Psychological Review
2001, Vol. 108, No. 4, 835–846

0033-295X/01/$5.00 DOI: 10.1037//0033-295X.108.4.835

A Multinomial Model for Short-Term Priming in Word Identification

Roger Ratcliff and Gail McKoon
Northwestern University

A simple multinomial model for short-term priming in perceptual word identification is presented. In the experiments to which the model is applied, prime words are presented just prior to a flashed target word, and subjects must decide which of 2 alternative words matches the target. The model assumes that on some proportion of trials, confusion among the words leads to the decision being based on 1 of the prime words instead of the target. In addition, it is assumed that subjects sometimes discount a prime that matches 1 of the test alternatives and so choose the alternative that does not match. With these assumptions, the model fits the data from 5 experiments (including 4 used to develop the model known as *ROUSE* [responding optimally with unknown sources of evidence]; D. E. Huber, R. Shiffrin, K. Lyle, & K. Ruys, 2001). The multinomial model fits the data about as well as the ROUSE model and so should lead to further development and critical testing of both models.

In this critique, we present a multinomial model to explain the short-term priming effects in word identification that have been observed by Huber, Shiffrin, Lyle, and Ruys (2001). In the priming paradigm used by Huber, Shiffrin, Lyle, and Ruys, two prime words are presented immediately prior to a briefly flashed target word. After the target, two alternative words are presented, and subjects must decide which of them matches the flashed target. The principal finding is that accuracy is reduced when one or both of the primes is similar in some way to the target. Huber, Shiffrin, Lyle, and Ruys explained this with the model known as *ROUSE* (responding optimally with unknown sources of evidence), an extension to short-term phenomena of the Bayesian decision framework developed recently for longer term priming effects (the model known as *REMI* [retrieving effectively from memory, implicit]; Schooler, Shiffrin, & Raaijmakers, 2001). According to the ROUSE model, the difficulty caused by similar primes comes about because words in short-term memory are represented as vectors of features, and features of the target word cannot be kept completely separate from features of the prime words or from noise; the mixing up of features leads to errors. In the model, word identification decisions are made using a Bayesian decision process that depends on calculating the likelihoods of features being from the target word. It is assumed that subjects can estimate the probability that the source of each feature was a prime word, the target word, or noise, and so they can discount those features that might have come from a prime and not the target. However, they cannot discount perfectly and so they make errors that reflect biases either for or against primes.

The multinomial model we present offers a simpler explanation. Identification of the target is assumed to take place by standard word identification processes, the same processes that operate in any situation in which identification of a word is required. Items in short-term memory are represented as words, not vectors of features, and all that happens in the short-term priming paradigm is that on some proportion of trials, subjects make a simple mistake: They confuse the primes and target and base their word identification decision on one of the prime words instead of on the target. The multinomial model accounts for Huber, Shiffrin, Lyle, and Ruys's (2001) data about as well as ROUSE does. As a competitor to ROUSE, the multinomial model is designed to provide impetus for further, theoretically based model testing with the goal of explaining short-term priming in word identification.

The Data to Be Explained

The experiments to which ROUSE is applied (Huber, Shiffrin, Lyle, & Ruys, 2001) use a two-alternative forced-choice word identification procedure. On each trial, the two prime words are presented simultaneously one above the other, then the target word is flashed, and then the two test alternative words are presented simultaneously side by side. One of the test alternatives is the same word as the target; the other (the "foil") is a different word. Subjects are asked to decide which of the two test alternatives matches the flashed target word. The amount of time for which the target is flashed is usually set so that performance is about midway between chance (50% correct) and ceiling (100% correct). This allows measurement of the effects of the prime words on performance, whether they hurt performance or help it.

The main questions of interest in Huber, Shiffrin, Lyle, and Ruys's (2001) experiments concerned how performance is affected by relationships among the primes, the target, and the test alternatives. Across Huber, Shiffrin, Lyle, and Ruys's experiments, the test alternatives were different words from the primes, the same words as the primes, or words that were semantically or ortho-

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

Preparation of this article was supported by National Institute of Mental Health Grant HD MH44640 and National Institute on Deafness and Other Communication Disorders Grant R01-DC01240. We thank David Huber for extremely useful suggestions about how to improve presentation and suggestions that led to better assumptions about fitting some data, and also for details of data and experiments from the Huber, Shiffrin, Lyle, and Ruys (2001) article. We also thank Rich Shiffrin and Mike Masson for useful comments on the research.

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

graphically similar to the primes. One question, for example, was whether, if one of the prime words was the same word as the target, subjects would be more likely to choose the test alternative that matched the target; if so, then performance would show a benefit. But if one of the prime words was the same as the foil, then subjects might be more likely to choose the foil instead of the target, and so there would be a cost to performance.

Other questions in Huber, Shiffrin, Lyle, and Ruys's (2001) experiments concerned how the effects of the primes on performance vary with the kind of attention subjects give to the primes. If subjects are induced to actively, as opposed to passively, process the primes, then they might better be able to discount features of the primes in their decisions about the test alternatives. For passive priming, the prime words were simply displayed for 500 ms and subjects were told they were a warning signal for the flashed target word. For active processing, the prime words were displayed twice. On their first presentation, they were displayed simultaneously and subjects were asked to make a decision about them (e.g., whether both words referred to animate entities, a decision that took 2–3 s). Immediately following this decision, the prime words were displayed a second time, for the same 500-ms duration as was used in the passive priming condition.

Table 1 shows experimental conditions typical of Huber, Shiffrin, Lyle, and Ruys's (2001) experiments. The first and second columns show four priming conditions. In the first condition, the baseline condition, neither the target nor the foil match either of the primes; in the second condition, the target matches one of the primes; in the third condition, the foil but not the target matches one of the primes; and in the fourth condition, the target and foil both match primes. The third column shows typical stimuli when the match relation among primes, target, and test alternatives is one of exact repetition. The fourth and fifth columns show typical stimuli for the same four conditions when the relation among the primes, target, and test alternatives is associative and orthographic, respectively. The stimuli in the sixth column will be explained later.

The patterns of data found by Huber, Shiffrin, Lyle, and Ruys (2001) present an intriguing puzzle for modeling. Consider first what happens when the four conditions ("neither primed," "target primed," "foil primed," or "both primed") are tested with passive processing of the primes (the case in which the primes are displayed only once, for 500 ms). As might be expected, there is a benefit in the target-primed condition (relative to baseline, the neither condition) and a cost in the foil-primed condition (relative to baseline). Accuracy is higher in the target-primed condition than baseline, and accuracy is lower in the foil-primed condition than baseline. The surprising finding is that there is also a cost in the both-primed condition, when both the target and the foil are primes. The finding is surprising because in long-term priming, when both primes were studied at some earlier time (e.g., tens of seconds or minutes earlier) in a list of other words, there is no decrement in performance (e.g., Masson & MacLeod, 1996; Ratcliff & McKoon, 1997).

