psychological Review
1997, Vol. 104, No. 2, 319-343

0033-295X/97/$3.00

A Counter Model for Implicit Priming in Perceptual Word Identification

Roger Ratcliff and Gail McKoon
Northwestern University

A model for the identification of briefly presented words is presented. The model accounts for data from naming and forced-choice experiments in which factors such as similarity of alternatives and stimulus presentation time are varied. The model assumes that counts are accumulated in counters that correspond to words and that a word is chosen as a response when the number of counts in its counter exceeds the maximum of the numbers of counts in other counters by a criterial value. Prior exposure to a word causes its counter to attract more counts than it otherwise would, and this yields priming effects. Ten experiments are presented, and the model provides correct predictions for the data. Implications of the model for research in implicit memory are considered.

Considerable research effort in the past decade has been devoted to implicit memory phenomena, and there have been many important empirical results. The results have come to be viewed from two sharply contrasting theoretical perspectives. The dominant explanation of the data is that there exist multiple independent memory systems: one system used for conscious recollection, and the other systems supporting the unconscious retrieval of implicit information. The other view of the data is that the empirical phenomena of implicit memory and the processes of retrieval without awareness can best be understood within an information-processing framework that does not postulate multiple implicit memory systems (Broadbent, 1958; LaBerge & Samuels, 1974; Morton, 1969; Posner, 1978). From both perspectives, the agenda for progress is the same: the development of detailed accounts of the mechanisms responsible for performance on implicit tasks. In this article, we offer a model of one implicit phenomenon, priming in the perceptual identification of words, and examine how the model relates to implicit memory system views of priming (Richardson-Klavehn & Bjork, 1988; Roediger & McDermott, 1993; Schacter, 1990, 1994; Schacter & Tulving, 1994; Squire, 1994) and information-processing and connectionist views of word identification (LaBerge & Samuels, 1974; McClelland & Rumelhart, 1981; Morton, 1969; Posner, 1978; Seidenberg & McClelland, 1989).

Perceptual identification was one of the first influential implicit procedures (Broadbent, 1967; Jacoby & Dallas, 1981; Morton, 1968; Winnick & Daniel, 1970). A typical experiment consists of two phases: first, study of a list of words and second, tests of those words mixed with tests of new words that were not studied. A test of a word consists of displaying the word visually for a brief amount of time, and the task for the observer is to attempt to identify the word. The time for which the word is flashed is so brief that performance is not perfect. The identification of a word does not require conscious retrieval of that word from the study list. Nevertheless, the probability of correct identification is increased if the word did appear in the study list. An increase in probability correct due to the repetition of a stimulus is called *priming,* and it is the finding around which the multiple implicit memory systems approach has revolved.

Priming in perceptual identification provides a cornerstone for the postulation of a visual word-form system of memory (Schacter, 1990, 1994), one of several proposed implicit perceptual-representation systems. The perceptual-representation systems are one of four proposed classes of independent long-term, memory systems: The others are procedural memory, semantic memory, and episodic memory. The visual word-form system is concerned with the perceptual identification of words, and it operates on a presemantic level of information. Like other perceptual-representation systems, it stores information in an implicit and not consciously accessible form (Schacter, 1990, 1994; Schacter & Tulving, 1994). It is this form of memory that supports priming, and the function of priming is to improve the perceptual identification of words (Tulving & Schacter, 1990). Placing "perceptual" identification in the context of a memory system emphasizes that "what is perceived is as much an expression of memory as it is of perception" (p. 302).

The primary goal of the research described in this article was to develop a model that would explain performance in perceptual identification and include an explanation of how priming comes about. The aim was to explain the mechanisms of identification and priming in sufficient detail to allow quantitative as well as qualitative predictions about performance. We anticipated that these mechanisms would not require the kind of multiple memory systems proposed by Schacter and Tulving (1994), and therefore a successful model would allow examination of how implicit phenomena can be understood from the standard infor-

---

Roger Ratcliff and Gail McKoon, Psychology Department, Northwestern University.

This research was supported by National Institute of Mental Health Grants HD MH44640 and MH00871, National Institute for Deafness and Other Communication Disorders Grant R01-DC01240, and National Science Foundation Grant SBR-9221940.

We would like to thank John Gabrieli for suggesting Experiment 3, Mark Seidenberg and Jay McClelland for making available their models, and David Thau for working on the Seidenberg and McClelland (1989) model. We also thank David Allbritton, Douglas Hintzman, and Bennet Murdock for their comments on this article.

Correspondence concerning this article should be addressed to Roger Ratcliff, Psychology Department, Northwestern University, Evanston, Illinois 60208-5190.

mation-processing tradition (e.g., LaBerge & Samuels, 1974; Morton, 1970; Posner, 1978, chap. 4).

The information-processing approach emphasizes the time course of the flow of information through cognitive processes and the transformations the information undergoes. Posner (1978) justified the approach by summarizing a great deal of research that demonstrates the evolution of different codes of information at different points through processing time. Generally, information-processing models emphasize what happens to a stimulus over time such that the culmination is a semantic representation: an "understanding" of the stimulus. For example, LaBerge and Samuels' (1974) model outlines the development over processing time of different codes in the course of understanding words in reading. Figure 1 shows a schematic version of their model; Figure 1 also has similarities to other models (e.g., McClelland & Rumelhart, 1981). In this framework, priming would be explained as a modification to some early stage of processing. The key is to develop a model of the mechanisms that produce performance in a task and then account for priming as modifications of those mechanisms.

Within the general framework described in Figure 1, there are several theories that attempt to describe how words are identified. One of the earliest was the *logogen model* proposed by Morton (1969, 1970, 1979; Murrell & Morton, 1974). In this model, words are represented in memory as counters or *logogens.* Incoming stimulus information for a word is accumulated in the counters, and identification of the stimulus occurs when the amount of evidence in a word's counter exceeds a threshold. Priming effects are explained as the temporary lowering of thresholds. Morton (1970, p. 216) views priming as simply a by-product of the main function of the logogen system, which is word identification.

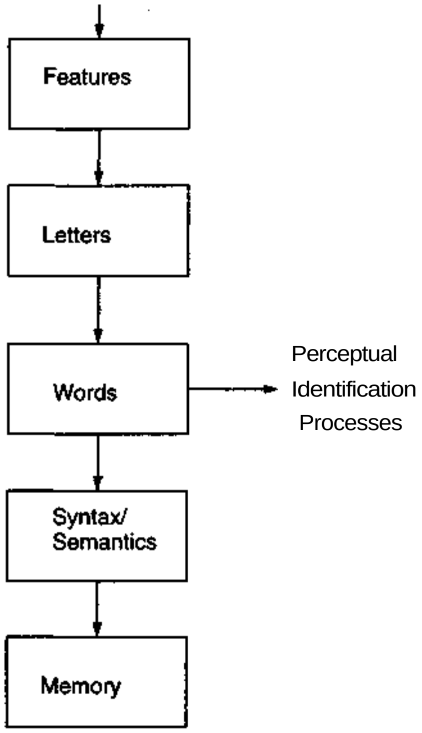

Figure 1. A traditional information-processing view of the stages or processes involved in word identification and memory. Reprinted from "Bias Effects in Implicit Memory Tasks," by R. Ratcliff and G. McKoon, 1996, *Journal of Experimental Psychology: General, 125,* p. 405. Copyright 1996 by the American Psychological Association.

More recently, connectionist models of word identification have been proposed by McClelland and Rumelhart (1981) and by Seidenberg and McClelland (1989). In McClelland and Rumelhart's model, incoming stimulus information sends activation spreading from feature nodes through letter nodes to converge on word nodes, and the stimulus is identified as the word with the largest relative amount of activation. This model can be seen as a generalization of the logogen model (McClelland & Rumelhart, 1981, p. 388), and might handle priming in a similar way, through the lowering of resting activation levels on word nodes. In Seidenberg and McClelland's model, words are represented in a distributed network of nodes. A stimulus causes activation to spread from a layer of nodes representing orthographic information to a layer of nodes representing phonemic information. Through repeated learning trials on which the orthographic features of words are presented to the orthographic layer, the model is trained to output the words' correct phonemic representations. For priming, there are several ways the influence of a single presentation might be implemented.

The logogen and connectionist models focus on the processes of word identification. Jacoby (Jacoby, 1983a, 1983b; Jacoby & Dallas, 1981; Jacoby & Witherspoon, 1982) has been more directly concerned with priming and how implicit priming effects relate to other forms of memory. Jacoby's proposal is that identification of a word can be facilitated by a single episode of prior perceptual processing of the word. The amount of facilitation depends on the amount of perceptual processing in the prior episode and the amount of overlap in perceptual features between the prior episode and the test episode.

The logogen model, the connectionist models, and Jacoby's (Jacoby, 1983a, 1983b; Jacoby & Dallas, 1981; Jacoby & Witherspoon, 1982) hypotheses have all been formulated in the context of the kind of information-processing view summarized by Figure 1. This tradition contrasts sharply in its goals with the implicit memory systems approach. By attributing implicit priming effects to representations constructed in multiple separate and independent memory systems, the memory systems approach emphasizes the laying down of traces in memory. By labeling the systems *memory* systems, it is implied that their primary function is memory. The consequence is research aimed at such issues as dissociations among the multiple systems and the locations in the brain of the systems. The theoretical constraints on hypothetical mechanisms for priming are not strong because the systems are only loosely linked to the computations being performed in an experimental task.

The alternative view we endorse, consistent with the older information-processing views, places the emphasis on developing models of processing that account for performance on a specific implicit task (see Ratcliff & McKoon, 1996). The effects of prior study that produce priming are modeled as residual effects that result from the computations involved in performing the task. As Morton (1970) put it, "The only part of the model which has memory as part of its function is the Cognitive System. The other parts of the model appear to store information for varying lengths of time and in varying amounts rather as a by-product of their main functions" (p. 216).

Morton's logogen model, the connectionist models, and Jacoby's (Jacoby, 1983a, 1983b; Jacoby & Dallas, 1981; Jacoby & Witherspoon, 1982) hypotheses all offer at least the possibility of explaining the processes involved in perceptual identification and priming in perceptual identification. They also have the potential to explain a second phenomenon, and that is bias: Prior processing has costs as well as benefits. In an earlier study (Ratcliff, McKoon, & Verwoerd, 1989), we changed the perceptual-identification task from naming the flashed word aloud to forced choice; after the target word was flashed, two alternative words were presented and the subject was asked to choose which of the two was the word that had been flashed. With the forced-choice procedure, identification of a word is facilitated by prior processing of exactly the same word and it is inhibited by prior processing of a similar but different word. When we began to examine models, we found that these two effects, by themselves, did not sufficiently constrain theory development. There was not enough empirical information to allow choices among competing hypotheses. Therefore, we conducted the series of experiments described in this article to collect additional parametric data.

In collecting parametric data, our main aim was model building, including examination of what kind of mechanisms could provide a quantitative account of word-identification and bias data. A simultaneous goal, however, was to bridge the gap that has existed between information-processing models on the one hand and the phenomena of implicit memory on the other. For the most part, implicit memory theorists have been concerned with the differences between implicit memory and other forms of memory and not with specifying the cognitive mechanisms responsible for implicit effects like priming. On the other side, theories of word identification have not usually had priming or bias effects as major targets for modeling. It was our hope that development of a quantitative model would inform both sides by demonstrating the leverage given by priming effects to learning about cognitive mechanisms and the leverage given by an understanding of potential mechanisms to the debate about multiple memory systems. Our approach was similar in spirit to Jacoby's: to emphasize the continuity between perception and memory through a unitary view of processing. Our intention was to provide such emphasis by providing a detailed model.

Before describing the model, we review what is already known about priming in word identification and then present a series of experiments that was designed to more fully characterize priming and bias than was possible from the earlier research. The results of the experiments led us to develop a new model for word identification and priming because none of the existing models could accommodate all of the data. After the experiments, we describe the model, discuss how it relates to other models, and consider what implications it might have for the debate about multiple memory systems.

Overview of Experiments

Our starting point for testing potential models of priming in word identification was the bias effect. Ratcliff et al. (1989; see also Johnston & Hale, 1984) found this effect for both word-identification *(naming)* and forced-choice paradigms. The experiments were similar to Jacoby's (1983a, 1983b; Jacoby & Dallas, 1981). Words were presented for study, and then at test, target words were flashed briefly. In the naming paradigm, subjects were asked to say aloud the word that had been flashed. In the forced-choice paradigm, subjects were asked to choose which of two words matched the one that had been flashed. When the flashed target had previously appeared for study, the probability of a correct response increased in both the naming and forced-choice paradigms. When a word visually similar to the flashed target, but not the target itself, had been studied, then there was an increase in the probability of the similar (studied) word as an incorrect response; this was true in both naming and forced choice. For example, suppose the target *died* was flashed, followed by the forced-choice alternatives *died* and *lied.* If *died* had been studied previously, then subjects would tend to choose *died,* a correct response. But if *lied* had been studied earlier, they would tend to choose *lied,* an incorrect response. In forced choice, this cost to performance in terms of incorrect responses almost exactly balanced the benefit in terms of correct responses. Thus, prior study did not lead to an overall improvement in perceptual identification.

The second critical point to be explained by a model is that prior study affects forced-choice performance only when the two alternative choices are similar to each other, not when they are dissimilar (Ratcliff et al., 1989; see also Masson & MacLeod, in press). If the flashed target is *died* and the two choices are *died* and *sofa,* then prior study of *died* does not increase the probability of correctly choosing *died* and prior study of *sofa* does not decrease the probability of correctly choosing *died.*

The bias effect and the fact that it holds only for similar alternatives in forced choice are two basic results that a model must explain. However, there are a number of additional pieces of information that are desirable. From some perspectives, a bias in word identification instead of an overall improvement in performance is surprising (although it might be expected from other views, e.g., Broadbent, 1967; Morton, 1968). The first two experiments reported in this article were designed to locate the forced-choice bias more firmly in identification processes. It might have been thought that bias came about from some explicit retrieval strategy, and Experiment 1 provided evidence against this possibility (see also Light & Kennison, in press; McKoon & Ratcliff, in press). It might also have been thought that, in addition to bias, there could still be an overall improvement in identification performance if both forced-choice alternatives were studied, and Experiment 2 evaluates this possibility. These experiments also served to replicate Ratcliff et al.'s (1989) previous results.

