Memory & Cognition
1995, 23 (5), 611-630

A multinomial modeling analysis of the recognition-failure paradigm

DAVID M. RIEFER
California State University, San Bernardino, California

and

WILLIAM H. BATCHELDER
University of California, Irvine, California

The recognition-failure paradigm has received much theoretical consideration, especially the Tulving-Wiseman function and its exceptions. We show that the Tulving-Wiseman function does a poor job of accounting for the data, both when its fit is measured with a model-based, goodness-of-fit statistic and when a logically equivalent reformulation of the function is compared with data. We then present a simple multinomial model based on retrieval-independence theory that is capable of measuring storage and retrieval processes in recognition failure. The model is used to conduct a meta-analysis of the recognition-failure paradigm, and shows that violations of the Tulving-Wiseman function occur under conditions in which weak storage is coupled with strong retrieval. In addition, if storage and retrieval are assumed to be positively correlated across conditions, the model produces a theoretically motivated, alternative equation to the Tulving-Wiseman function that provides a virtually identical fit to the data.

Memory theorists have long debated over the relation between recall and recognition, and in particular over what this relationship tells us about underlying memory processes. An important milestone in this debate was the discovery by Tulving and Thomson (1973) that recognition failure can occur for recallable words. This phenomenon is demonstrated using a special *recognition-failure* paradigm, in which weakly associated A-B pairs are presented for study. Memory is then tested by two successive procedures—a recognition test for the B-terms, followed by a cued-recall test of the B-terms, given the A-terms as cues. The data from the recognition-failure paradigm can be represented in the form of a 2 × 2 grid (see Table 1), where Rn indicates a correct response (hit) on recognition and Rc indicates a correct response on cued recall. The quantities N_{ij} $(i,j = 1,2)$ represent the response frequencies for each condition, aggregated over subjects and items. Tulving and Thomson used this paradigm to show that a substantial number of B-terms are not recognized, even when they are correctly recalled when their A-terms are given as cues. This result would be indicated by a nonzero value of N_{21} in the lower left cell of Table 1.

The recognition failure of recallable words is considered to be an important result because it apparently contradicts strong versions of generation-recognition theories of memory (Bahrick, 1970; Tulving & Wiseman, 1975). Such theories assume that successful recall of an item from memory is a two-stage process, whereby (1) a set of possible responses is generated from available memory cues, and (2) the correct response is recognized from this set. Recognition of the B-term is thus a subcomponent of successful cued recall, and therefore recognition failures of recallable items should not occur. Jones (1978) and Kintsch (1978) provide weaker versions of the generation-recognition theory that can account for these results; however, most contemporary theorists have abandoned the generation-recognition theory.

Since Tulving and Thomson's (1973) demonstration, recognition failure has proven to be a reliable phenomenon in a number of different experiments (see Nilsson & Gardiner, 1993). There has also been a considerable theoretical dialogue focusing on one aspect of the recognition-failure paradigm known as the Tulving-Wiseman function (Tulving & Wiseman,1975), which we will discuss shortly. But despite extensive experimentation with this paradigm, it is our contention that the literature on recognition failure has overlooked some basic theoretical and statistical issues, and instead has focused too strongly on complex theories and involved statistical artifacts centered on the Tulving-Wiseman function. In this article, we approach the issues in recognition failure from a different perspective. Our goal is to develop a very simple, multinomial processing-tree model (Hu & Batchelder, 1994; Riefer & Batchelder, 1988), and to show that this model is valid and can capture the main, reliable aspects of the recognition-failure paradigm viewed as a whole. We will demonstrate that the analysis of our multinomial model reveals new insights into the recognition-failure paradigm that have not been afforded

This research was supported by NSF Grant SBR-9309667. We thank Ece Kumbasar for her invaluable help on this project. We also thank Douglas Hintzman, Bennet Murdock, Richard Shiffrin, and John Wixted for their helpful review of an earlier version of this paper. Requests for reprints should be sent to D. Riefer, Department of Psychology, California State University, 5500 University Parkway, San Bernardino, CA 92407.

Copyright 1995 Psychonomic Society, Inc.