Although the pattern of data with passive priming alone provides a puzzle for modeling, the data obtained with active processing of the prime words add more constraints. When subjects actively process the primes (the case in which subjects make some decision about the two prime words), the benefit in the target-primed condition disappears (and in some experiments turns into a cost) and the cost in the foil-primed condition disappears (and sometimes turns into a benefit), but the cost of priming both the target and the foil does not disappear. A simple hypothesis would be that with active priming, subjects attempt to eliminate any effects of the primes on their decisions. This would explain the disappearance of the benefit from priming the target and the disappearance of the cost from priming the foil, but it would not explain why the cost from priming both test alternatives does not disappear. Something more is required.

The ROUSE Model

According to the ROUSE model, in the decision process for the short-term priming paradigm, words are represented as vectors. The vector for a word is made up of a series of lexical–semantic features—20 features in Huber, Shiffrin, Lyle, and Ruys's (2001) applications of the model. For the prime words and the test alternatives, display time is sufficient for all 20 of their features to be encoded. For the flashed target, display time is cut short and so only a subset of its features is perceived. It is assumed that the representation of a word is always the same. For example, if a

Table 1
Examples of Conditions in the Experiments

Priming condition	Stimulus sequence[a]	Repetition	Associative similarity	Orthographic similarity	Repetition with similar test alternatives
Neither primed	prime1 prime2	chef acre	chef acre	pier colt	pier colt
	target	shoe	shoe	hale	hale
	target foil	shoe frog	shoe frog	hale duel	hale hail
Target primed	target prime2	shoe acre	sock acre	hail colt	hale colt
	target	shoe	shoe	hale	hale
	target foil	shoe frog	shoe frog	hale duel	hale hail
Foil primed	prime1 foil	chef frog	chef toad	pier dual	pier hail
	target	shoe	shoe	hale	hale
	target foil	shoe frog	shoe frog	hale duel	hale hail
Both primed	target foil	shoe frog	sock toad	hail dual	hale hail
	target	shoe	shoe	hale	hale
	target foil	shoe frog	shoe frog	hale duel	hale hail

[a] Stimulus sequence for each priming condition: primes, flashed target, and test alternatives.

word is presented as both a prime and a test alternative, the same 20 features are encoded in both cases. If two words are similar to each other, either orthographically or semantically, then they have some features in common, and words that are defined in the experiments as being dissimilar to each other are assumed to have no features in common.

The features of the test alternatives are assigned values, which provide the basis for the decision. The initial value of all the features is 0, and the value of a feature is switched to 1 if the feature is the same as any of the lexical–semantic features that were encoded as features of the target word. Because of confusion and noise, not all the features that are encoded as features of a target actually come from the target itself. A feature that is encoded as a feature of a target word can come from any of three sources: It could have been a genuine feature of the target, it could have been a feature of a prime that was confused with the target, or it could have been generated by noise in the system. A genuine feature of the target word is encoded with probability β, a feature of a prime word is encoded as a feature of the target with probability α, and the probability of a feature being encoded into the target from noise is γ. No matter what the source of a target feature, the value of any test alternative feature that matches a target feature is set to 1.

Once the feature values for the test alternatives' vectors are set, a Bayesian decision rule is used to decide which of them should be the response. The idea is to determine whether the patterns of feature values in the vectors are more likely given that Alternative A was the flashed target or given that Alternative B was the flashed target.

To determine the overall likelihoods of the patterns in the vectors for the two test alternatives, the likelihoods are calculated for each of the individual features in the vectors. For each feature, the conditional probability of the feature's value is calculated given that the feature is a feature of the target word and given that it is not a feature of the target word. If the system had perfect knowledge of α, β, and γ, and if ROUSE made no further assumptions, the probabilities of feature values would be straightforward to lay out. For example, for a feature that is in a prime word and the target word, its probability of having the value 0 in the test alternative vector would be the probability that the feature was not encoded as part of the target, not confused between the prime and the target, and not encoded from noise: $(1 - \beta)(1 - \alpha)(1 - \gamma)$. If the feature is in the prime word but not in the target word, the feature could have the value 0 if it was not confused between the prime and target and if it was not encoded from noise: $(1 - \alpha)(1 - \gamma)$.

However, the decision system does not have perfect knowledge and ROUSE makes several further assumptions. First, consider the probabilities α, β, and γ. The cognitive system cannot have exact knowledge of these probabilities and so they must be estimated. The estimates are labeled α', β', and γ'. How α, β, and γ are estimated is not explained by ROUSE; this process is considered to be outside ROUSE's domain. However, ROUSE makes the assumption that β, the probability a feature of the target word itself is perceived and encoded, and γ, the probability a noise feature is encoded as part of the target, are accurately estimated. On the other hand, for α, the probability of a prime feature being encoded as a target feature, it is assumed that the estimate is not accurate. Therefore, in the equations for probabilities, β can be substituted for β' and γ can be substituted for γ', but α cannot be substituted for α'.

Second, a crucial mechanism of ROUSE is discounting. If a feature in a test alternative also occurs in a prime word, then it is possible that the feature has the value in the test alternative vector that it does because of the prime word and not the flashed target. ROUSE is designed to discount feature values that could have come about from confusion with prime words. This is done with α'. The assumption is that the estimate of α' is higher when the amount of discounting is high and lower when the amount of discounting is low.

Third, discounting should apply to features that a test alternative shares with a prime and not to features that the test alternative shares only with the target. To accomplish this, it is assumed that ROUSE can keep track of which features are shared between which words. How many features are shared between test alternatives and primes—that is, the probability that a feature is shared—is a parameter of the model, ρ. If the prime and test alternative are the same word, the probability of their sharing features is 1.00; if they are associatively similar, fits of the model to data give a probability of .07 or .30 (depending on the experiment); and if they are orthographically similar, fits give a probability of .70 to .80 (depending on the experiment).

Finally, if the test alternatives are similar to each other, they will have shared features. If a feature encoded as a feature of the target word (whether it comes from the target, noise, or confusion with a prime) is also a feature shared between the test alternatives, then that feature will have the same value in the two vectors. Therefore, the values of these shared features can be dropped from the probability calculations. The probability that a feature is shared between the two test alternatives is the same parameter ρ as the probability that a feature is shared between a prime word and a test alternative. For orthographically similar words, in Experiments 2 and 3 (Huber, Shiffrin, Lyle, & Ruys, 2001), the value of ρ for the probability of a feature being shared between the two test alternatives is the same as the value of the probability of a feature being shared by a prime and a test alternative. However, in Experiment 4, for orthographically similar words, the probability that a feature is shared between a prime and a test alternative is .80, whereas the probability that a feature is shared between the two test alternatives is .07. One reason for this large difference may be that in Experiment 4, orthographically similar primes and test alternatives shared four out of five letters and orthographically similar test alternatives shared only three out of five letters.