Experiments 3, 4, 5, and 6 examined the effects of two variables on perceptual processing. One was the amount of time for which the target word was flashed. The question for modeling was whether the amount of bias decreased with flash time. A model for which identification is strongly dependent on information accruing from the stimulus would predict that it should. The other variable was how closely the visual features of the flashed target word matched the visual features of the previously studied word. An explanation of priming mainly in terms of perceptual processes would predict a reduction in priming as a function of amount of mismatch. Finally, Experiment 7 investigated the effect of a delay of 30 min between study and test. We assumed that naming and forced-choice data would be explained with the same processing mechanism. If this were true, then delay should lead to equivalent effects in the two paradigms. We also assumed that a model for naming and forced choice should extend to other paradigms, and we tested a third task in Experiment 8, a task in which a single word was presented immediately following the flashed target word and subjects were asked to decide if the two words were the same.

The sections that follow begin with an overview of the experimental materials and procedures that were common to the experiments, and then the experiments are reported individually. After that, discussion proceeds to presentation of our counter model for priming in perceptual identification and to evaluation of the other current models, mentioned above, that can potentially explain priming.

General Method

Materials

Triples of words were constructed. All three words of a triple had the same number of letters, always between 4 and 7. Two of the words of a triple had the same visual shape, and the third was as different in shape as possible, where shape refers to the outline of the letters. For example, *d* would be similar to *l* but dissimilar to *s,* and *p* would be similar to *g* but dissimilar to *h.* The two words *data* and *date* have about the same shape but the third word of their triple, *club,* has a different shape. The words were chosen from the Brown University corpus (Kucera & Francis, 1967). For the first three experiments, there were 80 triples for which word frequency was not controlled. For each of these triples, the two similar words differed by one similarly shaped letter or (in 34 cases) by two or more similarly shaped letters (these were words of five or more letters). For the remainder of the experiments, there were 168 triples. For 61 of these, all three words had relatively high frequencies, between 78 and 10,601; for 54 triples, the words all had lower frequencies of either 4 or 5; and for the remaining triples, the three words were of varying frequencies. The similar words differed from each other in only one similarly shaped letter (154 words) or two or three similarly shaped letters (14 words of six or seven letters in length).

Equipment

Stimulus materials were displayed on an oscilloscope with a fast phosphor that was programmed to present words with stimulus presentation times ranging from 1 ms up, in increments of 1 ms. Words were written on the oscilloscope in letters produced by a character generator in the oscilloscope hardware. The oscilloscope was controlled by a PC, and responses were collected on the PC keyboard.

Procedure

In all experiments, words presented for study were displayed one at a time, for 1 s per word. Subjects were instructed to learn the words for a later (unspecified) test. Study lists were typically separated from test lists by a warning signal (a row of asterisks) displayed for 2.5 s. The sequence of events for each test item with the forced-choice procedure was as follows: A row of minus signs was presented for 400 ms as a warning signal; this was followed by a blank screen for 300 ms; then the test word was flashed; then a mask was displayed for 300 ms, covering where the test word had been; then two words were displayed side by side on the line below, and subjects had to choose which of them had been flashed. Subjects responded by pressing the Z key on the keyboard to indicate that the left-hand word was the flashed test word or the ?/ key to indicate the right-hand word. After the response, the warning signal began the sequence over again for the next item. The mask was a row of "@" characters that were displayed in a larger font than the letters of the flashed test words so that the mask completely covered all the space that the letters had covered. Which of the two alternatives was the correct choice was decided randomly. The sequence of events for each test item with the naming procedure was the same except that two choice words were not displayed following the masked target word. Instead, the mask was followed by a row of question marks that lasted until the subject responded. Subjects were instructed to name aloud the word that was flashed or, if they could not name the word, to say "no." In all the experiments, participants were instructed that the flashed target words were difficult to see and that they should try hard to do their best.

In some experiments (Experiments 1 and 2, forced-choice procedure), the first block of test items was used for calibration, to set individually for each participant the amount of time for which target words would be flashed. In the other experiments, the first block of test items was used for practice, with the same flash time (or times when flash time was a variable manipulated in the experiment) used for all subjects. Whether the first block was calibration or practice, the flash time for the first four test words was set very long, 100 ms, to orient the subject to the sequence of test events. The remainder of the block consisted of a series of 50 test trials. For the experiments in which flash times were calibrated for each individual subject, two flash times (typically 15 and 30 ms) were used for the 50 trials. Accuracy for each flash time was printed on the PC screen at the end of the trials, and the experimenter used these accuracy values to choose a flash time for the subject to be used for the remainder of the experiment, aiming at 75% correct performance.

All of the subjects took part in the experiments in order to fulfill a requirement of an introductory psychology course at Northwestern University. In all of the experiments, the assignment of triples to conditions, the choice of which of the similar words of a triple would be the flashed target, and the order of study and test items was random, with the randomizations changed after every second subject.

Experiment 1

Ratcliff et al. (1989) found that prior study biased responses towards the previously studied one of the two forced-choice alternatives when the two alternatives were similar (*died–lied*), but not when they were dissimilar (*died–sofa*). One interpretation of this result is that subjects were using some kind of explicit retrieval strategy by which similarity of the alternatives led the subjects to rely on memory for prior presentations of the words whereas dissimilarity of the alternatives led them to rely only on the stimulus itself. A strategy like this should be subject to manipulation. Making the strategy applicable to only a small proportion of test items should tend to eliminate its use.

This prediction was tested by varying the probability with which the forced-choice alternatives for a test item were similar versus dissimilar. For one group of subjects, the probability of similar alternatives was .8 and the probability of dissimilar alternatives was .2. The probabilities were reversed for the second group of subjects. If bias was a strategic retrieval effect, a high probability of dissimilar choices should reduce or eliminate bias for similar choices, and a high probability of similar choices might induce bias for dissimilar choices.

The conditions of the experiment are shown in Table 1. One of the similar words of a triple was the flashed target word. There were three possible study conditions for a target: The target itself was studied, a word similar or dissimilar to it was studied, or none of the words of the target's triple was studied. For example, if the target was *died,* then *died* was studied, the similar word of the triple, *lied,* was studied, the dissimilar word of the triple, *sofa,* was studied, or none of these was studied. A bias finding would be an increased likelihood of choosing the word that had been studied. This would be a correct choice if the flashed target was the one of the two forced-choice alternatives that had been studied but an incorrect choice if the other of the forced-choice alternatives had been studied.

Method

There were 16 subjects in the 80%-similar group and 16 subjects in the 80%-dissimilar group. The experiment began with the calibration block of test items. For the 80%-similar group, all the practice pairs used similar alternatives, whereas for the 80%-dissimilar group, all the practice pairs used dissimilar alternatives. The dissimilar alternatives led to better performance and because the experimenter was aiming for 75%-correct performance for the calibration pairs, the flash times were set to be shorter for subjects in the dissimilar group. The flash times assigned to individual subjects varied from 17 ms to 38 ms in the 80%-similar group and from 12 ms to 30 ms in the 80%-dissimilar group.

Following the calibration block, subjects began a sequence of four study-test blocks. In each block, there was a study list of 10 single words, followed by a test sequence of 20 test items.

A total of 80 triples was used. For each triple, one of the two similar words was designated the target to be flashed in the test phase (which word was decided randomly). There were two test conditions: The two words that were presented for forced choice were either the target and the similar word from its triple or the target and the dissimilar word from its triple. There were also three study conditions. In the first study condition, the target had been presented in the study list. In the second study condition, the other of the forced-choice alternatives had been studied. In the third study condition, neither the target nor the forced-choice alternative had been in the study list. For example, for the target *died* with similar forced-choice alternatives *(died lied),* either *died, lied,* or neither was studied. For the target *died* with dissimilar alternatives *(died sofa),* either *died, sofa,* or neither was studied. The two test conditions and the three study conditions were crossed to form the six conditions shown in Table 1.

Table 1
Probability Correct in Forced Choice, Experiment 1

	Study condition		
Similarity condition	Target studied	Distractor studied	Neither studied
Proportion of similar alternatives = .8			
Similar *(died* vs. *lied)*			
All responses	.85	.66	.75
Fastest two thirds	.90	.73	.82
Dissimilar *(died* vs. *sofa)*			
All responses	.83	.88	.87
Fastest two thirds	.93	.96	.93
Proportion of similar alternatives = .2			
Similar *(died* vs. *lied)*			
All responses	.74	.51	.62
Fastest two thirds	.83	.55	.76
Dissimilar *(died* vs. *sofa)*			
All responses	.78	.77	.79
Fastest two thirds	.85	.83	.84

For one group of subjects, 80% of the items were tested with similar alternatives and 20% with dissimilar alternatives, and for the other group, 80% were dissimilar and 20% similar. In each of the four study–test blocks, there were four test items from each of the high-probability conditions and one test item from each of the low-probability conditions, except that in the no-study baseline conditions, the number of test items was doubled (eight for high probability, two for low probability).

Results and Discussion

Table 1 shows the mean probability correct for each condition. In general, the results replicate those obtained by Ratcliff et al. (1989; note that in the Ratcliff et al. experiments, target words were studied as parts of sentences, whereas here they were studied in lists of single words). For the conditions in which similar words appeared in the forced-choice test, there were both costs and benefits from prior study: costs when the studied word was similar to but different from the flashed target and benefits when it was identical. The benefits were about balanced by the costs; the benefits averaged 11%, and the costs averaged 10%, so that there was no overall improvement in accuracy (d')as the result of prior study. This replicates Ratcliff et al.'s results. For the conditions in which dissimilar words appeared in the forced-choice test, there was no discernible effect of prior study. Most importantly, this pattern of results was not affected by the probability manipulation; bias effects were observed with similar forced-choice alternatives both when similar choices occurred in the experiment with high probability and when they occurred with low probability. Thus it appears that the existence of a bias effect with visually similar alternatives and the lack of such an effect with dissimilar alternatives are not the results of a special strategy adopted by subjects as a function of the type of alternatives.

Analyses of variance confirmed these conclusions ($p < .05$ throughout this article unless otherwise specified). With the similar forced-choice alternatives, there were significant differences across the three study conditions, $F(2, 30) = 15.04$ (standard error of the mean [SE_M] = .02) for the 80%-similar group and $F(2, 30) = 4.54$ for the 80%-dissimilar group ($SE_M = .06$). With the dissimilar forced-choice alternatives, there were no significant differences, $F(2, 30) < 1.0$ for the 80%-similar group ($SE_M = .04$) and $F(2, 30) < 1.0$ for the 80%-dissimilar group ($SE_M = .03$).

The data also provide another strong piece of evidence against the possibility that subjects choose a response by strategically accessing explicit memory for the prior study episode. Such a strategy is likely to be difficult or time consuming and so would only be employed when an initial attempt at identifying the stimulus had failed. It follows that bias should appear only for relatively slow responses (see Ratcliff & McKoon, 1995, who found slow explicit retrieval in the object-decision task). To examine this possibility, the slowest third of the responses (which were responses slower than the mean response time for the conditions with similar forced-choice alternatives) was eliminated and the data reanalyzed (see Table 1). The resulting might be that the smaller the quantity of incoming information (i.e., the faster the flash time), the less opportunity there would be for bias in the system to affect the outcome of identification processes. Alternatively, it might be that bias applied equally no matter how little information entered the system.

Across Experiments 3 and 4, flash time was varied from 10 ms to 45 ms. This range produced differences in performance from near chance to near ceiling. It allowed the collection of parametric data that show the functional relationship between flash time and the bias effect. Any potential model must account for the size of the bias effect as a function of stimulus duration.

Both experiments examined bias in forced choice with the three study conditions of studying the target, studying another word of the target's triple, or neither. In Experiment 3, only words similar to the target were used and there were four flash times. In Experiment 4, both similar and dissimilar words were used and there were three flash times.

Method

The experiments began with the block of practice test items. The flash times were the same for all subjects. In the practice block, the first four flash times were 100 ms and the remainder were either 20 ms or 40 ms. After this practice, in Experiment 3, there were five study-test blocks, with 16 words studied in each block and 32 tested. In Experiment 4, there were seven blocks of 12 study words and 24 test words.

In Experiment 3, 160 triples were used, and in Experiment 4, 168 triples were used. In Experiment 3, there were 3 study conditions: For each triple, the word designated the target was presented in the study phase, the word similar to it was presented in the study phase, or neither was studied. In the test phase, the target was flashed for 15 ms, 25 ms, 35 ms, or 45 ms. With study and test conditions combined, there were a total of 12 conditions, with 10 words tested in each condition except that there were twice as many words in each no-study condition. The two alternatives for each test item were always the target and the word similar to it. In Experiment 4, test conditions were added in which the two alternatives were the target and the word dissimilar to the target. There were 3 study conditions: study the target word, study another word of the target's triple (the similar word if the forced choice alternatives were similar or the dissimilar word if the alternatives were dissimilar), or study no word of the triple. The 3 study conditions combined with the 2 test conditions gave 6 conditions, which were combined with 3 flash times (10 ms, 20 ms, and 40 ms) to give a total of 18 conditions in the experiment. Seven words were tested in each condition except that there were twice as many in the no-study, baseline conditions. For both experiments, the assignment of conditions was equal across study-test blocks. There were 18 subjects in Experiment 3 and 32 subjects in Experiment 4.

Table 3
Probability Correct in Forced Choice, Experiment 3

Flash time (ms)	Study condition		
	Target studied	Similar word studied	Neither studied
15	.620	.444	.544
25	.750	.545	.660
35	.833	.678	.756
45	.883	.778	.831

Table 4
Probability Correct in Forced Choice, Experiment 4

Flash time (ms)	Study condition			
	Similar alternatives			Dissimilar alternatives (3 study conditions grouped)
	Target studied	Similar word studied	Neither studied	
10	.585	.408	.507	.551
20	.667	.564	.635	.675
40	.804	.686	.745	.880

Results

The main result of these experiments was that the amount of bias was not affected by flash time. For Experiment 3, the two alternatives for forced choice were the target and the word similar to it. The mean probabilities correct are shown in Table 3. The data show a benefit to prior study of a target and a cost to prior study of a similar word, and they show that the cost and benefit remain of about the same magnitude as flash time is decreased toward an average probability correct equal to chance (.50). Overall, there was no significant interaction between flash time and study condition, $F(6,\ 102) < 1.00$.