Table 1
2 × 2 Data Table From the Recognition-Failure Paradigm

	Recall	
Recognition	Rc	$\overline{\text{Rc}}$
Rn	N_{11}	N_{12}
$\overline{\text{Rn}}$	N_{21}	N_{22}

Note—Rc denotes recall, and Rn denotes recognition. The N_{ij} $(i, j = 1,2)$ represent the response frequencies in each condition.

by more traditional analyses. This is accomplished by conducting a model-based meta-analysis on the entire corpus of known experimental data from recognition-failure studies. Fortunately, Nilsson and Gardiner (1993) have recently documented 302 known data sets from 44 different studies that fit into this paradigm, making such an undertaking possible. Along the way, we will argue that the Tulving-Wiseman function does not fit the data well, and we provide a statistical explanation as to why others have come to the opposite conclusion.

Of course, one may well ask at this point why a new model is needed to study recognition failure—after all, a large number of memory models have already been applied to this area with some success. They include MINERVA (Hintzman,1987), CHARM (Metcalfe,1992), and SAM (Raaijmakers & Shiffrin,1981), to name a few. However, the approach we take in this article is different from that usually taken by others who analyze more complex models of memory. The usual strategy for testing complex models involves conducting a series of computer simulations and demonstrating that the models are capable of describing memory phenomena in a wide variety of different paradigms. By contrast, our approach involves using analytic methods to provide a precise and detailed statistical analysis of a very simple multinomial model applied specifically, and solely, to the recognition-failure paradigm. As Riefer and Batchelder (1988) point out, multinomial models are mathematical tools that can bridge the gap between strong theoretical models on the one hand and purely statistical approaches on the other. Because of their relative simplicity, these models are easy to work with analytically, and closed-form solutions for parameter estimates can often be obtained. Their statistical properties can also be explored—especially their robustness as approximations to more complete memory models.

We wish to make it clear at the outset that the statistical model we develop in this article is not meant as a competitor to other more sophisticated memory models that have already been extensively studied in the literature. Our model is psychologically based, but it is designed exclusively to understand data in the format of Table 1, and as such, it is not intended to be applied to a large array of potentially related memory phenomena such as false-alarm rates, confidence judgments, and forgetting functions. Rather than trying to extend our model to handle these and other basic effects, we will show that the model is a special case of a more complicated model, and that it can closely approximate other, more complex models when those models are restricted to the recognition-failure paradigm.

The article is organized as follows. First, we review the basic findings in the recognition-failure paradigm, and in particular, we focus on a major theoretical issue that has developed within this paradigm—the Tulving-Wiseman function and its so-called exceptions. We then show on statistical grounds that the Tulving-Wiseman function is not consistent with either the exceptions or the nonexceptions discussed in the literature. Following this, we present our multinomial model, and use the database from Nilsson and Gardiner (1993) to argue for its validity. Following that, we develop the model so that it can be used as a meta-analytic tool to provide a better understanding of data patterns across experimental conditions. In particular, we show that correlation between parameters across experiments can lead to a simplification of the model that helps explain the apparent good fit of the Tulving-Wiseman function. Finally we assess the impact of the model's simplifying assumptions on the analysis.

THE TULVING-WISEMAN FUNCTION

Tulving and his co-workers (e.g., Flexser & Tulving, 1978; Tulving & Wiseman, 1975) have argued that, within the recognition-failure paradigm, there is an orderly relationship across many studies between the proportion of recognition hits, P(Rn), and the conditional proportion of recognition hits given recall, P(Rn|Rc). This relationship is plotted in Figure 1, where each point represents one of the experimental conditions catalogued by Nilsson and Gardiner (1993) as conforming to the function (exceptions to the function are discussed later). As mentioned earlier, strong versions of generation-recognition theory predict that P(Rn|Rc) should be equal to one, because successful recall implies successful recognition. This is clearly not

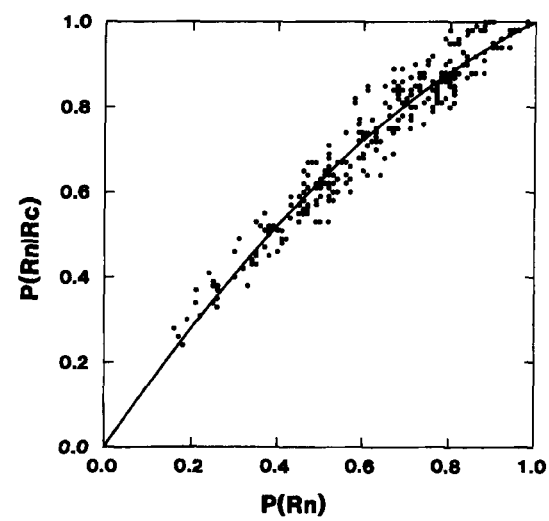

Figure 1. Relationship between probability of recognition, P(Rn), and probability of recognition given recall, P(Rn|Rc), plotted from Nilsson and Gardiner's (1993) database. The solid line through the points is the Tulving-Wiseman function from Equation 1. Only points that meet Nilsson and Gardiner's criterion for conforming to the function are plotted.