Given all of these assumptions, for each test alternative, the probability of each of its features' values can be determined given that the test alternative matches the target and given that the test alternative does not match the target, and the ratio of these probabilities can be constructed. In Table 2, the right-hand columns show the probabilities for a feature that is common to a prime and a test alternative; the left-hand columns show the probabilities for a feature that is not. Likelihood ratios are calculated from the probabilities shown in the table. In each of the four cases, the numerator of the ratio is the probability in the top row, and the denominator is the probability directly below in the bottom row. Suppose, for example, that a feature in the vector for Test Alternative A is not in a prime and has the value 0. The numerator of the ratio expresses the probability that the feature has the value 0 given that Test Alternative A was the flashed target, $(1 - \gamma)(1 -$

Table 2
Probability That a Test Alternative's Feature Has the Value 0 or 1

Probability	Feature not in a prime: Feature value = 0	Feature not in a prime: Feature value = 1	Feature in a prime: Feature value = 0	Feature in a prime: Feature value = 1
Probability of feature value given test alternative = target	$(1-\gamma)(1-\beta)$	$1-(1-\gamma)(1-\beta)$	$(1-\gamma)(1-\beta)(1-\alpha')$	$1-(1-\gamma)(1-\beta)(1-\alpha')$
Probability of feature value given test alternative ≠ target	$(1-\gamma)$	$1-(1-\gamma)$	$(1-\gamma)(1-\alpha')$	$1-(1-\gamma)(1-\alpha')$

β), and the denominator expresses the probability $(1-\gamma)$ that the feature has the value 0 given that Test Alternative A was not the flashed target.

It is assumed that each feature is an independent source of evidence, so to determine whether the overall patterns of feature values (0s and 1s) in the vectors are more likely given that A was the target or given that B was the target, the odds ratio in Equation 1 is calculated

$$\Phi\left(\frac{A}{B}\right) = \frac{\prod_{i=1}^{N} \frac{p(V\{A_i\}|A \text{ is target})}{p(V\{A_i\}|A \text{ is foil})}}{\prod_{j=1}^{N} \frac{p(V\{B_j\}|B \text{ is target})}{p(V\{B_j\}|B \text{ is foil})}}, \quad (1)$$

where $V(A_i)$ represents the value of the *i*th feature of **A** and N is the number of features, 20 in Huber, Shiffrin, Lyle, and Ruys's (2001) applications. The numerator is the product of the likelihood ratios for all the feature values in Vector **A** and the denominator is the product of the likelihood ratios for all the feature values in Vector **B.** If the value of the ratio is greater than 1, then the system decides in favor of Alternative A; if it is less than 1, the system decides in favor of Alternative B; and if it is exactly 1, the system guesses with probability .50 for each alternative.

ROUSE requires significantly large cognitive resources. First, it is not clear how α, β, and γ could be estimated. Consider β, the probability that a feature of the flashed target is perceived and encoded. To estimate β, features that are genuinely from the flashed target would have to be kept separate from prime and noise features, and the genuine target features would have to be counted across a number of trials. It is not obvious how this could be done. In fits of ROUSE to data, β is usually about .05. This means that, on average, on each trial of an experiment, only about 1 out of the 20 features of the target word is encoded. With the 80 trials of practice typically used by Huber, Shiffrin, Lyle, and Ruys (2001), the standard deviation in the estimate of β would be .005 (assuming $\beta = .05$). In other words, even with perfect information on all 80 trials about which features came from the target versus the primes or noise, the estimated value of β after 80 trials would vary between .04 and .06, and so, contrary to ROUSE's assumption, an accurate value of β could not be produced. The same problem arises in estimating γ and α.

A solution to these problems would be to assume that subjects set their estimates of the probabilities according to prior experience, not according to information gained during the experiment. This has the advantage of being more realistic; certainly, everyday situations in which information is flashed quickly are not repeated 80 times to allow for estimates of probabilities. For each of β and γ, this solution would add a parameter to the ROUSE model (the model would include parameters for both β and β' and γ and γ', as well as α and α'). However, the number of parameters could be minimized by using the same values of β' and γ' across all experiments (as should be the case if they are set from prior experience), and it is likely that good fits of ROUSE to data could be preserved. For α and α', it could be assumed that prior experience suggests likely probabilities for discounting—a larger value for more discounting and a smaller value for less discounting.

Another way in which ROUSE requires significantly large resources is that the decision system must hold full and complete knowledge of the features of the primes, target, and test alternatives. The rule for switching feature values in the test alternatives to 1 is that they must be features shared with the encoded representation of the target. This requires some process that holds the encoded target features (the features from the target itself, from confusion with the primes, and from noise) and identifies correspondences, that is, identifies which of the target features match which of the features in each of the test alternatives. Moreover, in calculating the likelihood ratios, the system must perform the calculations differently according to whether a feature of a test alternative is shared with a prime or is not shared with a prime (in the former case, the calculations include α', and in the latter case, they do not; see Table 2). Finally, the system must keep track of features shared between the two test alternatives to drop them out of the probability calculations. Overall, the system must keep track of the correspondences between all the individual features of all the pairwise combinations of the primes, target, and test alternatives, properly aligning each feature of each word with its matching features in each of the other words. Feature 3 in a test alternative, for example, might correspond to Feature 2 in the target, Feature 13 in one prime, and no feature at all in the other prime.[1] It is possible that some of these problems could be at least partly solved by moving to more standard lexical–semantic representations (e.g., latent semantic analysis; Landauer & Dumais,

[1] It might seem that the model could assign features to specific positions so that, for example, the features for the second letters of the prime words would line up with the features for the second letters of the test alternatives, but the combinatorial problem is too large. Consider the five 6-letter words *bakery*, *banker*, *barber*, *barker*, and *darker*. At each of the six positions, there could be 26 different letters. The two words *bakery* and *barker* share five letters and they are quite similar, but of the five letters they share, only the first two are in the same position. If features were aligned with letter positions, *bakery* and *barker* would be no more similar than *bakery* and *hungry* (which share their last two letters). Also, words vary in their lexical

1997) in which elements of meaning have fixed places in vectors so that individual elements line up from one vector to another. However, this would require significant reparameterization of ROUSE, and it is not clear if a revised version of ROUSE could fit the experimental data.

In its current form, the ROUSE model gives an accurate account of data from the short-term priming paradigm. The fits of the model to the data are generally within two standard errors. Moreover, the complete set of data provides new insights into short-term priming, and the experiments that generated these data would not have been conducted without the guidance of the model. However, ROUSE is not the only approach. The multinomial model presented below offers a simpler account and a different framework for understanding and interpreting the data, and it fits the data about as well as ROUSE. Our aim in presenting the multinomial model is to provide a basis for competitive model testing with the hope of generating new data and new theoretical insights. The best position to be in during the evolution of modeling in a domain is to have competing models. The competition can produce empirical tests that would not otherwise have been thought of, and because the multinomial model is so simple, research can focus clearly on exactly where the more sophisticated representations and calculations of ROUSE's Bayesian decision process are needed.

The Multinomial Model: Overview

In the main, the multinomial model is put together from "off-the-shelf" parts. The model assumes a standard word identification process for the flashed target. There are a number of models for word identification, including the logogen model (Morton, 1968, 1970), the counter model (McKoon & Ratcliff, 2001; Ratcliff & McKoon, 1997, 2000), and REMI (Schooler et al., 2001). For the purposes of explaining Huber, Shiffrin, Lyle, and Ruys's (2001) data, it is not necessary to choose among these models. Each can produce what the multinomial model requires for the short-term priming situation, namely some value p that is the probability of identifying a briefly flashed target word.