For the conditions in which the two alternatives were similar, the data from Experiment 4 (Table 4) show the same pattern as the data from Experiment 3: costs and benefits due to prior study that do not decrease as flash time decreases. However, when the two alternatives were dissimilar, there were no significant costs or benefits due to prior study, replicating the finding of Experiment 1.

In Experiment 3, the probabilities of correct responses increased significantly with flash time, F(3, 51) = 42.07, as would be expected, and there were significant differences due to study condition, F(2, 34) = 32.45. The *SE* of the probability correct was .033.

In Experiment 4, the main effects for all three variables were significant: Probability correct increased with flash time, $F(2,\ 62) = 149.51$, dissimilar alternatives led to more correct responses $F(1,\ 31) = 54.54$, and the probability correct differed with study condition, F(2, 62) = 9.44. Study condition affected probability correct for similar alternatives but not for dissimilar alternatives, $F(2,\ 62) = 6.11$; and this did not change across flash-time conditions, $F(4, 124) < 1.0$. There was an interaction such that increased flash time affected the probability correct for dissimilar alternatives more than for similar alternatives, $F(2,\ 62) = 6.08$. This effect would be expected because there are more distinguishing visual features for dissimilar than similar alternatives. The *SE* of the probability correct was .023.

Experiment 5

The results of Experiments 3 and 4 indicate that the bias effect does not depend strongly on the quantity of information coming into the system from the stimulus. Experiment 5 was designed to ask whether it depended on the quality of the match between features of the incoming information and features of the prior processing episode. Specifically, the flashed target word

Table 5
Probability Correct in Forced Choice, Experiment 5

	Study condition		
Case condition	Target studied	Similar word studied	Neither studied
Flash lower, choice lower	.734	.592	.663
Flash upper, choice lower	.810	.619	.685
Flash lower, choice upper	.774	.569	.690
Flash upper, choice upper	.803	.669	.753

Note. All study words were in lowercase. Standard error in the first two columns is about .04 and in the third column about .03. Mean probabilities for the four rows are .658, .695, .672, and .738.

was either in the same case (upper or lower) as the previously studied word or in the opposite case.

Previous research indicates that priming in the naming paradigm is not particularly sensitive to the match between the target and the previously studied words, so long as they are presented in the same modality (see Clarke & Morton, 1983, for a discussion of modality changes). With handwritten versus typewritten words, which have mismatches in visual features, Clarke and Morton (see also Jackson & Morton, 1984) found that the priming effect was the same whether the form matched between study and test or mismatched. Jacoby and Hayman (1989) used lowercase and uppercase words, for which there are also major mismatches in visual features (e.g., *died* vs. *DIED).* Under normal reading conditions, they found a small effect of mismatching case only in one condition of one experiment, although they did find larger effects with a type font that was very difficult to read or with subjects instructed to read the words one letter at a time. In general, previous research does not show the dramatic effects of visual mismatches that would be expected if priming in naming is completely based on perceptual aspects of the stimuli. The aim of Experiment 5 was to extend this generalization to the forced-choice paradigm.

Method

The experiment began with the practice test items, with the flash times fixed at 25 ms. After the practice, there were five study-test blocks, with 16 study words and 32 test words in each block.

For the experiment, 160 triples were used. There were three study conditions: A target test word had been presented in the study phase, the word similar to it had been studied, or neither had been studied. All words were presented for study in lowercase. There were four test conditions defined by crossing the case of the flashed target (uppercase or lowercase), with the case of the two alternatives presented for forced choice (both uppercase or both lowercase). The four case conditions were crossed with the three study conditions to yield a total of 12 conditions. Ten items were tested in each condition except the no-study conditions for which 20 were tested. Conditions were represented equally across study-test blocks. There were 15 subjects in the experiment.

Results

The mean probabilities of correct responses for each condition are shown in Table 5. The same bias effect due to study condition was found as in all the previous experiments, $F(2, 28) = 22.28$. Although there was a small but significant main effect of the four case conditions, $F(3, 42) = 4.14$, the case variations did not significantly alter the bias effect, $F(6, 84) < 1.0$. It appears that performance was somewhat better when the flashed word was in uppercase, especially when the forced-choice words were also in uppercase, but individual differences among pairs of case conditions were not significant, $Fs < 1.0$. The SE_M probability correct was .032.

Experiment 6

A model of priming in perceptual identification should account for priming in the naming and the forced-choice paradigms simultaneously. There is considerable variability across participants in their accuracy in performing the tasks, and a model should be able to predict the relationship between performance on the two tasks for individual participants. Given the level of performance on one of the tasks, the model should be able to predict the level of performance on the other. To obtain the appropriate data for testing models in this way, Experiment 6 mixed the naming and forced-choice tasks in the same experiment.

Method

For 11 subjects, the experiment began with the practice block, half forced choice and half naming, with flash times fixed at 35 ms (and the first four flash times at 100 ms). The experiment then continued with two study-test blocks. For the other 11 subjects, the study list of the first block was presented first in the experiment, then the practice block, then the test part of the first block, and finally the study and test lists of the second block. There were no other differences in the procedure for the two groups of subjects, and there were no differences in their performance, so their data were combined in the analyses presented below. There was a total of the two study-test blocks, with 32 study words and 64 test words in each block, and the flash time for all target test words was 35 ms.

In the experiment, 128 triples were used. There were three study conditions: The flashed target word had been presented in the study list, the word similar to it had been studied, or neither had been studied. There were two test conditions: The flashed target was followed either by the two similar words presented for forced choice or by a row of question marks that stayed on the screen until the subject responded. Subjects were instructed that the question marks were their cue to say aloud the word that had been flashed, or if they could not, to say "no." For each of the six conditions (three study conditions crossed with two test conditions), there were 16 test items except for 32 test items in each of the no-study conditions. The assignment of conditions was equal across study–test blocks.

Results

The data are shown in Table 6. The forced-choice responses showed the typical effect of study condition, $F(2, 40) = 12.81$, with a benefit from prior study of the target and a cost from prior study of the word similar to the target. The SE_M probability correct was .021.

Study condition also significantly affected the probability of correct naming responses, $F(2, 40) = 14.62$. Prior study of the target substantially increased the probability of a correct response to the target, but prior study of the similar word had little effect on target responses. The SE_M probability correct

Table 6
Probability Correct in Forced Choice and Naming

	Study condition					
	Experiment 6			Experiment 7		
Alternatives	Target studied	Similar word studied	Neither studied	Target studied	Similar word studied	Neither studied
Forced choice	.81	.67	.74	.77	.73	.76
Naming	.42	.28	.27	.36	.31	.29
Intrusion rate[a]	.01	.03	.01	.02	.04	.01

Note. In Experiment 7, with a 10-ms flash time, forced-choice performance was .55, .47, and .50 in the study-target, study-similar-word, and study-neither conditions, respectively.

[a] For the similar word in naming.

was .022. Subjects sometimes responded, incorrectly, with the similar word instead of the target, and the probability of such responses was higher, *F*(2, 40) = 4.68, when the similar word had been studied (see Jacoby & Dallas, 1981). The SE_M probabilities of a response of the similar word was .001.

The top panel of Figure 2 shows the probabilities of correct responses for the two tasks for individual subjects. The base of an arrow (*x* for a human subject and *o* for a theoretical prediction, to be discussed later) is the probability correct for the no-study condition, and the head of an arrow represents the amount of facilitation. The facilitation for forced choice was calculated as half the difference between the probability correct when the target was studied and the probability correct when the similar word was studied. The probabilities vary widely across subjects. Naming performance ranges from near zero to about 80% correct, while forced-choice performance ranges from around chance (50%) to about 90% correct. For the bottom panel, subjects were put into three groups as a function of naming performance (low, middle, and high performance). These data were used to test the model described in the sections that follow presentation of the experiments.

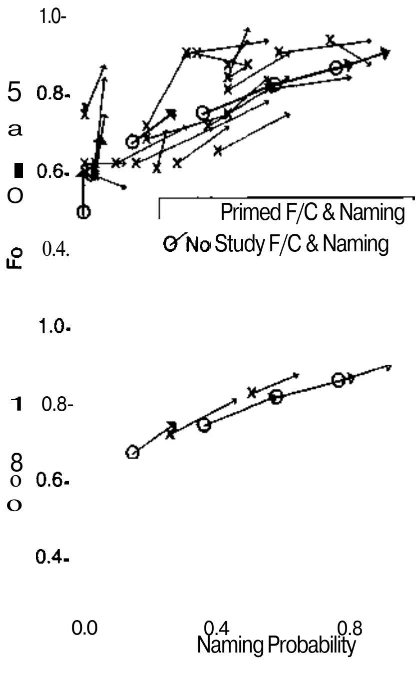

Figure 2. Probabilities correct in forced choice (F/C) and naming, plotted against each other. The top panel shows results for 25 subjects (the *X* symbols), and the bottom panel shows the same results for the subjects grouped into 3 groups based on naming performance (lowest 8, middle 8, and highest scoring 9 subjects). The counter model predictions are shown by the *O* symbols.

Experiment 7

If performance on naming and performance on forced choice are to be explained with the same processing mechanisms, then it would be expected that many variables would affect them in parallel ways. We tested one such variable, amount of forgetting. In general, a notable characteristic of implicit memories is that their effects on performance are not reduced as much by delay between study and test as the effects of explicit memories. Previous work on delay and priming in perceptual identification has shown that the priming effect can last several days (Jacoby, 1983a; Jacoby & Dallas, 1981; see Whitlow & Dalton, 1993, for a summary of delay effects). In Experiment 7, we used a delay of 30 min to compare naming and forced choice.

Method

The experiment began with a study list of 108 words. This was followed by an unrelated experiment on reading that took about 30 min. Then the practice block was presented with half naming and half forced-choice test items with 25 ms flash time (the first four tests were at 100 ms flash time), and then a test list of 162 items.

In the experiment, 162 triples were used. There were three study conditions: The flashed target word had been presented in the study list, the word similar to it had been studied, or neither had been studied. There were three test conditions: The target was flashed for 10 ms and followed by two words presented for forced choice, it was flashed for 35 ms and followed by two words for forced choice, or the target was flashed for 35 ms and followed by a row of question marks that continued until the subject responded by saying aloud the word that had been flashed, or if they could not, by saying "no." In the forced-choice conditions, the similar words were used for the alternative choices. For each of the nine conditions (three study conditions crossed with three test conditions), there were 18 test items except for 36 test items in each of the no-study conditions. The assignment of conditions was equal across study-test blocks. There were 25 subjects in the experiment.

Results

The forced-choice data showed smaller effects of prior study than had been observed in previous experiments (see Table 6). However, the effect of prior study was still marginally significant, $F(2, 48) = 2.60$, $p = .08$. A planned test showed the difference between prior study of the same word and prior study of a similar word significant for the 10 ms flash time conditions, $F(1, 48) = 4.54$. The main effect of better performance for the 35 ms flash time was significant, $F(1, 24) = 58.50$. The SE of the probability correct was .02.

The naming data (Table 6) showed significant effects of prior study in that correct responses of the target word and incorrect responses of the similar word were both increased by prior study, $F(2, 48) = 3.40$ and $F(2, 48) = 4.76$, with $SE = .020$ and .006, respectively.

In order to address the question of how naming and forced choice were affected by delay between study and test, we compared the results of this experiment for the 35 ms flash time to the results from Experiment 6. We added three additional subjects to that study in order to make a total of 25 in each of the two experiments. As the data in Table 6 show, the effect of prior study decreased with delay, and it decreased in about the same proportion for forced choice (.14 to .04) as for naming (.14 to .06). For forced-choice responses, the decrease in the effect of prior study across delay was significant, $F(2, 96) = 3.25$ ($SE_M = .015$). The decrease was also marginally significant for correct naming responses to the flashed target, $F(2, 96) = 2.75$, $p = .07$ ($SE_M = .014$). There were too few observations to provide enough power to test the delay effect for incorrect responses of the similar word. The overall amount of decay in naming was greater than is usually found after a 30-min delay (Whitlow & Dalton, 1993), but the level of performance was much nearer floor than in other studies. The important result, however, is that performance in naming and forced choice appear to decay at about the same rate.

Experiment 8

Just as quantitative models in other domains apply to multiple tasks, so should a model for perceptual identification. In memory, for example, the global memory models have been applied to recognition, recall, frequency judgments, recency judgments, and lexical decision (Gillund & Shiffrin, 1984; Hintzman, 1986, 1988; Lewandowsky & Murdock, 1989; Murdock, 1982, 1983; Ratcliff & McKoon, 1988). The standard paradigm for perceptual identification has been naming, and forced choice was a recent addition. Experiment 8 was designed to add one more procedure, one that required a "yes-no" decision. A word was flashed as for the other tasks. It was immediately followed by a single word, and subjects were asked to decide whether or not the single word was the same as the flashed word. As it turned out, this procedure provided data somewhat different from the naming and forced-choice paradigms, and in so doing provided a stringent test for possible models of perceptual identification.

Method

The experiment began with the practice block, and then 8 study-test blocks followed. The flash time was 25 ms for all words except the first four in the practice block. The procedure was the same as for forced choice except that at the point where two words would have been presented as alternative choices, only one word was presented. If this word matched the one that was flashed, subjects were instructed to press the ?/ key of the computer keyboard; if it did not match, they were instructed to press the Z key. In the experimental blocks, there were 14 words in each study list and 20 targets in each test list.

In the experiment, 160 triples were used. The word presented for the "yes-no" decision was chosen from one of the two similar words of a triple. There were four study conditions: The "yes-no" decision word was presented in the study list, the word similar to it was studied, the word dissimilar to it was studied, or none of them was studied. There were 3 test conditions: the "yes-no" decision word was flashed, the word similar to it was flashed, or the word dissimilar to it was flashed. The 10 combinations of study and test conditions that were used in the experiment are shown in Table 7 (with the triple *died, lied, sofa* as an example). For each of the 10 conditions, there were 16 test items. The assignment of conditions was equal across study-test blocks. There were 32 subjects in the experiment (plus seven who were excluded because they were unable to discriminate between target items and similar or dissimilar alternatives).

