., & Stanovitch, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory and Cognition, 14,* 95–103.

Whitlow, J. W., Jr. (1986). Nature of priming effects in semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 353–360.

Wickelgren, W. (1976). Network strength theory of storage and retrieval dynamics. *Psychological Review, 83,* 466–478.

Witherspoon, D., & Moscovitch, M. (1987). *Differential repetition effects in word fragment completion and perceptual identification.* Manuscript submitted for publication.

Wyer, R., & Srull, T. (1986). Human cognition in its social context. *Psychological Review, 93,* 322–359.

Yantis, S., & Meyer, D. (1988). Dynamics of activation in semantic and episodic memory. *Journal of Experimental Psychology: General, 117,* 130–147.

Received December 16, 1986
Revision received November 11, 1987
Accepted December 11, 1987 ■