Instead of the vectors of features used in ROUSE, the multinomial model treats the prime, target, and test alternative items as unitized words in short-term memory, a common assumption. Previous theoretical and empirical work (e.g., Lee & Estes, 1977) has made salient the notion that items in short-term memory are often confused with each other in terms of their order and identity. The multinomial model borrows this notion for the principal assumption in its explanation of performance in the short-term priming situation: On some proportion of trials, the flashed target word is confused with one of the prime words. Instead of the forced-choice decision being made on the basis of which of the two test alternatives matches the flashed target, the decision is made—by mistake—on the basis of which of the two test alternatives matches one of the prime words. This assumption is consistent with standard views of how position and order information is lost for recently presented items in short-term memory (see Neath, 1998, chap. 14).

semantic similarity; for example, *banker* and *barber* are both occupations. Overall, there are many more possible kinds of similarity combinations than can be represented by 20 aligned feature slots.

Making the decision between the two test alternatives on the basis of a prime word instead of the target word will sometimes help performance and sometimes hurt performance. If the prime is the same word as the target, it will help performance, but if the prime is not the same as the target—if it is the same as the foil—then the decision between the test alternatives will be wrong.

According to the multinomial model, confusion can occur between a prime word and the flashed target both when processing of the primes is passive and when it is active. The multinomial model also has a second mechanism that operates only when processing of the prime is active. This mechanism is similar to the discounting mechanism in ROUSE. The idea is that, with active processing of the primes, subjects tend to realize that prime–target confusions occur and so they attempt to correct for confusion errors by "discounting" prime–test alternative matches—that is, they choose a test alternative that does not match a prime word. It should be noted that, for ease of explication, we speak of discounting as a choice subjects make, but we intend no commitment to whether or not this is a conscious or strategic process.

Overall, the multinomial model gives a simple outline of how two-alternative forced-choice decisions are affected by prime words that immediately precede a flashed target word. If neither prime word matches the target or the foil, the probability of a correct response is p, a probability that is assumed to be the output of a word identification model. With passive processing of the prime words, there are two possibilities: If a prime word matches a test alternative but there is no confusion, then the probability of a correct response is still p. If confusion does occur, then the decision is based on the matching prime by mistake. With active processing of the prime words, subjects not only can be confused but also can sometimes choose to vote against prime words that match a test alternative.

There are several key differences between the multinomial model and ROUSE. For one, the primes, target, and test alternatives are represented in the multinomial model as words, not as vectors of features. In ROUSE, confusion can occur between all pairs of matching features, features in the primes, target, and test alternatives—400 possible combinations per pair of words—so all this information needs to be stored for use in the decision process. In the multinomial model, there are only two possibilities for confusion to occur: between each prime word and the target word. Little information needs to be stored. When a prime and target are confused, it may be that the prime information has simply overwritten the flashed target information in short-term memory. Although both ROUSE and the multinomial model have discounting, in ROUSE, discounting is done on a feature-by-feature basis, whereas in the multinomial model, it is a whole word that is either discounted or not discounted. Finally, the decision in ROUSE is based on the calculation of Bayesian probabilities. In the multinomial model, the decision is based on the probability that information is available from the target combined with the probabilities of confusion and discounting.

The multinomial model does a good job of accounting for Huber, Shiffrin, Lyle, and Ruys's (2001) data and for the data of the new experiment described below. It fits the data quantitatively about as well as ROUSE does, with about the same number of parameters. In the sections below, we present the multinomial model in more detail and then show how it fits the data.

The Multinomial Model: Details

Figure 1 displays the multinomial model by showing probabilities in trees. Each panel shows one of the four priming conditions. The dotted lines separate passive priming predictions (on the left) from active priming predictions (adding the branch on the right to the branches on the left). Each branch of each tree is independent of every other branch. This means that the model is neutral to order of processing, although we prefer to think of the flashed target processed first, followed by the possibilities of confusion with a prime word and, in the active priming conditions, discounting. For those branches that end in 1, the decision is to respond with the target alternative. For those branches that end in 0, the decision is to respond with the foil alternative.

When the target is flashed, information about it becomes available through the word identification processes of one or another of the models mentioned above. There are two mathematically equivalent (but psychologically different) ways to think about this. One is that identification is all or none; either the target is correctly identified from the stimulus (with some probability p_i) or no information about the target is available (with probability $1 - p_i$) and a guess is required. The other possibility is that perceptual processes produce some probability p that the target was the word that was flashed; $1 - p$ is the probability that the foil was the word