the case for the vast majority of studies represented in Figure 1. Instead, Tulving and Wiseman (1975) argued that the relation between $P(\text{Rn})$ and $P(\text{Rn}|\text{Rc})$ can be closely described by the following function:

$$P(\text{Rn}|\text{Rc}) = P(\text{Rn}) + c\,P(\text{Rn})\,[1 - P(\text{Rn})], \qquad (1)$$

with a value of the constant c approximately equal to .5. This function is represented by the solid line in Figure 1, and it expresses $P(\text{Rn}|\text{Rc})$ as the sum of a linear plus a quadratic component, both based on the recognition hit rate.[1]

Previous Analyses of the Function

Although some theorists (e.g., Nilsson & Gardiner, 1991; Nilsson, Law, & Tulving, 1988) have advocated that data conform so uniformly to the Tulving-Wiseman function that it should be labeled a psychological law, numerous exceptions to the function have been documented. According to Nilsson and Gardiner (1991, 1993), who have defined and catalogued them, these exceptions fall into two basic categories. One type consists of "encoding exceptions." These occur in conditions that lead to weak A-B encoding, such as those in which A-B pairs that are unrelated are used, or in which the rehearsal strategies that subjects are told to engage in are poor, leading to shallow levels of processing. The other type consists of "retrieval exceptions." These are created when the cues during recall contain much of the same information as the target items, which elevates the likelihood of successful retrieval given successful recognition. For example, Muter (1984) observed violations of the function when a semantic memory task was used in which the recall cues were detailed descriptions of famous last names (e.g., Nobel prize-winning physicist, Madame _____. Nilsson and Gardiner (1991) claim that by explicitly specifying the known circumstances that lead to violations, the boundary conditions within which the law is shown to hold are defined.

As Nilsson and Gardiner (1993) point out, explanations of the Tulving-Wiseman function need to account for violations of the function as well as for those points that conform to it. In general, attempts to explain the function and its exceptions have fallen into two categories: those that employ a strong, theoretical model of memory, and those that attempt to show that the function is largely a statistical artifact, thus meriting no special explanation. In the former case, a number of complex, formal models have been proposed, including ACT* (Anderson, 1983), MINERVA 2 (Hintzman, 1987), and CHARM (Metcalfe, 1992; see Nilsson & Gardiner, 1991; Ratcliff & McKoon, 1989, for reviews). As we indicated earlier, all of the above models are comprehensive models of human memory, designed to handle a wide range of memory phenomena in addition to the Tulving-Wiseman function. Because of their complexity, the basic approach used to test models of this type has involved conducting a few computer simulations and showing, under certain assumptions and parameter values, that the model is capable of producing simulated data that mimic the general shape of the Tulving-Wiseman function, rather than analytically deriving an actual relationship between $P(\text{Rn})$ and $P(\text{Rn}|\text{Rc})$ implied by the model.

A different type of account for the Tulving-Wiseman function has been proposed by Hintzman (1991, 1992, 1993), who argues that the orderliness of the function is a statistical artifact. Hintzman claims that mathematical constraints among the values of $P(\text{Rn}|\text{Rc})$, $P(\text{Rn})$, and $P(\text{Rc})$, principally when recall probability exceeds recognition probability, help to constrain data points into a region that is consistent with the function. Furthermore, he argues that parameter variability within each experimental condition, such as that caused by subject, item, or even subject–item differences, may also constrain data points to conform to the function. Hintzman's concerns have been challenged by Tulving and Flexser (1992, 1993), but the issues raised by Hintzman's critiques are important ones. It is clearly an approximation to assume that the subject-items that go into the 2 × 2 grid in Table 1 are independent and identically distributed.[2]