Results

Table 7 shows the mean probabilities of "yes" responses across the 10 conditions. Not surprisingly, the probability of subjects saying the test word matched the flashed target, a "yes" response, was low if the word presented for the "yes-no" decision was dissimilar to the flashed word (the four conditions on the right), higher if it was similar (the middle three conditions), and highest if was the same as the flashed word (the three conditions on the left). The most interesting result, and one that contrasts with the forced-choice procedure, was that a decision was affected by prior study of the decision word itself but not by prior study of a similar word. Using the examples in Table 7, when *died* was flashed, the probability of a "yes" response increased from .70 in the baseline condition to .77 when *died* had been studied, but there was no decrease when *lied* had been studied (there was actually a slight increase, to .73). The same pattern held when the flashed word was *lied* (an increase from .51 to .55) and when it was *sofa* (an increase from .21 to .23). Averaging over these conditions, the probability of a "yes" response to the decision word increased if it was previously studied, by about .04, but it did not decrease if a similar word was studied. This finding was independent of whether the "yes" response was correct or not. The increase in the probability of a "yes" response is an increase in probability correct if the decision word was actually the flashed word and a decrease in the probability correct if a different word was flashed.

For analysis of variance (ANOVA), the last condition in Table 7 was excluded in order to test a two-factor design (3 study conditions by 3 test conditions). The effect of which word was flashed was significant, $F(2, 62) = 125.78$, as was the effect

Table 7
Probability of Positive Response in Yes–No Task; Experiment 8

Word studied	Target word flashed	Similar word flashed	Dissimilar word flashed
Neither studied	0.70 (— *died died*)	0.51 (— *lied died*)	0.21 (— *sofa died*)
Target	0.77 (*died died died*)	0.55 (*died lied died*)	0.23 (*died sofa died*)
Similar	0.73 (*lied died died*)	0.49 (*lied lied died*)	0.21 (*lied sofa died*)
Dissimilar			0.18 (*sofa sofa died*)

Note. In each condition, the words in parentheses show the studied word, the flashed word, and the target yes-no decision word. — = neither studied.

of study condition, $F(2, 62) = 4.28$. Their interaction was not significant, $F < 1.00$, $SE_M = .02$.

Model Testing

There are four main features of priming in perceptual identification that a model must explain. First, forced-choice performance is biased by prior study such that subjects tend to choose a word that was previously studied over a word that was not. However, and this is the second critical piece of data, this bias obtains only when the two forced-choice alternatives are visually similar to each other, not when they are dissimilar. Bias for similar but not dissimilar words was shown by Ratcliff et al. (1989) and also in Experiments 1 and 4.

The third feature of priming that affects the choice of potential models is that varying the probability of similar versus dissimilar alternatives in forced choice has no significant effect on the amount of bias. This result (Experiment 1) means that bias cannot be explained as a consciously controlled strategy whereby subjects explicitly retrieve information about prior study. This conclusion is also supported by the finding that bias is observed when subjects respond quickly as well as when they respond slowly.

Fourth, bias is relatively unconstrained by the stimulus information that comes into the perceptual processing system. When the flash time for a word is reduced to the point where performance is near chance, the amount of bias does not decrease, and in fact is at about its largest value (Experiments 3 and 4; Tables 3 and 4). Bias is also not strongly affected by the amount of match in terms of visual features between a word presented for study and a word flashed for test (Experiment 5).

In the following sections, we show that the combination of these aspects of priming in perceptual identification cannot be explained by Morton's (1969) logogen model, by McClelland and Rumelhart's (1981) interactive action model, or by Seidenberg and McClelland's (1989) connectionist model. We then describe a new model, the counter model, that does encompass all four features of priming, as well as make correct predictions about several other aspects of perceptual identification.

Logogen Model

The logogen model (Morton, 1969, 1970; Murrell & Morton, 1974), one of the earliest models for word identification, proposes that each word in the lexicon is represented by a counter mechanism, a logogen. An incoming word stimulus causes information to be accumulated in the counters of all words that match the stimulus to some degree. A response is initiated when the amount of information accumulated in one of the logogens exceeds that logogen's threshold value. Priming due to prior exposure to a word is viewed as a temporary lowering of the threshold of the word's logogen (or equivalently as a raising of the resting level of the logogen, Morton, 1969, 1970). In the most current version of the model (Clarke & Morton, 1983; Jackson & Morton, 1984), there are three sets of logogens, one for visual stimuli, one for auditory stimuli, and one representing the output system.

The logogen model can explain bias in naming and in forced choice with similar alternatives as the temporary lowering of a word's threshold due to prior exposure. The features of a flashed target cause accumulation of evidence in the counters of all words that are similar to the target. If the target was previously studied, its lowered threshold tends to be reached sooner than the thresholds of similar words and so the likelihood that it is given as a correct response increases. If a word similar to the target was studied, its threshold is the one that is lowered and so it tends to be reached sooner than the target's, leading to an incorrect response.

The logogen model could also explain how the amount of bias in forced choice could remain unchanged as the flash time for the target was decreased to the point where performance was at chance. It could be assumed that as less and less information was available from the stimulus (as flash time decreased), the threshold required for a response was lowered for all words by an equivalent amount. Lowering the thresholds would enable a response even under circumstances in which stimulus information was extremely impoverished, and it would preserve the advantage of previously studied words. With no stimulus information at all, the logogen model (in an unmodified form without any assumptions about variability in resting levels or thresholds) would predict that a previously studied word would always be chosen over other words.

However, the logogen model (with the outlined assumptions) cannot explain why there is no bias in forced choice with dissimilar alternatives. A lowering of the threshold for a previously studied word should facilitate identification of that word whether the alternatives are similar or dissimilar. There is no mechanism by which the accumulation of evidence in one counter or its threshold can be affected by the evidence or threshold of another counter (unless a modification in the threshold causes a large change in the proportion of trials for which responses are produced from that counter and so produces changes in the probabilities of responses from other counters). Morton (1970) has directly stated that the logogen model was designed so that a response would be based on the absolute numbers of counts in independent counters and not on "a comparison of the counts in different logogens'' (p. 208).

Interactive Activation Model

The interactive activation model designed by McClelland and Rumelhart (1981) for word recognition is a local connectionist model. Information is represented in a network of three levels of nodes: letter feature (line segment) nodes, letter nodes, and word nodes, plus facilitative and inhibitory connections among nodes in different levels and inhibitory connections among nodes in the same level. The model represents feature, letter, and word nodes for 1,179 four-letter words (see McClelland & Rumelhart, 1989). A stimulus is input to the system as activation to the feature nodes. Activation is then propagated through the system until the stimulus is turned off or masked, at which point input activation is set to zero. With experience, a response mechanism learns the point in time at which the level of activation in the output layer reaches a maximum value. At that time, activation values on the word or letter nodes (either of these can serve as the output layer and which depends on the task and materials) are used to make decisions about what word was input to the system. Decisions are made using the Luce (1959) choice rule, which computes the probability that a node is given as the response; the probability is the strength of the node divided by the sum of the strengths of all nodes. Strength of a node is defined as the exponential transform of a weighted average across time of the activation values of the node. Although the model correctly predicts a number of results in word identification (see McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), it has not been applied to repetition effects.

One way it might be supposed that an effect of prior study could be added to the model would be to increment the initial resting activation levels of the nodes of the letters and features that make up a previously studied word. However, this cannot correctly predict bias in forced choice. Which of the two alternative words is chosen in forced choice is determined by the Luce (1959) decision rule applied to the letter nodes (because in McClelland and Rumelhart's (1981) applications, the two alternative words always differed by only one letter, so a decision on this letter dictated a decision on the word). The choice is determined by which of the two alternatives has the greater strength relative to the strengths of all other nodes and so, because strength depends on the resting levels of activation, an advantage would be predicted for a word studied previously. However, an advantage would be given to the previously studied alternative whether the other alternative was similar to it or dissimilar to it, in contradiction of the data. There is no way for the model, under its current assumptions about how choices are made, to make the presence or absence of priming depend on the similarity of the alternatives.

The model also cannot explain why the amount of bias in forced choice does not decrease as the flash time of the target decreases. In the model, if the letter node with the greatest amount of strength relative to all letter nodes cannot be used to identify which alternative to choose (i.e., the letter is either in both words or neither of them), then the choice is determined by guessing. Because this outcome becomes more likely as flash time decreases, the size of the effect of prior study should decrease.

We discovered a third problem when we experimented with the computer program for the model (see McClelland & Rumelhart, 1989) to try to produce priming effects by incrementing the resting activation levels of words to simulate prior study. To simulate a test phase, activation was input to the system, and target words were presented to the system for forced choice. We found that the number of words that were similar and how they were similar (e.g., "friends" versus "enemies," McClelland & Rumelhart, 1981) to the target had a large effect on the results, and that the effect was different for different words. There were no consistent effects of modifying the resting levels of activation on target words or on words similar to the target words. For some words, there was a small bias to respond with the previously studied word along with a large bias to respond with a similar word; for other words, the pattern was exactly the opposite; and for some words, there was little effect of prior study (as well as cases in which both were facilitated or both inhibited). Thus, the effects on particular individual words were too highly variable to produce the systematic effects that are found in forced choice experiments. This conclusion about variability is the same as that reached by McClelland and Rumelhart (1981) for neighborhood effects.

Seidenberg and McClelland's Distributed Connectionist Model

The interactive activation model has no learning mechanism and so the only way to simulate repetition effects is to "hardwire" into the system the changes to a word's representation that might result from prior exposure. The Seidenberg and McClelland (1989) model for word recognition and naming was a major advance in that it has a learning mechanism as a central component. The model is a distributed connectionist network in which one layer of nodes provides an orthographic representation of a word, another layer provides a phonemic representation of a word, and a third (hidden) layer connects the other two. The model was designed to learn a mapping between the orthographic and phonemic representations of a word by a training process in which both representations are presented to the system. Activation cycles among the layers until the weights on connections among nodes arrive at values that allow the correct phonemic representation of a word to be produced in response to the word's orthographic representation. The model was trained on a large set of monosyllabic words, including all the words in the Kucera and Francis (1967) norms, and it accounts for a wide range of word-identification data.

We obtained the model (Seidenberg & McClelland, 1989) and attempted to simulate priming (see also Rueckl, 1990). Presentation of a word in a study list was treated as a learning trial, which resulted in slightly altered connection weights. The rapid presentation of a target test word was modeled by presenting only a subset of the orthographic units of the word to the system and then allowing the system to produce output at both the phonological and orthographic layers. We discovered that, because the network of connections is so highly overlearned, no small number of additional learning trials could alter the behavior of the system enough to produce changes even a fraction as large as those observed for priming in forced choice or naming. What would be needed is modification of the model to allow relatively large but temporary changes in performance; however, that would require a major modification of the model (but see Ratcliff, 1990).

The Counter Model for Word Identification

The counter model was designed to explain priming in masked word identification along with other major phenomena of masked word identification. Morton's (1969, 1970) logogen model was the starting point. The mechanisms of that model were modified and new mechanisms were added to explain how bias comes about, why it is obtained for similar but not dissimilar alternatives in forced choice, and why it occurs even when the flash time for a target word is so short that performance is near chance. We first give a verbal description of the counter model and then describe how it accounts parametrically for the data from the experiments presented above.

The model is a counter model, with one counter for each word in the system. The counters can be thought of as decision counters in that their accumulation of counts is the mechanism by which a decision is made about what response should be the output of the system. The system moves continuously towards a decision: Counts are accumulated at a constant rate such that, for each unit of time, one count is accumulated to one (and only one) counter. This is an important assumption of the model. Under the impoverished stimulus conditions of perceptual-identification experiments, it implies that, across time, there are counts that are determined by perceptual features of the stimulus, but there are also counts that are not determined by the stimulus because not enough stimulus information enters the system. The counts that correspond to perceptual features are accumulated by an appropriate counter. A count from a perceptual feature of the letter *d,* for example, might be accumulated by the counter for *died.* The counts that are not determined by the stimulus (null counts) can be accumulated to any counter. Essentially, a null count represents random noise so that the decision counter "ticks over" toward a decision even when the stimulus provides no perceptual information. Random noise is needed to allow the system to respond even when there is little or no perceptual information from the stimulus.

The key aspect of the model that allows it to explain priming is that counters can become attractors. At the time of test, prior exposure to a word causes the word's counter to attract a few counts more than it otherwise would, stealing them away from the counters of other similar words. It is assumed that this attractive force is quite weak and that its influence extends only through the space of the word's similar "neighbors," its "cohort," and not to faraway, dissimilar words. Also, just as a counter corresponding to a studied word can attract counts away from counters for similar words, so counters for the words in a cohort can attract counts away from similar cohorts. For visually presented words, it is assumed that similarity is defined in terms of abstract visual features.

When a target word is flashed, the accumulation of counts in the decision system begins. If the flash time is long enough, most of the counts are determined by the stimulus, but if the flash time is extremely short, there may be very few counts determined by the stimulus. In typical perceptual-identification experiments, the flash durations are short relative to the time taken to reach a decision and consequently the counters continue accumulating counts after presentation of the stimulus has terminated. Accumulation of the counts into counters continues until the total number of counts in one of the counters exceeds the maximum of the others by a criterial amount, *k.* This multicounter model can be seen as a generalization of the random walk process (Laming, 1968; Link, 1975; Link & Heath, 1975; Ratcliff, 1978, 1988; Stone, 1960) to multiple responses, and it differs from the logogen model, which uses an absolute criterion and so is a member of the accumulator class of models (Luce, 1986). The multiple counter mechanism has considerable similarity to a winner-take-all connectionist decision mechanism (Feldman & Ballard, 1982).