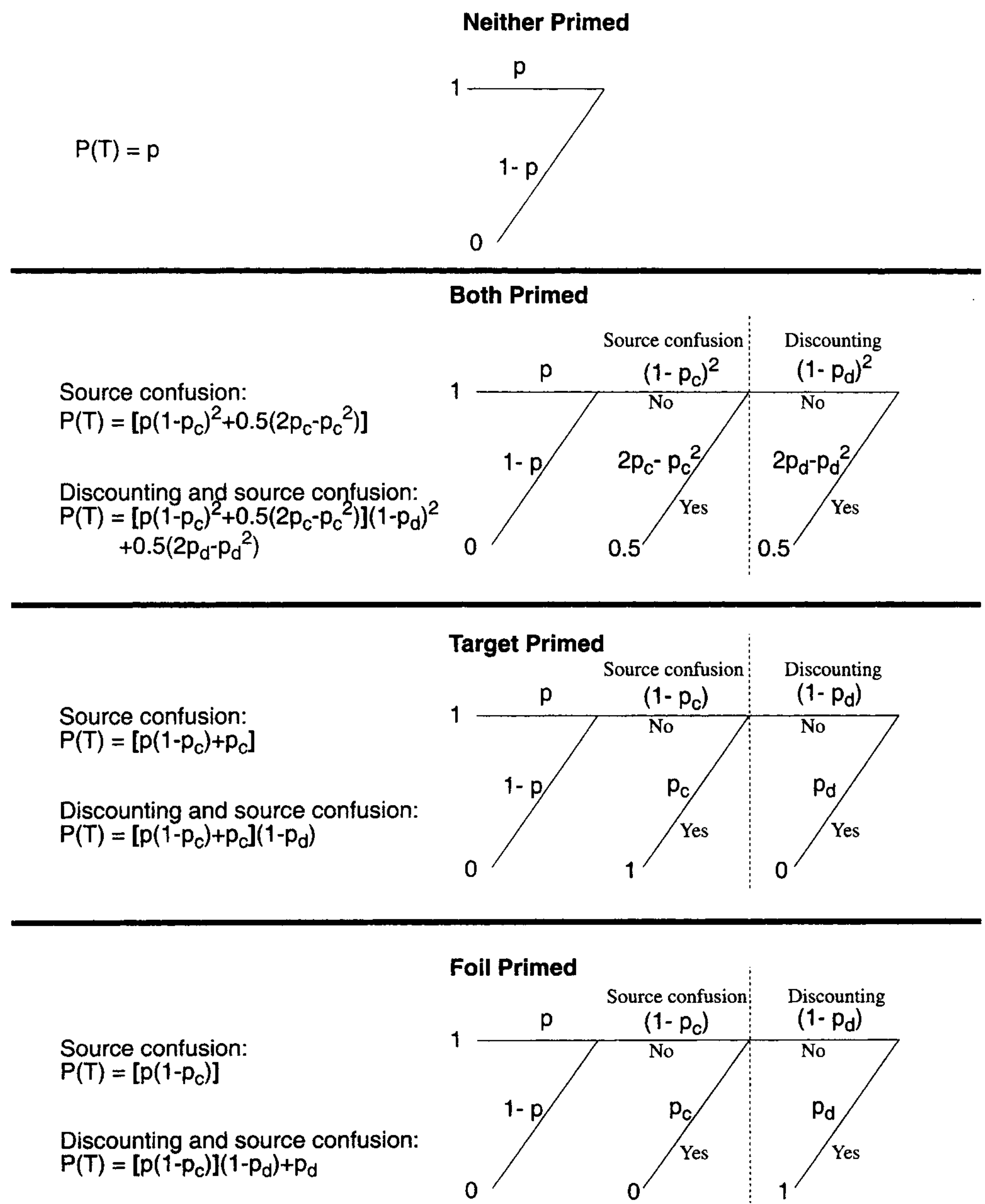

Figure 1. An illustration of the multinomial model showing the four priming conditions. P(T) is the probability of selecting the target from the two test alternatives, and the numbers 0 and 1 at the ends of the branches of the trees represent the probability of choosing the target given that the branch was selected.

flashed. In this latter way of thinking about the model, the system is never in a state with no information. This fits better with REMI (Schooler et al., 2001) and the counter model (Ratcliff & McKoon, 1997). Mathematically, p_i and p are related by $p_i = 2(p - 1)$, so the multinomial model could use either assumption (although the two versions of the model would be slightly different because with the p_i assumption, some proportion p_i of trials would require a guess, i.e., the leftmost 0 in each panel of Figure 1 would change to .5). For the remainder of this article, we adopt the assumption that word identification processes produce a value of p, that is, a probability that the target was the flashed word, but all the results we report would be equivalently consistent with the p_i assumption.

In the neither-primed condition (first panel, Figure 1), neither prime matches either test alternative, so there is no source confusion or discounting, and so the probability of selecting the target is $P(\text{T}) = p$. In the tree, the branch with probability p ends in 1, indicating that the target is chosen as the response, and the branch with $1 - p$ ends in 0, indicating the target is not chosen as the response.

Panel 3 of Figure 1 shows the target-primed condition. First, consider passive priming (the branches to the left of the dotted line). There are two ways that the target can be chosen as the response: Word identification processes favor the target and there is no confusion, $p(1 - p_c)$, or there is confusion and so the decision is based on the matching prime (p_c). Adding these together, the total probability of choosing the target with passive priming is shown at the left of the panel. With active priming, there are two ways the target can be chosen: (a) No discounting, no confusion, and word identification processes favor the target, or (b) no discounting and confusion, $(1 - p_d)(1 - p_c)p + (1 - p_d)p_c$. Panel 4 of Figure 1 shows the foil-primed condition, and probabilities are calculated by following the branches of the tree in the same way as for the target-primed condition.

The case where both primes match test alternatives is more complicated because there are two potential sources of confusion. The probability that there is confusion for each of the prime–test alternative pairs is p_c, so the total probability of confusion is $p_c + p_c - p_c p_c$. The probability that there is no confusion for each pair is $1 - p_c$, so the total probability of no confusion is $(1 - p_c)^2 = 1 - (p_c + p_c - p_c p_c)$. For discounting, the probability of discounting for each prime–test alternative pair is p_d, and so the probability for both is $p_d + p_d - p_d p_d$. The probability that there is no discounting for each prime–test alternative pair is $1 - p_d$, so the total probability of no discounting is $(1 - p_d)^2$. These expressions, following the branches that end in 1s, lead to the expressions for $P(\text{T})$ shown at the left of the panel.

When processing of the prime is passive, the word identification and source confusion mechanisms are the only two mechanisms needed to fit the data from all four priming conditions, and so the model has only the two parameters p and p_c. The probability of confusion, p_c, varies according to whether the relations among the primes, target, and test alternatives are repetition, associative similarity, or orthographic similarity. Confusion is more likely (the value of p_c is larger) for repetition and orthographic similarity than for associative similarity. From the equations for $P(\text{T})$, the model's predictions for the relative values of probability correct across the priming conditions for passive priming can be derived: $P(\text{T})$ for the both-primed condition should be less than p (which is $P[\text{T}]$ in the neither-primed condition), and $P(\text{T})$ for the foil-primed condition should be less than $P(\text{T})$ for the target-primed condition. These two predictions are consistent with Huber, Shiffrin, Lyle, and Ruys's (2001) data, as we show later. When processing of the prime is active, the multinomial model needs all three of its mechanisms to fit the data: word identification, confusion, and discounting. There are three parameters, the p and p_c parameters needed for passive priming plus the p_d parameter for discounting.

Flash-Time Experiment

One empirical question we addressed before fitting the multinomial model to data was whether the probability of confusion (the value of p_c) is a function of the amount of perceptual information entering the word identification system. Confusion should be less likely as the amount of perceptual information increases. To examine this question, we varied the amount of perceptual information by varying the flash time for the target. This experiment was run as a replication of one condition in an experiment by Huber, Shiffrin, Lyle, and Quach (2001).

The experiment used Huber, Shiffrin, Lyle, and Ruys's (2001) four priming conditions: neither primed, both primed, target primed, and foil primed. Processing of the primes was always passive. The relation among the primes, target, and test alternatives was repetition; example stimuli are shown in the third column of Table 1.

In the multinomial model, with passive priming, the difference between $P(\text{T})$ for the target-primed condition and $P(\text{T})$ for the foil-primed condition is $p(1 - p_c) + p_c - p(1 - p_c) = p_c$. Thus, with passive priming, the difference in probability correct between the target-primed and foil-primed conditions is a measure of p_c. The question for the experiment was whether this difference decreased as flash time increased.

Method

Materials. There were 228 quadruples of words. The words of a quadruple each had the same number of letters and about the same word frequency (Kucera & Francis, 1967). For each trial, the primes, target, and foil were chosen randomly from the four words of the quadruple. For 60 of the quadruples, the Kucera–Francis frequency of the words was 0 or 1. For the other 168 quadruples, the frequency of the words varied between 4 and 10,601.

Design, procedure, and subjects. The words were displayed on the PC screen of a Pentium class computer with the refresh rate on the screen set to 8 ms (Von Brisinski, 1994). Responses were collected on the PC's keyboard. Thirty-four students from an introductory psychology class participated in the experiment for credit in the class.

There were 12 conditions in the experiment: 3 flash times for the target (8 ms, 16 ms, and 24 ms) crossed with 4 priming conditions. The priming conditions were neither primed, target primed, foil primed, and both primed (for examples, see Table 1, column 3). Quadruples were assigned randomly to each condition with the restriction that there be equal numbers of quadruples of low-frequency words in each condition and equal numbers of quadruples of high-frequency words in each condition.