Alternative Analyses of the Function

The literature on the recognition-failure paradigm reviewed above has focused almost exclusively on the Tulving-Wiseman function's failures and successes in representing the paradigm. This is unfortunate because, as we show next, the apparent fit of the function is not substantiated when the data are transformed and examined in different forms. For example, consider the data quantity

$$\text{D} = P(\text{Rc}|\text{Rn}) - P(\text{Rc}|\overline{\text{Rn}}), \qquad (2)$$

which can be viewed as a measure of learning on the recognition test (see Humphreys & Bowyer, 1980). It is easy to use the laws of probability to derive

$$\text{D} = \frac{P(\text{Rn \& Rc}) - P(\text{Rn})\,P(\text{Rc})}{P(\text{Rn})\,[1 - P(\text{Rn})]}. \qquad (3)$$

Next, note from Equation 1 that

$$P(\text{Rn \& Rc}) = P(\text{Rn})\,P(\text{Rc}) + c\,P(\text{Rn})\,[1 - P(\text{Rn})]\,P(\text{Rc}), \qquad (4)$$

which is an equivalent expression of the Tulving-Wiseman function created by multiplying both sides of Equation 1 by $P(\text{Rc})$. This can be substituted into Equation 3, and the result is

$$\text{D} = c\,P(\text{Rc}). \qquad (5)$$

In other words, the Tulving-Wiseman function exactly fits the data in an experimental condition if and only if Equation 5 holds exactly.

To test the relationship in Equation 5, we examined those conditions in Nilsson and Gardiner (1993) that the authors catalogued as nonviolations of the Tulving-Wiseman function, and we constructed the scatterplot between D and $P(\text{Rc})$ for those conditions. This is presented in Figure 2. Equation 5 is represented by the line through the points in Figure 2, based on a least-squares estimate of $c = .476$. This equation yielded a correlation between predicted and ob-

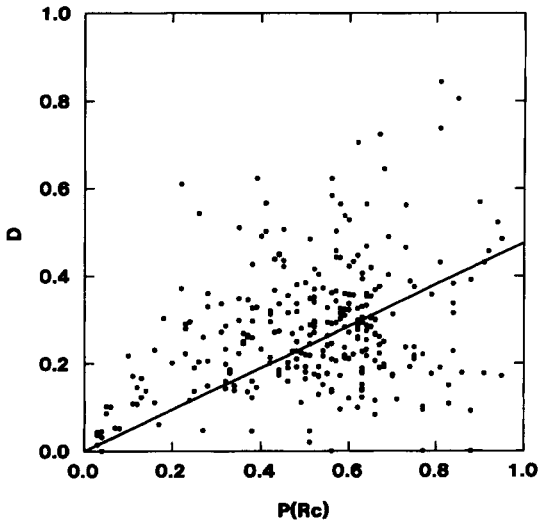

Figure 2. Relationship between probability of recall, P(Rc), and D, which equals P(Rc|Rn) – P(Rc|$\overline{\text{Rn}}$), plotted from Nilsson and Gardiner's (1993) database. The line through the points is the best-fitting line from Equation 5, with c = .476.

served values of D of r = .235. While the best-fitting value of c in Figure 2 is close to the traditional value of .5, the proportion of variance explained (r^2 = .055) is very low compared with the value obtained for the Tulving-Wiseman function in Figure 1 (r^2 = .916). This is an important result, because it suggests that the Tulving-Wiseman function is merely a convenient curve-fitting function for a very particular way of plotting the data across conditions, as in Figure 1. When the very same data are plotted in a different way, the function in Equation 5, derived directly from the Tulving-Wiseman function, appears to do a poor job of fitting the data. Thus we have an apparent paradox—namely, that two theoretical relationships can be mutually consistent, yet the data may appear to fit one of them much better than they fit the other.

Fortunately, it is not hard to understand this apparent paradox. Suppose that an experimental condition yields data that fail to fit the Tulving-Wiseman function by a certain amount of error, denoted by e. We can represent this situation by the equation

$$P(\text{Rn}|\text{Rc}) = P(\text{Rn}) + c\,P(\text{Rn})\,[1 - P(\text{Rn})] + e, \quad (6)$$

where e is a nonzero value consistent with the constraints of probability theory. Similarly, Equation 5 can be expressed with its own error term, e^*:

$$\text{D} = c\,P(\text{Rc}) + e^*. \quad (7)$$

From Equations 2–4, we can compute an expression for e^* in terms of e given by

$$e^* = \left[\frac{P(\text{Rc})}{P(Rn)[1 - P(Rn)]}\right] e. \quad (8)$$

Note that P(Rn) [1 − P(Rn)] has a maximum possible value of .25, and Nilsson and Gardiner's (1993) database shows that most studies on the recognition-failure paradigm yield values of P(Rc) that are greater than .25 (cf. Figure 4). Thus e^* will generally be larger than e, and sometimes by a great amount. Deviations from the Tulving-Wiseman fit in Figure 1 will therefore tend to be magnified when the equivalent relationship from Equation 5 is used, as can be seen in Figure 2. In general, we can expect that the data will fit Equation 5 more poorly if the measure of fit is a simple error-variance measure, rather than a model-based goodness-of-fit statistic.

Fortunately, it is possible to develop a model-based goodness-of-fit statistic for the Tulving-Wiseman function. One aspect of the function that has generally been overlooked is that it can actually be viewed as a statistical model—specifically, a model for the nonindependence in an aggregate data table such as Table 1. To see this, note that Equation 4 gives an expression for P(Rn & Rc), which is the probability in the upper left cell of Table 1. The relationship in Equation 4 therefore implies that the cell frequencies in Table 1 can be reproduced, within sampling variability, by a knowledge of the two marginal probabilities P(Rn) and P(Rc), along with the value of c.

We can analyze data in the form of Table 1 in a manner similar to a chi-square test of independence, except that the expected frequencies are computed from Equation 4 rather than from an independence assumption. First, the two marginals are estimated by

$$\hat{p}(\text{Rn}) = \frac{N_{11}+N_{12}}{N} \quad (9)$$

and

$$\hat{p}(\text{Rc}) = \frac{N_{11}+N_{21}}{N}, \quad (10)$$

where $N = N_{11} + N_{21} + N_{12} + N_{22}$. Next, the expected frequencies are generated from Equation 4 by

$$F_{11} = N\hat{p}(\text{Rn})\hat{p}(\text{Rc})\,[1 + c - c\hat{p}(\text{Rn})] \quad (11a)$$

$$F_{12} = N\hat{p}(\text{Rn}) - F_{11} \quad (11b)$$

$$F_{21} = N\hat{p}(\text{Rc}) - F_{11} \quad (11c)$$

$$F_{22} = N\,[1 - \hat{p}(\text{Rn})] - F_{21}. \quad (11d)$$

Finally, a chi-square test can be conducted by the formula

$$\chi^2(1) = \sum_{i,j=1}^{2} \frac{(F_{ij} - N_{ij})^2}{F_{ij}}. \quad (12)$$

We analyzed data in the form of Table 1 for all the data sets in Nilsson and Gardiner (1993) that were classified by the authors as nonexceptions and that had expected frequencies of five or more in each cell. The results of this analysis were enlightening. There were a total of 253 cases from Nilsson and Gardiner that met the above restrictions. Of these, a full 35% produced significant deviations of observed frequencies from predicted frequencies from the Tulving-Wiseman function, using a significance level of α = .05. At α = .01, this percentage was still 24%. This seems to us to be a fairly substantial percentage of cases

for which the function performs poorly, especially considering that these cases were all classified as nonexceptions by Nilsson and Gardiner. Moreover, the average χ^2 value for these 253 cases is 5.19, which is well above the expected value of 1.00 for a random χ^2 variable with one degree of freedom.[3]

We think that the results of the χ^2 tests are cause to statistically reject the Tulving-Wiseman function as a valid representation of the data in the recognition-failure paradigm. In light of these results, it is useful to revisit Hintzman's (1992, 1993) analyses of the function. Hintzman is correct in showing that, under certain circumstances, constraints among P(Rn), P(Rc), and P(Rn|Rc) can reduce the error in fitting the function and help to create a positive correlation over conditions, as seen in Figure 1. However, we argue that these constraints only pose a problem if one evaluates the Tulving-Wiseman function with a measure that is not model-based, such as by "eyeballing" its fit or conducting correlation tests. In contrast, the χ^2 test is based on known statistical sampling theory, and it reflects constraints of the type that Hintzman discusses through the expected and joint probabilities of each of the four cells in Table 1. Relationships such as that obtained by Equation 1 basically reduce the data structure by one degree of freedom and, as such, are perfectly plausible statistical models for 2 × 2 data tables. By their very nature, such relationships will entail constraints on the data. However, in the case of the recognition-failure paradigm, the relationship proposed by Tulving and Wiseman (1975) is not supported when the resulting model is subjected to a valid statistical test.