Figure 3 illustrates the processes of the counter model in the forced-choice, naming, and single-word decision tasks. In the forced-choice task, the flashed target word is immediately followed by the two alternative choices. As soon as the two alternatives are available to the system, the decision process is restricted to those two words by means of restricting the accumulation of counts to the counters for those two words. Every count is accumulated by one or the other of these two counters. Some of the counts correspond to features that are a part of both words, for example the features that are a part of the letter

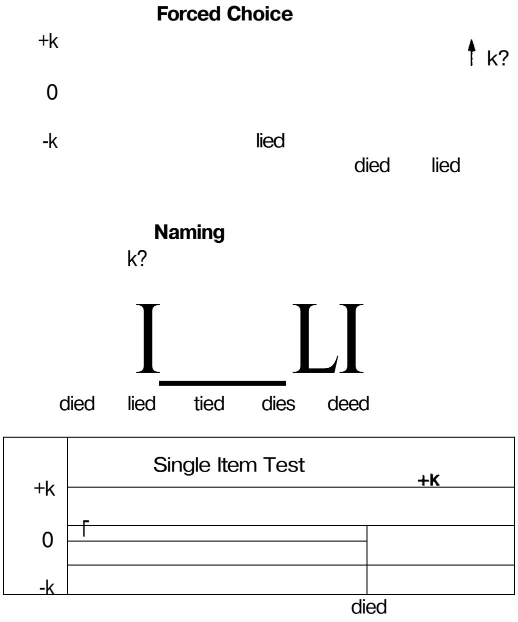

Figure 3. An illustration of the counter model for forced-choice, naming, and the "yes-no" single test word tasks. For forced choice and naming, *k* is the number of counts by which one counter must exceed the maximum of the others for a response to be made. For the single-item, "yes-no" task, *k* positive counts need to be accumulated for a positive response or *k* negative counts for a negative response. The random walks illustrate the translation from the counter description to the random walk description of the forced-choice and single test word models.

i in *died* and *lied.* This kind of count is accumulated by the counter for the target word with probability 0.5 and by the counter for the other alternative with probability 0.5, if neither of the words was studied previously. Some of the counts are null counts (random noise), and these are also accumulated by one of the two counters, each with probability 0.5, if neither alternative was studied. A few counts correspond to features that are features only of the flashed target word and not the alternative word (e.g., features of the first *d* in *died),* and these counts are always accumulated by the target's counter. A count of the last kind is called "diagnostic" for the target. Diagnostic counts are much less likely if the two alternatives are highly similar to each other (e.g., *died lied)* than if they are highly dissimilar (e.g., *died sofa*). The probability that a count is of this diagnostic sort is a parameter of the model. For similar alternatives, the probability of a diagnostic count is ps, which in applications to data was less than .1. For dissimilar alternatives, the probability is pd, which must be greater than ps and in applications was $1.9ps$. Overall, the probability that a count is accumulated into the target word's counter when the alternatives are similar and neither was studied is $p = ps + 0.5(1 - ps)$. When the alternatives are dissimilar, the probability is $p = pd + 0.5(1 - pd)$. (It should be noted that ps is small, always smaller than .10, so that multiplication by 1.9 to produce pd does not result in a value of pd greater than 1; this is also true for the other parameters such as pc provided later.) Accumulation of counts continues until the total in one of the counters exceeds the total in the other by the criterial amount, k, which was set to 10 for the data examined later.

Prior study of a word causes the counter for that word to steal nondiagnostic counts away from the counters of other words similar to it. In the forced-choice task, this means stealing counts away from a similar alternative that was not studied. Instead of the two counters for the alternative choices distributing nondiagnostic counts evenly between them (i.e., with probability .5), nondiagnostic counts are more likely to be taken by the counter for the previously studied word. In applying the model to data, the value .50 was increased to .51 for the previously studied alternative when the two alternatives were similar (the size of this increase could vary as a function of the similarity of the two alternatives, but this was not done in the implementations presented here). This theft of nondiagnostic counts yields bias: an increased probability of a correct response if the flashed target was previously studied and a decreased probability of a correct response if the similar alternative was studied. Because attraction of nondiagnostic counts by the previously studied word extends only through the word's similar neighbors, bias occurs only for similar alternatives and not dissimilar ones. Because attraction applies to all nondiagnostic counts including null counts (random noise), bias occurs even when the flash duration is so short that overall performance approaches chance. Null counts allow the system to make a response (as subjects in experiments must) even when flash time is so short that little stimulus information is available.[1]

In the naming task, there are no forced-choice alternatives to restrict the decision process. Null counts are randomly accumulated to counters, but the probability that one of them is accumulated to the target word's counter instead of any one of the other counters in the system is extremely small. Of those counts that are not null, some of them correspond to features that distinguish the target from words similar to it (e.g., the features of the *c* in *cage* distinguish *cage* from *rage, page,* and *wage*), and these "target diagnostic" counts are taken by the target's counter. The probability that a count is of this target diagnostic sort is the same as for the forced-choice task, ps. Other counts, "cohort diagnostic" counts, correspond to features that distinguish the target and words similar to it from all the other words in the lexicon (e.g., the features of *a, g,* and *e* distinguish *cage* and the words similar to it from other words like *whom).* There are many more of these counts than there are target diagnostic counts, so the probability that a count is cohort diagnostic, pc, is greater than the probability that a count is target diagnostic. In applications to data, $pc = 4.0\,ps$. All words similar to the target, that is all words in the target's cohort (including the target), have an equal chance of accumulating a cohort diagnostic count, if none of them was studied previously. The probability of a word taking the count, where n is the number of words in the cohort, is $(pc - ps)ln$ (the target diagnostic counts are also cohort diagnostic, so their probability is subtracted from pc). Therefore, the probability that a count is accumulated to the target's counter is the probability that the count is target diagnostic plus the probability that it is cohort diagnostic and accumulated by the target, $p = ps + (pc - ps)ln$, as shown in Table 8.

In forced choice, all counts are accumulated to one of two counters, and sufficient counts are accumulated such that the total count in one exceeds the other by the criterial amount (see Feller, 1968). However, in naming, counts are divided up among all the words of the system. When the evidence from the stimulus is weak, it may be that even after thousands of counts, the total in no one of the counters exceeds the maximum of the others by the criterial amount. The stopping rule that was adopted for naming was based on the absolute number of counts that had been accumulated. The idea was that, if very few counts had been accumulated in any one counter, then it should be clear that no response was forthcoming and processing should terminate (by the subjects saying "no" in the experiments). In simulations of the model, the rule was: When 30 counts have entered the system, terminate if the number of counts in the counter with the maximum number of counts is less than k; when 100 counts have come in, terminate if the number of counts in the counter with the maximum number of counts is less than $4k$; when 300 counts have come in, terminate if the number of counts in the counter with the maximum number of counts is

---

[1] The assumption that counts are accumulated at a constant rate over the course of decision making is a strong assumption. It may be that, with brief masked stimulus presentation, perceptual information is available only briefly. The counter model can be modified so that early counts are based mainly on stimulus information and later counts more likely to be null ones. We tried such a modification by setting a number of counts m, before which ps was a constant value (e.g., .1) and after which the probability ps was set to zero. This modification produced quantitatively similar results to the original model. A more realistic model would add parameters to capture a rapid rise in stimulus information (ps) followed by a constant rate of information followed by a gradual decay (e.g., activation values in the interactive activation model, McClelland & Rumelhart, 1981).

Table 8
Expressions for Naming Probability

Study word	Flashed word	Value of pc	Probability that a feature is accumulated in the counter of the flashed word
Neither studied	*died*	$4.0ps$	$ps + (pc - ps)/n$
died	*died*	$4.8ps$	$ps + 0.1\,(pc - ps) + 0.9\,(pc - ps)/n$
lied	*died*	$4.8ps$	$ps + 0.9\,(pc - ps)/n$
Neither studied	*lied*	$4.0ps$	$(pc - ps)/n$
died	*lied*	$4.8ps$	$0.9\,(pc - ps)/n$

Note. pc = cohort diagnostic count probability; ps = target diagnostic probability; n = number of words in the cohort.

less than $10k$. For the applications to data provided later, k was 3. In a more complete version of the model, the stopping rule would be more continuous to allow the possibility of stopping without a response at more than just three positions in the course of count accumulation.

In naming, as in forced choice, prior study leads to theft. The cohort of words similar to a studied word steals counts that would have gone to other words in the system, and the studied word itself steals counts that would have gone to other words in the cohort. The probability, pc, that a count is taken by one of the words in the studied word's cohort (instead of some word outside the cohort) increases in applications to data from $4.0ps$ to $4.8ps$. In addition, the words in the cohort do not have an equal chance at the count, as they would without prior study; instead the previously studied word takes the count with a higher probability than the other words. In applications, the probability was .10 that the count would be accumulated by the previously studied word and .90 split among all the words in the cohort (including the target). Table 8 shows the total probabilities of a count being accumulated by previously studied targets and by previously studied other words in the cohort. For a flashed target word that was previously studied, for example, the probability that a count is accumulated in the word's counter increases as a result of prior study by $0.1(pc - ps)$, which yields bias toward that word.

The third task to be modeled is the "yes–no" decision task, the task in which the flashed stimulus is followed by a single word for which the subject must decide whether or not it was the same as the flashed word. In forced choice, the flashed word is followed by two alternative choices, and the accumulation of counts is restricted to their two counters. In the "yes-no" task, the accumulation of counts is restricted to the counter for the single word that is presented for a decision. For the forced-choice and naming tasks, evidence is either for a word or for some other word. When the decision process is restricted to only a single word, then the evidence is either for the word or against it. In the situation where the flashed word and the word presented for decision are the same, then counts determined by features from the flashed word are accumulated into the decision word's counter. If the flashed and decision words are not the same, then features from the flashed word are evidence against the decision word and counts from these features cause the word's counter to lose counts. Null counts are taken as positive or negative evidence randomly, if the decision word was not previously studied.

The probability that a count increments the counter for the decision word is q, and the probability that a count decrements the counter is $p = (1 - q)$. For mathematical tractability, q is the total probability, combining the probability of a null count (random noise) incrementing the counter with the probability that a count comes from the flashed word and matches the decision word. A decision is reached either when the total number of counts in the decision word's counter increases to k greater than the starting point (a "yes" response) or decreases to k less than the starting point (a "no" response; see Figure 3). The single counter is a single random walk process (Feller, 1968), and the probability of a "yes" response is

$$\Pr(\text{yes}) = [(q/p)^{2k} - (q/p)^{k}]/[(q/p)^{2k} - 1]).$$

In the naming and forced-choice tasks, prior study leads to a counter taking a small proportion of null counts (random noise) and nondiagnostic counts as positive evidence for itself. Prior study has the same effect in the "yes–no" task. However, in forced choice and naming, the null counts are stolen away from other counters. In the "yes-no" task, the decision word's counter already accumulates all the counts (either positively or negatively). So all that changes is an increase in the probability that it takes null counts as positive evidence.

In psychological terms, the priming effects that have been so important to the development of the implicit memory enterprise are conceptualized quite differently by this model than by previous interpretations. Priming and bias are the tendency of a counter to attract counts as a result of prior study. Prior study of a word causes that word to accumulate more information favoring itself than otherwise would be the case. In the forced-choice and naming tasks, this occurs at the expense of other similar words. The conceptualization of the effects of prior study offered by the counter model is not one of the processes by which a new representation about a word is formed but rather one of alteration of standard information processes. Bias is a small alteration in the standard mechanisms of word identification. The main ways the model differs from Morton's (1969, 1970) logogen model are the assumption that prior study makes a counter attract counts from neighboring counters with slightly greater probability than it otherwise would and the assumption that the response criterion is relative to counts in other counters as opposed to an absolute criterion.

The counter model could be described using a different metaphor than the theft of counts, for example, a "greased pathways" metaphor. Study of a word would lead to small alterations in the processing of subsequent perceptual information so that there was a tendency to follow the same pathways as for prior processing. We prefer the theft metaphor because it conveys the notion of the counter for a previously studied word capturing nondiagnostic counts from similar neighbors and because it is a new metaphor with its meaning unclouded by past usage.

Applications of the Counter Model to Data

For priming in the forced-choice task, the main phenomena to model are the bias effect due to prior study, the fact that the effect appears for similar but not dissimilar alternatives, and the fact that it is not decreased in size as overall performance approaches chance. For naming, the phenomenon to model is the increase in response probability due to prior study. For the "yes-no" task, the phenomenon is that the probability of a "yes" response to the decision word is increased by prior study of that word but not decreased by prior study of a similar word.

Figure 4 shows the probability of a correct response as a function of flash time for the conditions of Experiment 3 (see Table 3) in which the forced-choice alternatives were similar to each other. Prior study of a word increases the probability that a nondiagnostic count is accumulated by that word's counter from .50 to .51 (this increase is one parameter of me model, the parameter that gives bias from prior study). Decreasing flash time decreases the probability *ps* that a count is diagnostic of the target (i.e., the probability that a count is null increases). To produce the model predictions across flash times, *ps* was varied over a range from .0 to .083, with a decrease of .0167 in *ps* equivalent to a decrease of 9 ms in flash time. As can be seen in the figure, the model predictions provide good fits to the data. With only the one parameter *ps* varying, the model captures the effect of bias across the four flash durations and, in particular, the slight increase in the amount of bias as the flash duration decreases.

Figure 5 (data) and Figure 6 (model predictions) show how well the model fits data for an experiment with both similar and dissimilar forced-choice alternatives (Experiment 4; Table 4). With dissimilar alternatives, there were no differences across study condition (see Table 4), so the probabilities were collapsed into one line in the figure. With similar alternatives, as with Experiment 3 (Figure 4), prior study of a word increased the probability that a nondiagnostic count would be accumulated by that word's counter from .50 to .51 (yielding bias from prior study), and *ps* varied across flash times (a decrease of .0167 in *ps* was equivalent to a decrease of 7 ms in flash time). Performance on dissimilar alternatives depends on *pd*, the probability of a diagnostic count when the alternatives are dissimilar. The probability *pd* must be greater than *ps*, and it was set to 1.9*ps*. The figures show good fits of the model to the data, with bias increasing as flash time decreased and also with accuracy

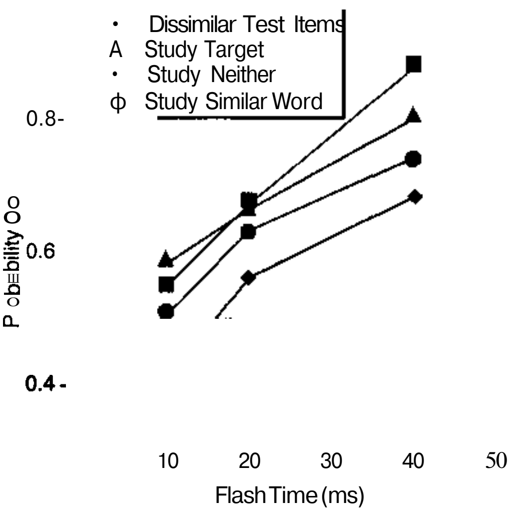

Figure 5. Experimental data from Experiment 5. The experimental conditions, with similar forced-choice alternatives, are (a) the octagon is baseline (no study), (b) the triangle is the condition in which the target was studied earlier, and (c) the diamond is the condition in which a word similar to the target was studied. The square combines all the conditions in which the forced-choice alternatives were dissimilar.