There were 228 trials in the experiment (preceded by 35 practice trials with 16- and 24-ms flash times and 5 initial practice trials with 120-, 80-, and 40-ms flash times to orient subjects to the sequence of events). Items in the 12 experimental conditions were presented in random order. The sequence of events for each trial was (a) a warning signal (a row of + symbols) for 240 ms, (b) a 500-ms delay, (c) the two prime words side by side for 248 ms, (d) the target word flashed for 8, 16, or 24 ms, (e) a mask (made up of random line segments about the same length as line segments in the fonts used for the words) for 200 ms, and then (f) the two alternatives side by side. The alternatives remained on the screen until the subject responded, pressing the "?/" key to indicate the right-hand alternative and the "z" key to indicate the left-hand alternative. A correct response was followed by the word *CORRECT* presented for 160 ms and an error response was followed by the word *ERROR* presented for 160 ms followed by a 500-ms delay. The right versus left positions of the primes and test alternatives were chosen randomly.

Results

Table 3 shows the data. The effects of flash time and priming condition were significant, $F(2, 66) = 159.32$, and $F(3, 99) = 13.16$, $p < .05$, respectively, and the interaction of the two was significant, $F(6, 198) = 5.81$, $p < .05$. The standard error of the means was .0177.

The main result is that the difference between the probabilities correct in the target-primed and foil-primed conditions (the value of p_c) decreased as flash time increased. With the shortest flash time, the probability of confusion was .219. The probability decreased to .112 for the 16-ms flash time and .031 for the 24-ms flash time. This confirms our expectation about the multinomial model's confusion mechanism: The probability of confusion decreases as flash time increases. In fitting the multinomial model to the data from this experiment, described below, we assumed that p_c decreases linearly from a maximum value of p_m when the probability of correct identification of the target is at chance ($p = .5$) to a value of 0 when the probability of correct identification is 1.0.

Fitting the Multinomial Model to Empirical Data

In the Huber, Shiffrin, Lyle, and Ruys (2001) article, data from four experiments were presented. Across the experiments, primes matched test alternatives by exact repetition, by orthographic similarity, or by associative relationship. Table 1 shows examples of the stimuli.

Experiment 1, Huber, Shiffrin, Lyle, and Ruys (2001). The 18 priming conditions are shown in Table 4. In one of the two mixed conditions, one of the primes was the target word and the other prime was associatively related to the foil, and in the other mixed condition, one of the primes was associatively related to the target and the other prime was the foil word. The same pool of words was used for all the conditions. Passive versus active priming was a between-subjects variable, as it was in all four of the Huber, Shiffrin, Lyle, and Ruys experiments discussed here.

We point out important aspects of the data using the repetition conditions as examples. First, with passive priming, probability correct is .69 in the neither-primed (baseline) condition, and probability correct goes up in the target-primed condition (to .77) and down in the foil-primed condition (to .57). Probability correct in the both-primed condition also goes down relative to the baseline condition (to .63). With active priming, the pattern of data for two of the conditions reverses: Probability correct in the target-primed condition decreases relative to the baseline, and probability correct in the foil-primed condition increases (slightly) relative to the baseline. However, for the both-primed condition, probability correct is still lower than the baseline.

The multinomial model has three parameters to fit the data from the nine passive priming conditions: The probability that word identification processing favors the target, p; the probability of confusion in the repetition conditions, one value of p_c; and the probability of confusion in the associative conditions, a different value of p_c. To fit the additional nine active priming conditions, the model adds one parameter for the probability of discounting, p_d. The best fits of the data were obtained when discounting was assumed to occur only in the repetition conditions, with no discounting in the associative conditions. The model was fit to the data using a general SIMPLEX (Nelder & Mead, 1965) minimization routine that adjusts the parameters of the model to find the parameters that give the minimum sum of squared differences between the data and the theoretical predictions of the model.

The results from fitting the model to the data are shown in Table 4 and the parameter values are shown in Table 5. The fits to the data are good, all within two standard errors of the data (see Huber, Shiffrin, Lyle, & Ruys, 2001, for a table containing the standard errors). In three cases in the active priming conditions, the model misses the data by about 4%, but these are still within two standard errors of the data. Comparing the multinomial model to ROUSE, chi-square goodness-of-fit values are 11.05 for ROUSE and 14.35 for the multinomial model, a reasonably small difference in quality of fits. The total number of parameter values, seven, was the same for ROUSE and the multinomial model.

Experiment 2, Huber, Shiffrin, Lyle, and Ruys (2001). For Experiment 2, primes matched test alternatives either exactly (repetition priming) or in terms of orthographic similarity (see Table 1). Different target items were used for the repetition conditions than the orthographic similarity conditions.

The fits of the multinomial model to the data were generally good (see Table 6), with only one point missing by more than two

Table 3
Probability Correct: Data and Fits of the Multinomial Model for the Flash-Time Experiment

Flash time (ms)	Data and fits from the multinomial model	Priming conditions: Neither primed	Both primed	Target primed	Foil primed
24	Theory	.890	.843	.896	.835
	Data	.859	.829	.909	.864
16	Theory	.810	.749	.830	.725
	Data	.812	.733	.843	.731
8	Theory	.627	.580	.704	.498
	Data	.611	.610	.711	.492

Table 4
Probability Correct: Data and Fits of the Multinomial Model for Experiment 1, Huber, Shiffrin, Lyle, and Ruys (2001)

		Repetition			Associatively similar			Mixed repetition and associative	
Data and fits from multinomial model	Neither	Both	Target	Foil	Both	Target	Foil	Repeat target	Repeat foil
Passive processing of primes									
Theory	.692	.637	.740	.584	.686	.697	.681	.745	.573
Data	.692	.626	.770	.567	.671	.733	.670	.712	.594
Active processing of primes									
Theory	.787	.688	.688	.738	.780	.789	.777	.690	.730
Data	.760	.647	.716	.781	.793	.790	.776	.686	.720

standard errors. Because there were different items in the repetition and orthographic similarity conditions, there were different values of p, p_c, and p_d for these conditions (see Table 5).

Experiment 4, Huber, Shiffrin, Lyle, and Ruys (2001). In this experiment, there was only one prime word instead of two. The match between test alternatives and prime was either orthographic or associative, different items in the two cases. The design required that the test alternative pairs be different across conditions. For example, with associative matches for the target *happy*, the prime, target, and test alternatives in each condition were, respectively: Both—*smile, happy, happy frown*; Neither—*table, happy, happy frown*; Target match—*smile, happy, happy above*; Foil match—*below, happy, happy above*.

The fits of the multinomial model were again good with all but three points falling within two standard errors of the data. There were three parameters, p, p_c, and p_d, with different values for the orthographically similar and associatively similar conditions (see Table 5). The chi-square goodness-of-fit values were 9.33 for ROUSE and 61.22 for the multinomial model, with a total of 11 different parameter values for ROUSE and 10 for the multinomial model.