As mentioned earlier, Hintzman (1980,1991,1993) has also raised concerns about applying models to aggregate data in the form of Table 1 when they entail the assumption that the observations are independent and identically distributed. The χ^2 test that we adopt does in fact make these assumptions. However, as with other applications of χ^2 tests to contingency tables in the social and behavioral sciences (see, e.g., Bishop, Fienberg, & Holland, 1975), one has reason to hope that the test is robust under small violations of these assumptions. In any event, it is up to proponents of the Tulving-Wiseman function to show that the failure of the χ^2 tests to support the function can be explained by the fact that cell entries within a condition are not independent and identically distributed. Ironically, this situation is just the opposite of that proposed by Hintzman in his argument that these assumptions may have resulted in spurious support for the function.

A Multinomial-Modeling Analysis

It is evident from the above analysis that research on the recognition-failure paradigm has focused too narrowly on correlational tests of just one aspect of the data in Table 1—namely, the Tulving-Wiseman function. As we have indicated, this is not an appropriate way to test a statistical model of the entire data structure. Such an approach can be very deceptive, as revealed by the bad fit when the same aspect of the data is plotted against a completely equivalent relationship, such as that given by Equation 5 and Figure 2.

It is clear that alternative approaches can be useful in exploring aspects of the recognition-failure paradigm that have not been revealed in the more traditional analyses based on the Tulving-Wiseman function. In the remainder of this article, we will explore one such approach, based on a multinomial-modeling analysis. Specifically, our goal is to develop a simple model that incorporates the retrieval-independence theories of Begg (1979) and Flexser and Tulving (1978), and to use this model to conduct a comprehensive meta-analysis of the database in Nilsson and Gardiner (1993). We will use the database to argue for the validity of the model, as well as to understand both the exceptions and nonexceptions to the Tulving-Wiseman function. We will also show that a reduced version of the model that incorporates parameter covariation across experimental conditions in a particular way can explain some of the apparent success of the Tulving-Wiseman function.

THE MULTINOMIAL MODEL

The multinomial model presented next is consistent with a commonly held view of recognition and cued recall called *retrieval-independence theory*. Retrieval independence assumes that the underlying processes of recognition and cued recall are independent of each other, in the sense that they involve different cognitive mechanisms. Thus, the success or failure of one process does not affect the likelihood of success for the other. Dependencies in empirical measures of recall and recognition occur because both processes access a common memory trace that varies probabilistically in encoding strength (Flexser & Tulving, 1978) or likelihood of retention (Begg, 1979).

Data Representation

The basic experimental task behind the model is the recognition-failure paradigm described earlier in Table 1. Subjects can be successful or unsuccessful on either of these memory tests, resulting in four data events: E_{11}—successful recognition and recall (Rn & Rc); E_{12}—successful recognition but unsuccessful recall (Rn & $\overline{\text{Rc}}$); E_{21}—unsuccessful recognition but successful recall ($\overline{\text{Rn}}$ & Rc); and E_{22}(unsuccessful recognition and recall ($\overline{\text{Rn}}$ & $\overline{\text{Rc}}$). The quantities N_{ij} are the frequencies for each E_{ij} data event.

Model Development

The multinomial model assumes that subjects' recognition and recall performance for an A-B pair is a function of four hypothetical, dichotomous processes:

Storage of the A-B pair. A representation of an A-B pair is assumed to be stored within memory and maintained to the time of testing with probability s, $0 \leq s \leq 1$.

Retrieving stored A-B pairs during recognition. During the recognition test, when subjects are presented with B-item probes, the model assumes that a stored A-B pair is accessed with probability r_1, $0 \leq r_1 \leq 1$. This results in correct recognition of the B item.