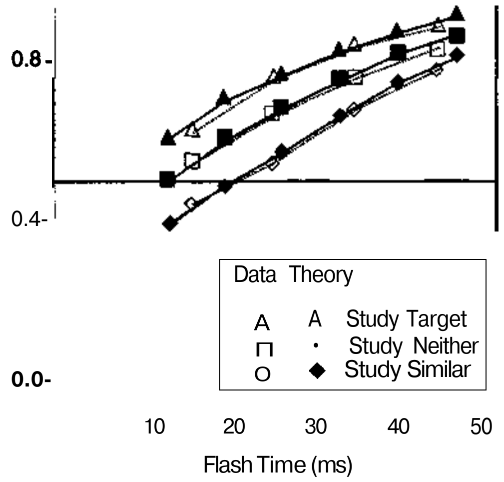

Figure 4. Experimental data and model predictions for probability correct in forced choice as a function of flash time in Experiment 4. The predictions are the solid lines, and the dotted lines are the experimental data. Square symbols are the baseline conditions, triangles are conditions in which the target was studied, and diamonds are conditions in which a word similar to the target was studied.

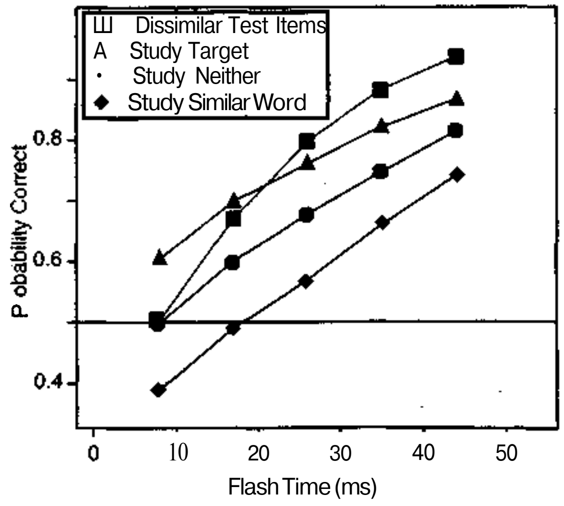

Figure 6. Predictions of the counter model for the data from Experiment 5. The conditions, with similar forced-choice alternatives, are (a) the octagon is baseline (no study), (b) the triangle is the condition in which the target was studied earlier, and (c) the diamond is the condition in which a word similar to the target was studied. The square combines all the conditions in which the forced-choice alternatives were dissimilar.

decreasing at a greater rate with flash time for dissimilar than similar alternatives.

In Experiments 6 and 7, the forced-choice task with similar alternatives was mixed with the naming task, so that the same subjects performed both tasks. In addition, flash duration was varied for the forced-choice task. Only one parameter in the model, *ps*, can vary to accommodate the data from all of these conditions. It must account for all the data simultaneously and it must account for individual differences in performance.

Table 9 shows a range of values of *ps* from .017 to .083. A value .042 produces a good fit to the averages across subjects of the naming data from Experiment 6 (the data shown in Table 6; note also that a slightly smaller value of *ps* would capture the data from Experiment 7 where there was a 30-min delay between study and test). The absolute values of the probabilities generated with *ps* = .042, and their values relative to each other are about the same as in the data. Prior study of the target increases the probability of the target being given as a correct response, and prior study of a word similar to the target increases the probability of that word being given as an incorrect response. This probability of the studied word being given as an incorrect response is so low because most of the evidence being accumulated is evidence for the word that was actually flashed, and the attraction toward the word that was studied is very small.

With the *ps* value of .042, the model captures the average performance across subjects for naming. In reality, *ps* should vary across subjects (i.e., subjects vary in how well they can "see" the target word). However, for a given subject, the same value of *ps* must apply to both naming and forced choice (as long as the flash time is the same for the two tasks). A subject who cannot see the target word very well for forced choice cannot see it very well for naming either. Figure 2, top panel, shows the probabilities of correct responses for naming and forced choice for individual subjects. The *xs* mark the no-study conditions for the individual subjects, and the heads of the arrows mark the priming from prior study. The circles mark the predictions of the model for the no-study conditions, and the heads of those arrows mark the predictions for priming. With only *ps* varying as probability correct increases, the model provides a good account of the data. The data from the subjects show about the same trends as the theoretical predictions; the arrows point up when naming performance is near zero (indicating a large bias in forced choice) and point to the right when forced-choice performance is near ceiling (indicating a large effect in naming). The bottom panel collapses the data from the individual subjects into thirds (based on naming performance) and shows a remarkably accurate prediction of the theory: The theoretical predictions and data lie almost on top of each other. The intuition captured by the negatively accelerated function in the model is that getting only a few features of the stimulus might be enough to allow an accurate forced-choice decision on some trials while producing a large bias effect, but obtaining a only a few features of a word does not provide enough information to name it. In the model, this is because there are not enough counts to exceed the stopping criterion. When *ps* increases, extra features help naming and forced choice more equally.

It is noteworthy that the sizes of the bias effects in the model are similar for the naming and forced-choice tasks. In forced choice, the probability that a nondiagnostic count is accumulated by a word's counter is increased by .01 if the word was previously studied. In naming, the equivalent increase in probability is $.1(pc - ps)$, which for *ps* = .05 was .015. This is in the same range as the value .01 for forced choice. In both cases, these changes in probability may seem quite small, but they work through an iterative and additive process with the result that over the number of counts required to reach the response criterion, their effects cumulate to produce effects of the size of those observed in the data.

In the "yes-no" task, the accumulation of counts is restricted to the single counter for the word presented for decision. The probability of a "yes" response is determined by the parameter *q*, the probability that a count increments the counter. Without prior study of the decision word, *q* depends on how the decision word matches the flashed word: If the two words are the same, the counts determined by the flashed word increment the counter; if the two words are different but similar, then the counts determined by the flashed word frequently but not always increment the counter; and if the two words are dissimilar, the counts determined by the flashed word rarely increment the counter. For the data from Experiment 8 (Table 7), the *q* values for these three different conditions were .53, .50, and .45, with *k* set to 7, producing probabilities of "yes" responses of .70,

Table 9
Predicted Probability of Naming Responses

	Target responses			Similar word responses		
Parameter *ps*	Study target	Study similar word	Study neither	Study target	Study similar word	Study neither
.017	.05	.02	.02	0	0	0
.033	.27	.15	.15	0	.02	0
.042	.43	.24	.24	0	.03	0
.050	.58	.35	.36	0	.05	.01
.067	.82	.57	.58	.01	.07	.01
.075	.88	.68	.67	.01	.08	.01
.083	.93	.75	.77	0	.09	.01

Note. *ps* = target diagnostic count probability.

.50, and .20 in the three conditions of decision and test words matching, similar, and dissimilar, respectively. Prior study of the decision word increases q because null counts (random noise) are more likely taken as positive evidence. The increase in q was .01, the same amount of bias as for the forced choice and naming tasks. With this increase in q, the predicted probabilities of "yes" responses were .75, .57, and .24, in close agreement with the data.

The differences in q reflect the differences in the probability that a count is diagnostic for the decision word; q increases as the similarity of the flashed target and the decision word increases. The increase in q from the target being similar to the decision word to the target being the same as the decision word was .03, and the increase from the target being dissimilar to it being the same as the flashed word was .08. These increases were much the same as in the forced-choice task. The flash time for the "yes-no" task in Experiment 8 was 25 ms. To predict the data for this flash time for forced choice (Experiment 3, Figure 4), ps, the probability that a count is diagnostic between the flashed target and a similar word was .033, and pd, the probability that a count is diagnostic between the target and a dissimilar word was .063. The similarity in the parameter values across tasks is strong support for the counter model.

Overall, the counter model fits the data very well. With only one parameter varying, the model accounts for performance on naming and forced choice across flash times at the level of the individual subject. About the same values of the parameters predict performance in all three tasks, naming, forced-choice, and the "yes-no" decision task. The next step was to extend the model to two variables that have figured prominently in the word-identification literature, the frequency of a word and the number of similar neighbors the word has in the lexicon.

The Counter Model and Word Frequency

In the counter model, bias effects in perceptual identification come about because prior study of a word causes its counter and the counters of similar words to attract counts a little more strongly than they otherwise would. In all other respects, a word is processed by the same representations and mechanisms as if it had not been previously studied. The counter model was developed from the logogen model, and thus it should apply to the same word-identification effects as that model.

One of the primary effects driving research on word identification historically has been the effect of word frequency (e.g., Broadbent, 1967; Morton, 1968): High-frequency words are easier to identify than low-frequency words. The logogen model was specifically designed to deal with word frequency by assuming that the criterion for a response was lowered for high-frequency words relative to low-frequency words (Morton, 1969, 1970, 1979). Broadbent (1967) and Morton (1968) provided evidence supporting this assumption. For example, Broadbent used auditory presentation of high- and low-frequency words in noise and found that a high-frequency word was often produced by mistake in response to a similar low-frequency target, whereas a low-frequency word was rarely produced in response to a similar high-frequency target (see the discussion in Morton, 1970, p. 208). McClelland and Rumelhart's (1981) interactive activation model explains the word-frequency effect in much the same way as the logogen model; differences in frequency are represented by differences in the resting level of activation of word nodes.

The counter model also offers the same account as Morton's (1970). Word-frequency effects are modeled as variations in the resting levels of the counters in the system. For the applications of the model to the bias effects that were described above, the resting levels of all counters were assumed to be zero. The qualitative patterns of bias effects are not altered when the resting levels are set to reflect word frequency.

An important problem with the logogen model is the assumption that word-frequency effects and repetition (priming) effects are mediated by the same mechanism. Word frequency is a function of the resting level of the logogen, whereas repetition effects are modeled by reducing the response criterion for the logogen. Because a single presentation of a word lowers the criterion and because the empirical effect lasts over 24 hr, Jacoby and Witherspoon (1982) argued that the model must predict that the thresholds of all logogens should be permanently lowered and consequently frequency advantages for high-frequency words should disappear. In the counter model, repetition effects are decoupled from word-frequency effects such that study does not affect the components of the model that deal with word frequency.

With the assumption that resting levels are set to reflect relative word frequencies, the counter model predicts that overall accuracy in forced choice is not a function of word frequency when the two forced-choice alternatives have equivalent frequencies. In Experiments 3, 4, 6, and 7, responses were coded into two frequency conditions: a condition in which the flashed target and the other word presented as an alternative were both high-frequency words and a condition in which they were both low-frequency words. In the first condition, the higher resting level due to high frequency is the same for the two counters, and neither should get an advantage over the other because of frequency. Similarly, in the second condition, the resting levels of the two words are about the same, so again neither should get an advantage over the other. Averaging over all the other conditions in the four experiments (study condition, flash duration, similar versus dissimilar alternatives), the probability correct when the flashed target was a high-frequency word was .701 and the probability correct for a low-frequency word was .696. Just as the model predicts, overall accuracy was about the same for the high- and low-frequency targets.

The logogen model incorrectly predicts that a forced choice between high-frequency words is less accurate than a forced choice between low-frequency words. This is because the resting levels of high-frequency words are nearer their absolute response criterion, and only a few counts misplaced in the counter of a high-frequency word can cause that counter to exceed its threshold in error. For example, consider the counters for two high-frequency words: one of them, Counter 1, the correct response, and both counters one count below their thresholds. If the probability of getting a count in Counter 1 is .7 and the probability of getting a count in Counter 2 is .3, then Counter 2 exceeds its threshold in error with probability .3. In contrast, consider two counters for low frequency words, both 20 counts from their thresholds. With the probability of getting a count in Counter 1, the correct response, equal to .7 and the probability of getting a count in Counter 2 equal to .3, an error could be made only after 20 counts in Counter 2; the probability of this is near zero. Thus the probability of an error is higher for high-frequency pairs than for low-frequency pairs. The reason the counter model makes a correct prediction and the logogen model does not is the difference in their decision rules: The counter model uses a relative response criterion (one has to beat the other by some fixed amount), whereas the logogen uses an absolute response criterion.

Both the counter and the logogen models predict that when the forced-choice alternatives are a high- and a low-frequency word, there should be a bias to select the high-frequency word because its resting level is higher than for the low-frequency word. For the no-study conditions from Experiments 3, 4, 6, and 7, when the high-frequency word was flashed, the probability of correctly choosing it over the low-frequency alternative was .727, whereas when the low-frequency word was flashed, the probability of correctly choosing it over the high-frequency alternative was .613. This is the bias predicted by the model (and by Broadbent, 1967, and Morton, 1968).

For naming, the counter model predicts that high-frequency words have an advantage over low-frequency words. In naming, if no counter exceeds the maximum of the others by the criterial amount *k* before the absolute number of counts exceeds one of the three stopping criteria, then processing terminates without identification of the target. If the target is a high-frequency word and therefore has a higher resting level of counts, then it is less likely that the absolute number of counts exceeds the stopping value before the target is identified. Averaging over all the other conditions in the two experiments with naming (Experiments 6 and 7), the probability of a correct response was .350 for high-frequency words and .288 for low-frequency words. Priming is also predicted to be somewhat larger for low-frequency words than for high-frequency words. Data relevant to this prediction are presented with Experiments 9 and 10 below.

The Counter Model and Neighborhoods: Experiments 9 and 10

The success of the counter model in explaining bias effects depends on the assumption that counts for visually similar words are distributed across their counters. If there are more words in a neighborhood (see Andrews, 1992; Coltheart, Davelaar, Jonasson, & Besner, 1977; and Goldinger, Luce, & Pisoni, 1989, for a discussion of neighborhood effects), then there are more counters to share the counts, and therefore the less likely it is that a count is accumulated to any one counter. Thus, the model predicts that the probability of successfully naming a word is inversely related to the number of words in its neighborhood.