Two of the misses that contributed most to the chi-square value for the multinomial model occurred with associative primes: Probability correct was higher, by about 3%, in the both-primed condition than in the neither-primed condition (this is the opposite pattern for these two conditions from what was found in all the other experiments). Probability correct being higher in the both-primed condition with associative primes suggests associative facilitation (semantic priming), something that has not been incorporated into either ROUSE or the multinomial model. However, the ROUSE model was able to come close to the observed data by allowing the probability of prime features being confused with target features to be about three times larger than the value in the other experiments and the value in the active priming condition ($\alpha = .378$) and allowing the probability of features being shared between primes and test alternatives to be very low ($\rho = .073$). These parameter values allowed ROUSE to produce no difference between the neither-primed and both-primed conditions but not a

Table 5
Parameters for the Fits of the Multinomial Model to the Data From Experiments 1, 2, 3, and 4, From Huber, Shiffrin, Lyle, and Ruys (2001) and the Flash-Time Experiment

Experiment and type of prime processing	p	p_c repetition	p_c associative	p_c orthographic	p_d
Experiment 1 passive	.692	.156	.015		
Experiment 2 passive	.832	.114			
Experiment 2 passive	.793			.159	
Experiment 3 passive (dissimilar test pair)	.749	.152			
Experiment 3 passive (ortho. related test pair)	.671	.187			
Experiment 4 passive	.770		.078		
Experiment 4 passive	.697			.147	
Experiment 1 active	.787	.121	.012		.153
Experiment 2 active	.793	.127			.144
Experiment 2 active	.808			.114	.097
Experiment 3 active (dissimilar test pair)	.799	.080			.072
Experiment 3 active (ortho. related test pair)	.709	.187			.072
Experiment 4 active	.765			.040	.100
Experiment 4 active	.814		.036		.000
Flash-time experiment (8 ms)	.628	.207			
Flash-time experiment (16 ms)	.811	.105			
Flash-time experiment (24 ms)	.890	.061			

Note. For the flash-time experiment, the value of p_c when $p = .5$ is $p_m = 0.278$; the values of p_c in the table are $2(1 - p)p_m$. In the active priming condition for orthographically related test alternatives there was an extra parameter, $p_t = .344$, which represented the probability of choosing the wrong test alternative to discount. ortho. = orthographically.

Table 6
Probability Correct: Data and Fits of the Multinomial Model for Experiments 2, 3, and 4 in Huber, Shiffrin, Lyle, and Ruys (2001)

Prime-test alternative or test alternative relationship	Fits from the multinomial model and data	Neither	Both	Target	Foil
	Experiment 2—passive				
Repetition	Theory	.832	**.760**	.851	.737
	Data	.833	.699	.860	.782
Orthographic	Theory	.793	.707	.826	.666
	Data	.766	.745	.845	.648
	Experiment 2—active				
Repetition	Theory	.793	.691	.701	.736
	Data	.780	.655	.722	.765
Orthographic	Theory	.808	.718	.749	.743
	Data	.802	.702	.759	.757
	Experiment 3—passive				
Repetition: Test choices dissimilar	Theory	.749	.679	.787	.635
	Data	.732	.705	.797	.624
Repetition: Test choices similar	Theory	.671	.613	.733	.545
	Data	.694	.602	.717	.542
	Experiment 3—active				
Repetition: Test choices dissimilar	Theory	.799	.735	.756	.754
	Data	.793	.719	.766	.768
Repetition: Test choices similar	Theory	.709	.618	.733	.581
	Data	.734	.609	.721	.568
	Experiment 4—passive				
Associative	Theory	**.770**	**.731**	.787	.712
	Data	.738	.766	.807	.696
Orthographic	Theory	.697	**.643**	.741	.594
	Data	.717	.591	.739	.618
	Experiment 4—active				
Associative	Theory	.814	.791	.820	.784
	Data	.811	.805	.821	.774
Orthographic	Theory	.765	.720	.697	.761
	Data	.757	.700	.710	.779

Note. The entries in bold are predictions that miss by more than 2 standard errors.

priming effect. The prediction of no difference resulted in a much lower contribution to chi-square compared with the multinomial model. If a semantic priming mechanism were added to the multinomial model, it could fit the data better, but more comprehensive experiments would be needed to test such an addition both for the multinomial model and for an equivalent addition to ROUSE.

Experiment 3, Huber, Shiffrin, Lyle, and Ruys (2001). In this experiment, the test alternatives were orthographically similar to each other or dissimilar (Table 1, 6th column). Different items were used in the two sets of conditions.

ROUSE uses the parameter ρ to represent the probability that test alternatives have a feature in common. We took the same tack for the multinomial model. In general, when subjects are discounting, they discount a test alternative that is similar to a prime word, so they reduce the difference in performance between the target-primed and foil-primed conditions. For example, in Experiment 3, probability correct in the target-primed and foil-primed conditions (repetition conditions) was .80 versus .62 with passive priming but .77 versus .77 with active priming. However, when the test alternatives are similar to each other, subjects may sometimes discount for the wrong test alternative. For example, if the test alternatives were *hail* and *hale* and only *hail* was a prime, subjects might choose against *hale* instead of *hail.* In this case, discounting would not reduce the difference in performance between the target-primed and foil-primed conditions as it usually does. This is what happens in Experiment 3: The probabilities correct for the target-primed and foil-primed conditions are .72 and .57, much the same as with passive priming. Following the reasoning just outlined, and like ROUSE's use of the ρ parameter, we introduced a new parameter (p_t) for the probability of discounting the wrong test alternative when the two test alternatives are similar to each other.

With dissimilar test alternatives, there were five parameters for the multinomial model, p and p_c for passive priming, p, p_c, and p_d for active priming (see Table 5). With similar test alternatives, there were three parameters, a value of p for passive priming conditions, a different value of p for active priming conditions, and one value of p_c for both the active and passive priming conditions (we set this parameter to be the same for these conditions to reduce the number of free parameters). The value of p_d for similar test alternatives was set to be the same as the value of p_d for dissimilar test alternatives, again to reduce the number of free parameters. With these parameters, the multinomial model fit the data with no deviations larger than two standard errors (see Table 6). The best-fitting value of the parameter p_t was .344. (We also fit the model with p_t to the data from the active priming condition with orthographically similar test items in Experiment 4, and p_t was estimated to be 0. This is similar to Huber, Shiffrin, Lyle, and Ruys's [2001] finding that ρ for orthographically similar test words was near 0, i.e., .070, in that experiment.)

For Experiments 2 and 3 combined, the chi-square goodness-of-fit values were 96.59 for ROUSE and 78.22 for the multinomial model, with a total of 17 different parameter values for ROUSE and 19 for the multinomial model.

Flash-time experiment. The experiment showed that p_c decreases as a function of flash time. The decrease in p_c was roughly linear with p, so we fit the data using the following equation: $p_c = 2(1 - p)p_m$, where p_m is the maximum value of p_c at $p = .5$. The model fits the data with all theoretical points falling within two standard errors of the data points ($SEM = .0177$; see Table 3 and the Results section). The values of p_c for the three flash times (see Table 5) were derived from the one value of p_m, and thus the 12 data points are fitted with four free parameters: three values of p and one value of p_m. We expect that ROUSE could also fit the data by allowing the probability of encoding features from the target (β) to vary with flash time.