Retrieving stored A-B pairs during recall. During the cued-recall test, when subjects are presented with A-item probes, the model assumes that if there is a repre-

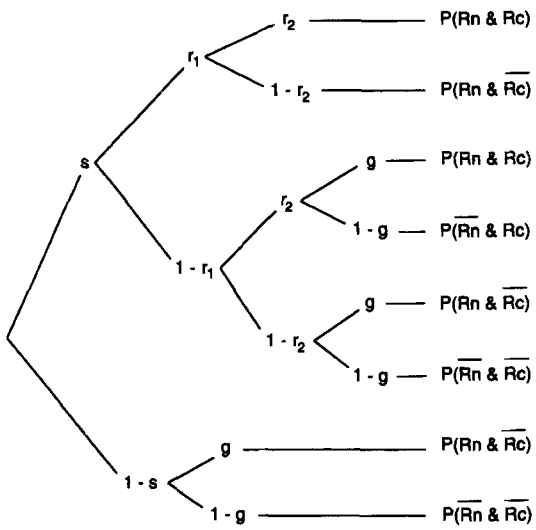

Figure 3. Multinomial processing-tree model based on retrieval-independence theory.

sentation of A-B in memory, the cue provides successful retrieval of its target with probability r_2, $0 \leq r_2 \leq 1$. This results in correct recall of the B item.

Guessing on recognition. If an A-B pair is not stored or retrieved on recognition, a correct recognition is still possible with probability g, $0 \leq g \leq 1$.

Multinomial models such as the one above are sometimes referred to as *general processing-tree models* (Hu & Batchelder, 1994), because they can be expressed as a tree diagram with a special parametric form. Figure 3 presents the tree structure for the current model. The model in Figure 3 is similar to a model for proactive inhibition developed by Greeno, James, DaPolito, and Polson (1978), and presented as a general processing-tree model in Riefer and Batchelder (1988, Figure 1). Greeno et al. applied their model to a proactive-inhibition paradigm, in which a list of A-B associates is presented for learning, followed by a second list with the same stimuli paired with new responses (A-C). Subjects then attempt to recall both B and C, when given A as a cue. Retrieval independence is also an issue in interference theory, because the recall of B and C is often stochastically independent even when strong interference effects are evident—a finding known as the *retrieval-independence phenomenon* (Martin, 1971). As it turns out, if $g = 0$ in Figure 3, then our retrieval-independence model is equivalent to the Greeno et al. model in terms of its basic structure, although, of course, the two models postulate a different interpretation of the parameters.

From the tree structure in Figure 3, it is straightforward to write theoretical expressions for the probabilities of each data event:

$$p(\text{Rn \& Rc}) = s\, r_1\, r_2 + s\,(1 - r_1)\, r_2\, g \tag{13a}$$

$$p(\text{Rn \& } \overline{\text{Rc}}) = s\, r_1\,(1 - r_2) + s\,(1 - r_1)\,(1 - r_2)\, g + (1 - s)\, g \tag{13b}$$

$$p(\overline{\text{Rn}} \text{ \& Rc}) = s\,(1 - r_1)\, r_2\,(1 - g) \tag{13c}$$

$$p(\overline{\text{Rn}} \text{ \& } \overline{\text{Rc}}) = s\,(1 - r_1)\,(1 - r_2)\,(1 - g) + (1 - s)\,(1 - g). \tag{13d}$$

If $g = 0$, these expressions reduce to the same structure as Equation 31 in Riefer and Batchelder (1988) for the Greeno et al. (1978) model.

In the theoretical literature on the recognition-failure paradigm, the possibility of a correct response on recognition due to guessing is not often discussed. For example, in Flexser and Tulving's (1978) model, failure to adequately store the A-B pair necessarily results in a recognition failure. The assumption of no guessing on recognition is consistent with only a few studies that have used stringent recognition criteria (e.g., Gardiner, 1988; Nilsson et al., 1988); however, there are many recognition paradigms in which such guessing processes are likely to occur. One method for examining these guessing processes is to look at false-alarm responses to distractor items. It is straightforward to expand the data to include distractors, and then it would be necessary to expand the model in Figure 3 to include a tree for distractor items. There are several ways to incorporate a model for distractors; however, because data for distractors are not reported in many of the articles on the recognition-failure paradigm, we hereafter concentrate only on the model in Figure 3. Further, to simplify our analyses, we will also assume that $g = 0$ in the model. This approximation not only makes the model identifiable for a given condition (with three parameters and three degrees of freedom), it also substantially facilitates the analyses to follow. Later, we will discuss the impact of $g > 0$ on our conclusions, as well as the issue of completing the model to include distractors.