We tested this prediction in Experiment 9, using the naming procedure from Experiment 6. For each of the words of the triples used in the preceding experiments, we approximated its number of visually similar neighbors by finding the number of other words that differed from it by one letter in the Kucera and Francis (1967) word norms.

Experiment 10 was the same as Experiment 9 except that the flash time was increased from 30 ms to 65 ms. This manipulation was motivated by a discrepancy between the results of Experiments 6 and 7 and the results of an experiment by Rueckl (1990). In his experiment, prior study of a word similar to the target increased the probability that the target would be named correctly (and the increase was large, 10-15%). Experiments 6 and 7 did not show this effect; only prior study of the target itself increased the probability with which it would be named. A possibly important difference in Rueckl's experiment was that the probability of correct responses in the no-study baseline condition was .5 to .6, whereas it was only .2 to .3 in Experiments 6 and 7. Increasing the flash time to 65 ms was expected to increase the base level of performance, perhaps leading to a replication of Rueckl's results.

Method

The experiments began with the practice block, and then there were two study-test blocks, each with 32 words to study and 64 test items. The flash time was 30 ms for all targets in Experiment 9 and 65 ms in Experiment 10. In Experiment 9, 28 subjects were tested, but 8 were eliminated because they failed to identify any target. In Experiment 10, there were 16 subjects.

In the experiment, 128 triples were used. Each word was coded both by its frequency (Kucera & Francis, 1967) and its number of neighbors. There were three study conditions: The flashed target word had been presented in the study list, the word similar to it had been studied, or neither had been studied. All test items required naming responses. Triples were assigned randomly to the study conditions, and therefore levels of word frequency and number of neighbors were assigned to the study conditions randomly. The naming procedure was the same as in the earlier experiments. There were 64 test items in the no-study condition and 32 in the other conditions.

In Rueckl's experiment, subjects were required to generate a response. They could reasonably be expected do this even with very little information from the stimulus because each word was exactly four letters long. Word lengths varied in our materials such that a response could not be required in Experiment 10, but subjects were encouraged to guess a response if they were not sure (but not to guess when they had no information at all from the flashed item).

Results and Discussion

The data, shown in Tables 10 and 11, are similar for the two experiments. As in Experiments 6 and 7, prior study of the target increased the probability of a correct response, and prior study of a similar word increased the probability with which that word was given as an incorrect response. The probability of a correct response to the target was greater for high-frequency words than for low-frequency words. Most importantly, the prediction that the probability of a correct response to the target would decrease as its number of neighbors increased was confirmed.

Table 10
Response Probabilities in Naming Averaged Over Word Frequency and Number of Neighbors, Experiments 9 and 10

Flash time (ms)	Response	Study target	Study similar word	Study neither
30	Target word	.43	.26	.22
30	Similar word	.01	.08	.01
65	Target word	.89	.71	.73
65	Similar word	.00	.06	.02

Table 11
Response Probabilities in Naming Averaged Over Study Conditions, Experiments 9 and 10

Flash time (ms)	Response (word)	Frequency of target word	Number of neighbors 1-4	5-8	9 and over
30	Target	Low	.34	.21	.16
30	Target	High	.37	.33	.30
30	Similar	Low	.06	.02	.02
30	Similar	High	.04	.01	.01
65	Target	Low	.75	.65	.61
65	Target	High	.89	.84	.77
65	Similar	Low	.03	.05	.05
65	Similar	High	.01	.02	.00

ANOVA for correct responses to the targets showed the main effects of study condition, frequency, and number of neighbors significant, $F(2, 38) = 26.40$, $F(1, 19) = 11.87$, and $F(2, 38) = 10.20$, respectively, for Experiment 9 and $F(2, 30) = 6.39$, $F(1, 15) = 3.59$, and $F(2, 30) = 10.51$, respectively, for Experiment 10. There were no significant interactions among the three variables, all $Fs < 1.90$.

As predicted by the counter model, the increase in the probability of correct responses to the target due to prior study was larger for low-frequency words than high-frequency words. To gain sufficient data, Experiments 6 and 9 were combined (the amount of priming in Experiment 7 was too small to test for interactions of word frequency and priming). For low-frequency words, the baseline, no study, probability was .20, increasing to .40 with prior study. For high-frequency words, the baseline was .30, increasing to .43 with prior study. For Experiment 10, the corresponding probabilities for low-frequency words were .63 and .86, and for high-frequency words, .85 and .89.

Experiment 10 failed to replicate Rueckl's (1990) result: The probability of correctly reporting a target word was not significantly increased over baseline when a word similar to the target had been studied; in fact it decreased slightly. Apparently, simply increasing the baseline level of correct responses does not bring about the effect he found. Inspection of the results of Experiments 6, 7, and 9 and those in Ratcliff et al. (1989) also showed no evidence of the effect. It might be that the differences in results are due to differences in display equipment. Our apparatus used a fast oscilloscope with 1 ms time resolution and a fast phosphor. The Apple II system used by Rueckl had a slow-display phosphor, and it also used an interlaced display (adjacent lines were displayed on successive sweeps of the tube beam). Also, the mask characters in his experiment were the same size as the stimulus characters, and this might allow portions of the stimuli to be viewed after the mask was displayed (see the effects of different masks in Experiment 2 above). The most likely possibility may be that Rueckl's effect was brought about by his requirement that subjects produce a response. This requirement would lead subjects to guess, and they might base their guesses on words they had studied. This would lead to an increased probability of giving words similar to the target as responses as well as an increased probability of giving the target as a response. Guessing based on explicit memory for studied words would likely be a slower process than the usual processes of naming. An analysis of Experiments 6, 7, 9, and 10 in which the slowest responses were eliminated showed no change in the pattern of results indicating that no slow, special strategy was being used. A similar analysis for Rueckl's data might show the locus of the differences between his results and ours.

Despite the lack of a clear understanding of the reason for the difference in the results of our experiments and Rueckl's (1990), the difference is not significant for the counter model. Whether or not the probability of a correct response to a target is increased when a similar item is studied is not a strong test of the counter model. The model can accommodate a moderate range of differences in response probability on the target word as a function of study of the similar word by varying the amount of attraction of counts into the cohort. But the model does make the strong prediction that there are intrusion errors, as found in all our experiments; intrusion probabilities were not reported in Rueckl. In contrast, an increase in response probability for a target word as a function of study of a similar word is predicted by Rueckl's theoretical analysis and so our data provides problems for that analysis.

The Counter Model and Encoding-Specific Features of Stimuli

In the counter model, the features that determine which counters accumulate counts are abstract entities, but it is assumed that they are tied to the perceptual features of a stimulus. Exactly how they are tied to perceptual features is beyond the scope of the model. It might be possible to investigate amount of similarity as a predictor of amount of bias, using methods like those developed by Tversky (1977) to produce representations of similarity among words and thus a theoretical framework for understanding similarity effects. However, the counter model was developed to explain the decision mechanisms for naming, forced-choice, and the "yes–no" task, not the ways visual features translate information into counts.

Nevertheless, some aspects of representation can be fit into the framework of the counter model. One example is modality information. Switching modality between study and test (auditory study to visual test) reduces the amount of priming to zero or to a small and usually barely significant value (Clarke & Morton, 1983; Jacoby & Dallas, 1981; Winnick & Daniel, 1970). This and related results led Morton (1979) to propose a revised logogen model in which there are two sets of input logogens, each set specific to one modality, visual or auditory. The connectionist models also separate visual and auditory processes, although the two sets of processes interact with each other (e.g., Seidenberg & McClelland, 1989). The counter model can adopt the same kind of solution and assume that there are separate sets of counters, one for auditory information and one for visual information. Neighborhoods (similarity cohorts) would be assumed to depend on similarity defined with respect to modality (e.g., Goldinger et al., 1989; Ratcliff, Allbritton, & McKoon, 1997). Some transfer between the two sets of counters might occur, perhaps by rehearsal or by imaging a word or by internal mechanisms that link together the decision units from different modalities.

Within the visual modality, study to test changes in the physical form of the stimulus have small or nonexistent effects on performance (e.g., Rajaram & Roediger, 1993). For example, Winnick and Daniel (1970; see also Clarke & Morton, 1983) found equivalent amounts of priming from study of handwritten words versus typed words to tests of typed words, and Jacoby and Hayman (1989) found little effect of case change on priming (except for high-frequency words tested in lower case). In terms of the counter model, the fact that changes in visual form have no effect or only a very small effect suggests that the decision counters work with relatively abstract visual information. Jacoby and Hayman did obtain larger effects with unusual fonts (see also, Graf & Ryan, 1990), but these might involve special, slow decoding processes that would be additional to the mechanisms of the counter model. Certainly, the effects of physical form are not strong and pervasive like the effects of repetition.

The Counter Model: Summary

The counter model is designed to explain the identification of words under the suboptimal conditions of the typical perceptual identification experiment in which flash times are too fast to allow perfect performance. The mechanisms of the model account for the effects of several salient variables. Prior exposure to a word facilitates its identification because, in the decision process, evidence is more likely to be taken as favorable to that word than it otherwise would be. However, in forced choice and naming, this results in evidence being taken away from competitors, and they are put at a disadvantage. The decision counters accumulate evidence at a constant rate, whether it is "good" evidence from the perceptual stimulus or null counts (random noise), so responses can be generated even when the overall level of performance is near chance.

The main reason that other existing models cannot explain priming effects is their assumption that prior exposure to a word changes some property of the representation of the word itself. For example, the resting level of activation for the word might be changed or, in a connectionist model, weights on connections involving the word might be changed. When a property of the word itself changes, then processing of the word should always show facilitation relative to processing of other words. The fact that facilitation is observed only relative to other similar words but not relative to other dissimilar words shows that prior exposure cannot simply change a property of a word independently of all other words.

This same argument applies to models that might be developed to attempt to explain priming by assuming that prior study leads to a new representation of the stimulus. If a new representation produces a facilitation in processing, it should do so when the forced-choice alternatives are dissimilar as well as similar. Also, when the amount of perceptual information is reduced to near zero by use of a short flash time, there will be no perceptual features to contact the new representation and so there should be no bias, contrary to the data.

Any model that attempts to explain priming with a lateral inhibition mechanism also faces problems. Presentation of a previously studied item leads to facilitation of the tested item (on-center) and inhibition of near neighbors (off-surround) relative to baseline. This leads to facilitation of the previously studied item over all other items in forced choice, even those that are dissimilar, thus predicting a d' effect for dissimilar test pairs in forced choice.

Whether other models could be changed to include a decision process like that of the counter model is an open question. As pointed out earlier, the counter model is most compatible with the logogen model, replacing that model's decision process with a more interactive decision process that accumulates features competitively. It might be possible to add something like the counter model's decision process on to the end of the interactive activation model, but such a radical change would require completely new tests of that model against all the phenomena it now explains. It is more difficult to see how the Seidenberg and McClelland (1989) model could be modified to be like the counter model, because representations of words are distributed across a whole network of nodes. This does not mean that connectionist models could not be developed to account for these phenomena. In fact, one message from this article is that models designed to deal with word identification should include repetition priming as a target phenomenon for modeling.

It is a good bet that the developers of any successful model will have to place most of the consequences of prior study on mechanisms that have their major effect on retrieval and decision processes. The counter model's theft mechanism applies at retrieval, and this provides the ability to produce different effects depending on the kind of task or on the alternatives in forced choice.

As with any model, it is important to understand the scope of the counter model and the range of phenomena it is designed to explain. It is concerned with retrieval aspects of information processing and how to relate the naming, forced-choice, and "yes-no" tasks with bias effects in a traditional information-processing framework. However, the model is mute about visual variables such as the type of mask, and about letter-level constraints, nonword decisions, and so on, variables that are primary, for example, for the McClelland and Rumelhart (1981) model.

The counter model is consistent in its goals with a proposal made by Jacoby (Jacoby, 1983a, 1983b; Jacoby & Dallas, 1981) about how prior episodes affect perceptual and memory processes. He has suggested that presentation of an item leads to the creation of a representation of the item in memory. This representation is assumed to have multiple aspects, some related to meaning, some related to perceptual features, some related to associations, and so on. Performance on the item in a subsequent task depends on the degree to which the task taps different aspects of the memory representation. His proposal is in agreement with the standard information-processing framework of the counter model, that is, that memory is not made up of a number of independent systems. However, his specific proposal that prior episodes create new representations is at odds with the counter model's assumption that prior episodes change only the ability of counters to attract counts as positive evidence for themselves.

Implications of the Counter Model for Implicit Memory and Memory Systems

The hypothesis that there exist a number of different memory systems (see Schacter, 1994; Schacter & Tulving, 1994) was put forward to explain priming effects and how priming is independent of explicit memory. For priming in perceptual identification, the relevant system is the visual word-form system. Because of the focus of implicit memory theorists on the separation of systems responsible for priming from the episodic system responsible for conscious retrieval, there has been little development directed towards the mechanisms of word identification themselves or on how the mechanisms support priming. However, it is reasonable to ask what these mechanisms might be, especially in light of the success of the counter model.

One possibility is that the word-form system could incorporate the mechanisms of some other model, such as the interactive activation model (McClelland & Rumelhart, 1981). Exactly how this could be done is problematic at the outset because the two lines of research have proceeded completely independently; work on the interactive activation model has not been concerned with priming effects and implicit memory theories have not been concerned with specific processing models. Moreover, the success of the interactive activation model has come without the need for postulating separate memory systems. And, even if there could be some agreement about how a model like the interactive activation model could become an implicit memory system, priming effects would still not be explained, for the reasons mentioned above. That is, without substantial changes, the interactive activation model cannot account for the fact that bias applies to similar forced choice alternatives but not dissimilar ones.

Another possibility is suggested by the claim of implicit memory theorists that prior processing of a stimulus improves its perceptual identification. More specifically, Schacter (1990) has hypothesized that "visual processing of a word creates a representation of its particular visual features in the word form system" (p. 552). This hypothesis is contradicted by the data, in two main ways. It predicts an overall improvement in performance in forced choice rather than the bias effect that is actually observed. The problem is that the newly created representation for a word should help processing of the word but not hurt processing of its neighbors in forced choice. Schacter's proposal also predicts an improvement in performance independent of whether the other alternative in forced choice is similar or dissimilar. The problem here is that the newly created representation should help processing independently of what the alternative choice is.