A test of the multinomial model. The multinomial model predicts that the average of the probabilities of a correct response for two of the experimental conditions should be almost equal to the average of the probabilities of a correct response for two of the other conditions. Specifically, the average for the both-primed and neither-primed conditions should be within 1% of the average for the target-primed and foil-primed conditions. For passive priming, using the values of $P(\mathrm{T})$ in Figure 1, the average of the neither-primed condition and the both-primed condition minus the average of the target-primed condition and the foil-primed condition is $0.5(p - .5)p_c^2$. Using typical values of the parameters from Table 5, this is about .004; as predicted, less than 1%. For active priming, the expression for the difference between the averages is more complicated, but again using typical parameter values, it is about .006. The prediction is also confirmed when the data from the experiments and the fits to the data are examined: Across all four experiments from Huber, Shiffrin, Lyle, and Ruys (2001), for both passive and active priming conditions combined, the two averages are .722 and .731, respectively, for the data, and .730 and .724, respectively, for the fits of the multinomial model to the data, both differences of less than 1%. Averaging over flash times in the flash-time experiment, the two averages for the data are .742 and .758, and for the multinomial model fits, they are .750 and .749; again, the differences are less than 1%. Thus, at least for this one test, the multinomial model makes a reasonably strong prediction and the data are consistent with this prediction.

Discussion

The parameter values for the multinomial model fits to the data are displayed in Table 5. For most of the experiments, the parameters had different values for different groups of subjects and different groups of items, just as in ROUSE. Stronger tests of the models, looking for parameter invariances across manipulations, will require within-items and within-subjects experimental designs.

The parameters for the multinomial model have sensible values. The parameter p is simply the probability correct in the neither-primed condition. The parameter p_c is the probability of confusion, the probability that the decision is based on a prime instead of the flashed target. When a test alternative repeats a prime word exactly, the values of p_c for the various experiments range from .08 to .19. The values of p_c are about the same when a prime and test alternative are orthographically similar. This is not surprising, and the fits of ROUSE provide the same conclusion. When a prime and test alternative are associatively similar, the value of p_c is smaller, near 0 in most fits. The discounting parameter, p_d, ranges between about .07 and .15. This is about the same as the range of the confusion parameter, p_c, indicating that subjects are reasonably accurate at estimating the amount of their confusion and attempting to compensate for it, although they do not accurately identify on which trials discounting is needed and which it is not. Again, a similar conclusion is produced by the ROUSE model.

Overall, for the multinomial model, there were 4 misses out of 78 data points (see Table 6), near the expected chance level of .05. Two of these misses are for the facilitation effect with associative primes in the both-primed condition in Experiment 3, for which neither the multinomial model nor ROUSE can account. The ROUSE model has 3 out of 66 misses, again about the .05 level. In other words, both models account for the data about equally well and the obtained misses between theory and data are about what might be expected from chance.

ROUSE, the Multinomial Model, and Long-Term Priming

One of the future aims for ROUSE is to have it explain long-term priming effects in perceptual identification—that is, the effects that are observed when prime words precede flashed targets by delays of several minutes, filled with other words. ROUSE will then become a competitor to the REMI model (Schooler et al., 2001) within the Bayesian framework of this class of models, and it will become a competitor to the counter model (McKoon & Ratcliff, 2001; Ratcliff & McKoon, 1997). In contrast, the multinomial model simply adds two assumptions to existing models of perceptual identification, either REMI or the counter model. The short-term priming effects discussed in the current article are explained with a standard assumption about the loss of order and position information for items in short-term memory (Neath, 1998) and the assumption that subjects sometimes discount confusing items.

Compound Cue Models

Huber, Shiffrin, Lyle, and Ruys (2001) discuss the relationship between spreading activation models, compound cue models (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988), and the ROUSE model. ROUSE is not consistent with many spreading activation models for several reasons. In ROUSE, items are not kept separate from each other in memory and activation cannot flow from item to item along association chains.

The relationship between compound cue models and ROUSE is interesting. Compound cue models have usually been applied to paradigms in which two items, a prime and target, are presented in sequence, and a decision is required about the second item (e.g., lexical decision or item recognition). Compound cue models assume that the representations of the prime and target are combined in short-term memory to form a compound, in much the same way as features are combined in the test alternative vectors in ROUSE. In the standard paradigms to which compound cue models have been applied, information must be retrieved from memory. The compound formed in short-term memory is the probe to long-term memory, and the familiarity of the probe determines the decision. In the paradigm to which ROUSE is applied, retrieval from memory is not required for the task at hand—that is, to decide between the two test alternatives. For retrieval from long-term memory, the compound cue models calculate the familiarity of the probe in such a way that there is a boost to familiarity when the prime and target are associatively or semantically related. No such boost is required in the paradigm under discussion in this article. The decision between the two test alternatives does not require retrieval from memory and the data show no boost to performance for associatively related items.

Conclusions

The multinomial model is offered as a simple, commonsense alternative to the ROUSE model. The aim is to provide a competitor model to guide model testing and theoretical development. The fact that such a simple model can fit data that were generated to develop ROUSE suggests that the sophisticated mechanisms of ROUSE have not been fully exploited. On the one hand, ROUSE appears to be more flexible than the multinomial model, but on the other hand, it may turn out that experimental data will be as constrained as the multinomial model predicts.

References

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 2,* 191–211.

Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Quach, R. (2001). *Source confusion and discounting 2: Prime similarity and target duration.* Manuscript submitted for publication.

Huber, D. E., Shiffrin, R., Lyle, K., & Ruys, K. (2001). Perception and preference in short-term word priming. *Psychological Review, 108,* 149–182.

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Lee, C., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior, 16,* 395–418.

Masson, M. E. J., & MacLeod, C. M. (1996). Contributions of processing fluency to repetition effects in masked word identification. *Canadian Journal of Experimental Psychology, 50,* 9–21.

McKoon, G., & Ratcliff, R. (2001). Counter model for word identification: Reply to Bowers (1999). *Psychological Review, 108,* 674–681.

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

Neath, I. (1998). *Human memory: An introduction to research, data, and theory.* Pacific Grove, CA: Brooks/Cole.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308–313.

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

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104,* 319–343.

Ratcliff, R., & McKoon, G. (2000). Modeling the effects of repetition and word frequency in perceptual identification. *Psychonomic Bulletin and Review, 7,* 713–717.

Schooler, L., Shiffrin, R. M., & Raaijmakers, J. (2001). Theoretical note: A model for implicit effects in perceptual identification. *Psychological Review, 108,* 257–272.

Von Brisinski, I. S. (1994). Ultrafast display buildup with standard VGA on MS-DOS computers. *Behavior Research Methods, Instruments, & Computers, 26,* 335–336.

Received April 17, 2000
Revision received November 16, 2000
Accepted November 17, 2000 ■