A third, perhaps most obvious, possibility is that the word-form system incorporate the counter model. In doing so, it would incorporate a working model for perceptual identification, including priming effects. But, aside from attaching a new label to the counter model, nothing would be gained. Whatever proposals the implicit theory might make about perceptual identification would simply collapse into proposals about the mechanisms of the counter model: where they might be located in the brain (e.g., late stages of visual information processing for visual word identification), how they might be damaged, and why they might sometimes dissociate from other brain mechanisms. Furthermore, simply to label a processing system as the location of a memory system is at odds with the emphases of the memory systems approach. It is also at odds with the claims of discovery of new and functionally distinct memory systems.

Dissociations and Stochastic Independence

Research on implicit memory has directed attention towards dissociations between performance on implicit tasks and performance on tasks that require the conscious retrieval of episodic information. It has been shown that such dissociations do not force the postulation of separate and independent processing systems (e.g., Hintzman, 1990; Jacoby & Dallas, 1981; Jacoby & Witherspoon, 1982; Nosofsky, 1988; Ostergaard & Jernigan, 1993). It has also been argued that dissociations are to be expected from the perspective of standard and single memory information-processing views (Hintzman, 1990; Hintzman & Hartry, 1990; McKoon & Ratcliff, 1995; Ratcliff & McKoon, 1995; see also Schacter & Cooper, 1995). The counter model is an instantiation of part of such an information-processing system.

It is easy to see why, in the counter model, performance on perceptual identification is likely to dissociate from performance on an episodic task like recognition. In terms of the framework displayed in Figure 1, information relevant to recognition is found at a different level of the system than information relevant to identification. From this perspective, there is no apparent reason for a variable that affects the amount of bias in perceptual identification, a variable such as the visual similarity of two forced-choice alternatives, to affect recognition. Likewise, there is no reason for a variable that affects recognition, like semantic versus rhyming rehearsal, to affect perceptual decision counters. In general, variables might affect one of the tasks or both tasks, in the same or different ways; the only means of predicting variables' affects is through the mechanisms of specific models for the different levels of the system.

The counter model, in the context of a standard information-processing framework, is also compatible with findings of stochastic independence between performance on episodic tasks and amount of priming in perceptual identification. Like functional dissociations, stochastic independence has been taken as providing a major source of evidence for multiple memory systems. However, it too has been shown to provide only weak support because of lack of empirical power (Ostergaard, 1992 and because of averaging artifacts (Hintzman, 1990; Hintzman & Hartry, 1990). For the counter model, the reasoning b which findings of stochastic independence are explained is th same as for functional independence: The aspects of a word ma distinguish it perceptually from other words are likely to b independent from aspects of the word that make it easy to con sciously retrieve.

Generality Across Tasks

In research on implicit memory, experimental tasks like per ceptual identification and word-fragment completion are ofte grouped according to their similarity in showing priming effect (see Roediger & McDermott, 1993, for a review). The questio this grouping raises is what are the implications of the counte model for priming effects in other tasks? The answer is th there are no direct implications; the model does not apply t stem completion or fragment completion or any task other tha perceptual identification.

What the model does show is that it is possible to understan priming effects in detailed quantitative way by understanding underlying mechanisms. The counter model explains priming in perceptual identification by explaining the mechanisms of perceptual identification. To understand priming in fragment completion or stem completion, for example, it is necessary to develop models for performance in those tasks. Only with reasonably adequate models can tasks be more properly grouped according to the mechanisms they invoke, rather than superficial resemblances. The need for regrouping is evident from within the implicit memory framework: Witherspoon and Moscovitch (1989) found stochastic independence between perceptual identification and word fragment completion, and Perruchet and Baveux (1989) found, for several tasks, that the degree of correlation in performance did not properly divide the tasks into implicit versus explicit. Thus, the need for specific models as a means of generalizing across tasks is apparent even if the models are to be incorporated into implicit memory systems.

The Counter Model and Making Decisions

The counters in the counter model are devices for producing decisions about words when perceptual information is limited. A natural question is whether other functions or tasks such as reading and retrieving information from memory require these counters to identify words as one stage in processing prior to the information being used in later processes. The answer is a firm ''possibly.'' It could be that each word needs to be identified such that a discrete code can be passed on to the next stage of processing. The counter mechanism would provide exactly the mechanism for this. However, it is also possible that in a task such as reading, continuously available information is needed. Then, decision counters would only be needed when a response is required that is based on the information computed at the relevant point in the information-processing sequence flow. For example, for word identification, the relevant point is the point specified by the counter model, whereas for retrieval from memory, a later point in processing would be relevant. However, the same general class of decision mechanism can be used at different points in the flow of information. The random walk or diffusion model class of mechanisms has been used both for the counter model presented here and for models of recognition memory and lexical decision (Ratcliff, 1978, 1981, 1985, 1987, 1988; Ratcliff & van Zandt, 1996).

The counter model applies to three word-identification tasks: naming, forced choice, and single word "yes-no" decision. For these tasks, the decision process uses multiple counters or two counters or one counter, respectively. This points to the flexibility that must be present in decision making in these tasks and in human information-processing in general. The system must be capable of applying a decision mechanism at the level that discriminative information is required. In addition, the kind of decision process is determined by the demands of the task (e.g., single test item or select among many alternatives). The counter-diffusion model is a good candidate for this general retrieval mechanism because it fits a range of data from a range of experimental paradigms, and because it can deal with binary decisions and naming tasks (using multiple counters).

Conclusions

The counter model provides a straightforward account of the mechanisms that produce responses in perceptual identification. It correctly predicts a range of data across a number of variables, and its success contrasts with the apparent inability of other models to accommodate the data. However, in the traditional terms of cognitive psychology, the counter model is not easy to understand. How is memory represented in the model? Not in the traditional way of information being stored and available for use in subsequent tasks. Rather, memory is that a word comes to attract more than its fair share of the perceptual evidence entering the system from a stimulus. Should the operations that are affected by this kind of memory be described as encoding operations or retrieval operations? In the counter model, this binary distinction, the kind of distinction that has grounded so much of cognitive psychology (cf. Newell, 1973), loses its precision. Identification of a word is certainly an encoding operation, but the identification depends on an attraction of counts to counters that is influenced by prior experience. We believe, as Hintzman (1990) has pointed out, that these kinds of labels and distinctions become irrelevant once a model has been spelled out; then "the explanatory burden is carried by the nature of the proposed mechanisms and their interactions, not by what they are called" (p. 121).

Priming in an implicit task has been interpreted as demonstrating the need to postulate multiple memory systems. The counter model, and the fact that it provides a reasonable account of priming in perceptual identification, show that, in at least one instance, multiple memory systems are not required. At one level, the memory systems view has been a very comfortable view for a broad range of constituents in cognitive neuroscience. The assumption of functionally distinct systems seems compatible in easily obvious ways with research on patients with head injuries, patients with amnesia, and brain-imaging techniques. But at another level, it is not immediately obvious why there should be a brain system like the word-form system which, in evolutionary terms, could not have developed to deal with the reading of words. These kinds of issues have led proponents of the memory systems view to an agenda of making lists of empirical phenomena to separate different brain systems (e.g., Schacter & Tulving, 1994, p. 16).

In contrast, by spelling out the mechanisms of perceptual identification, the counter model illustrates a different agenda. It is a more complicated agenda because it necessitates a focus on the mechanisms that compute information, the processes that produce responses, and the representations that hold information. It is an agenda by which theory must be totally open to new possibilities; for example, what look like memories might be simply alterations to processing (Kolers & Roediger, 1984). Different alterations to processing might be carried out by the same mechanism operating in different circumstances. One location in the brain might carry out a variety of different computations, and one kind of process might be the result of operations across a number of locations. However, if specific quantitative models can be developed, not just for perceptual identification but also for other implicit tasks, then they can be combined with neuropsychological efforts and it may be possible for theories to advance beyond qualitative descriptions of empirical phenomena.

The difference in emphasis of the two approaches is made evident by an examination of recent research (see Ratcliff & McKoon, 1996). The implicit memory systems approach has not resulted in detailed explanations of processing for the experimental tasks used in implicit memory research. For example, there are no recent models to explain how the tasks of stem completion, fragment completion, or perceptual identification are performed and how a memory system might contribute to performance. The memory systems approach has provided a wealth of experimental data, and now it is time to begin trying to understand the mechanisms and processes underlying the empirical phenomena. The counter model is detailed and specific enough that it can be tested both qualitatively and quantitatively. New data should be generated to support it or falsify it, and competing models should be developed and competitively tested against it. It is our hope that any model that replaces the counter model will provide an explicit mechanistic account of the phenomena documented in this article and lead to greater empirical coverage and better explanations of existing data.

References

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234-254.

Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.

Broadbent, D. E. (1967). Word frequency effect and response bias. *Psychological Review, 74,* 1-15.

Clarke, R., & Morton, J. (1983). Cross modality facilitation in tachistoscopic word recognition. *Quarterly Journal of Experimental Psychology, 35A,* 79-96.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Domic (Ed.), *Attention and performance, VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205–254.

Feller, W. (1968). *An introduction to probability theory and its applications.* New York: Wiley.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effect of competition and inhibition. *Journal of Memory and Language, 28,* 501–518.

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 978-992.

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

Hintzman, D. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528-551.

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology Annual, 41,* 109-139.

Hintzman, D. L., & Hartry, A. L. (1990). Item effects in recognition and fragment completion: Contingency relations vary for different subsets of words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 955-969.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory and Cognition, 12,* 568-574.

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

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

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

Jacoby, L. L., & Hayman, G. C. (1989). Specific word transfer as a measure of processing in the word-superiority paradigm. *Memory and Cognition, 17,* 125-133.

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

Johnston, J. C., & Hale, B. L. (1984). The influence of prior context on word identification: Bias and sensitivity effects. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance, X.* (pp. 243-255). Hillsdale, NJ: Erlbaum.

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

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

Laming, D. R. J. (1968). *Information theory of choice reaction time.* New York: Wiley.

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review, 96,* 25-57.

Light, L. L., & Kennison, R. F. (in press). Guessing strategies, aging, and bias effects in perceptual identification. *Consciousness and Cognition.*

Link, S. W. (1975). The relative judgement theory of two choice response time. *Journal of Mathematical Psychology, 12,* 114-135.

Link, S. W, & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77–105.

Luce, R. D. (1959). *Individual Choice Behavior.* New York: Wiley.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

Masson, M. E. J., & MacLeod, C. M. (in press). Contributions of processing fluency to repetition effects in masked word identification. *Canadian Journal of Experimental Psychology.*

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, 88,* 375–407.

McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises.* Cambridge, MA: MIT Press.

McKoon, G., & Ratcliff, R. (1995). How should implicit memory phenomena be modeled? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 777-784.

McKoon, G., & Ratcliff, R. (in press). Separating implicit from explicit retrieval processes in perceptual identification. *Consciousness and Cognition.*

Morton, J. (1968). A retest of the response bias explanation of the word frequency threshold effect. *British Journal of Mathematical and Statistical Psychology, 21,* 21–33.

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. Wrolstad, & H. Bouma (Eds.), *Processing visible language 1* (pp. 259-268). New York: Plenum.

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.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology, 102,* 963-968.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700-708.

Ostergaard, A. L, (1992). A method for judging measures of stochastic dependence: Further comments on the current controversy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 413-420.

Ostergaard, A. L., & Jernigan, T. L. (1993). Are word priming and explicit memory mediated by different brain structures? In P. Graf & M. E. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology* (pp. 327–349). Hillsdale, NJ: Erlbaum.

Perruchet, P., & Baveux, P. (1989). Correlational analyses of explicit and implicit memory performance. *Memory and Cognition, 17,* 77–86.

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

Rajaram, S., & Roediger, H. L., III. (1993). Direct comparison of four implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 765–776.

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. (1987). More on the speed and accuracy of positive and negative responses. *Psychological Review, 94,* 277-280.

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

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review, 97,* 285-308.

Ratcliff, R., Allbritton, D. W., & McKoon, G. (1997). Bias in auditory priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 143-152.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385-408.

Ratcliff, R., & McKoon, G. (1995). Bias in the priming of object decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 754-767.

Ratcliff, R., & McKoon, G. (1996). Bias effects in implicit memory tasks. *Journal of Experimental Psychology: General, 125,* 403-421.

Ratcliff, R., McKoon, G., & Verwoerd, M. (1989). A bias interpretation of facilitation in perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 378-387.

Ratcliff, R., & van Zandt, T. (1996). *Comparing connectionist and diffusion models of reaction time.* Manuscript submitted for publication.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475-543.

Richardson-Klavehn, A., Lee, M. G., Joubran, R., & Bjork, R. A. (1994). Intention and awareness in perceptual identification priming. *Memory and Cognition, 22,* 293–312.

Roediger, H. L., III, & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boiler & J. Grafman (Eds.), *The handbook of neuropsychology.* (Vol. 8, pp. 63-131). New York: Elsevier.

Rueckl, I. G. (1990). Similarity effects in word and pseudoword repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 374-391.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60-94.

Schacter, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.), *Annals of the New York Academy of Sciences: Vol. 608. Development and neural bases of higher cognitive functions* (pp. 543-572). New York: New York Academy of Sciences.

Schacter, D. L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 233-238). Cambridge, MA: MIT Press.

Schacter, D. L., & Cooper, L. A. (1995). Bias in the priming of object decisions: Logic, assumptions, and data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 768-776.

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 1–38). Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523-568.

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems 1994* (pp. 203-231). Cambridge, MA: MIT Press.

Stone, M. (1960). Models for choice reaction time. *Psychometrika, 25,* 251-260.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247,* 301-306.

Tversky, A. (1977). Features of Similarity. *Psychological Review, 84,* 327-352.

Whitlow, J. W., Jr., & Dalton, P. (1993, November). *On the persistence of priming.* Paper presented at the 34th annual meeting of the Psychonomic Society, Washington, DC.

Winnick, W. A., & Daniel, S. A. (1970). Two kinds of response priming in tachistoscopic recognition. *Journal of Experimental Psychology, 84,* 74-81.

Witherspoon, D., & Moscovitch, M. (1989). Stochastic independence between two implicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 22–30.

Received June 26, 1995
Revision received May 15, 1996
Accepted June 27, 1996 •