Parameter Estimation

With $g = 0$, there are three parameters in the model, and the data structure in Table 1 contains three degrees of freedom. Riefer and Batchelder (1988, p. 327) derive closed-form maximum-likelihood estimates (MLEs) for the parameter values of the model. They are

$$\hat{s} = \frac{(N_{11}+N_{12})(N_{11}+N_{21})}{N_{11}N} \tag{14a}$$

$$\hat{r}_1 = \frac{N_{11}}{N_{11}+N_{21}} \tag{14b}$$

$$\hat{r}_2 = \frac{N_{11}}{N_{11}+N_{12}}. \tag{14c}$$

These MLEs are valid as long as $N N_{11} > (N_{11} + N_{12})(N_{11} + N_{21})$. If this inequality does not hold, it is possible to use a computer search method developed by Hu and Batchelder (1994) to find MLEs within the unit interval.

It is also possible to express the estimates in Equation 14 in terms of the three data statistics provided in the Nilsson-Gardiner (1993) database—namely, $P(\text{Rn})$, $P(\text{Rc})$, and $P(\text{Rn}|\text{Rc})$. The resultant equations are

$$\hat{s} = \frac{P(\mathrm{Rn})}{P(\mathrm{Rn}|\mathrm{Rc})} \tag{15a}$$

$$\hat{r}_1 = P(\mathrm{Rn}|\mathrm{Rc}) \tag{15b}$$

$$\hat{r}_2 = \frac{P(\mathrm{Rc})\,P(\mathrm{Rn}|\mathrm{Rc})}{P(\mathrm{Rn})}. \tag{15c}$$

Equation 15 provides MLEs for the parameters within the open interval (0,1), provided that $P(\mathrm{Rn}|\mathrm{Rc}) > P(\mathrm{Rn})$.

Relation to Other Models

The multinomial model just presented attempts to represent the basic assumption of retrieval independence in a simple mathematical form. The processes of recognition and recall are independent in the model because, given storage of A-B, the probability of recall retrieval (as measured by r_2) does not depend on whether recognition retrieval occurred (as measured by r_1). Certain strong theoretical models of memory, when applied to the recognition-failure paradigm, can also be thought of in this framework. For example, Metcalfe's (1992) CHARM model assumes that A-B pairs are stored in a single composite memory trace that is accessed by an A-vector for cued recall, and separately by a B-vector for recognition. SAM (Raaijmakers & Shiffrin, 1981) also incorporates retrieval independence, because cued recall occurs through a sequential search and recognition occurs through global activation, both accessing the same stored strengths.

The one theory that perhaps bears the closest resemblance to our current model is Flexser and Tulving's (1978) goodness-of-encoding model. In fact, our model (with $g = 0$) can be seen as a special case of Flexser and Tulving's model. Their model represents memory traces as feature vectors, and the multinomial model is equivalent to the simplest case in their model, in which vectors consist of a single component. In their general model, integer thresholds on the number of component matches to achieve successful recognition and cued recall are also parameters. In our special case, the threshold parameters would be set to one, which represents the number of components.

Model Validity

An important step in the development of any model consists of assessing its validity. One way to accomplish this is to determine that the parameter estimators of the model reasonably measure the processes that they are designed to represent. This typically involves conducting a series of experiments, manipulating relevant independent variables, and seeing whether these variables influence the parameter values in a logical or theoretically reasonable manner. Fortunately, it is not necessary to conduct new experiments to test the validity of the current model, because the large array of previously published research on the recognition-failure paradigm provides an ample corpus of data for this task.

As mentioned previously, Nilsson and Gardiner's (1993) database contains the aggregated values of $P(\mathrm{Rn})$, $P(\mathrm{Rc})$, and $P(\mathrm{Rn}|\mathrm{Rc})$ for each of their 302 conditions, which can be inserted directly into Equation 15 to derive estimates of the parameters. For convenience, we have computed the estimates of s, r_1, and r_2 for each of these data sets, and they are presented in the Appendix. Each set of estimates is catalogued using the same condition number as used in Nilsson and Gardiner (1993). One advantage of multinomial modeling is that approximate asymptotic confidence intervals can also be computed for each parameter, using methods described in Riefer and Batchelder (1988, pp. 327–328). The 95% asymptotic confidence intervals for each parameter are also included in the Appendix.

With these parameter estimates, we can first explore the relationship between recall and recognition. To start with, Figure 4 plots the relation between $P(\mathrm{Rn})$ and $P(\mathrm{Rc})$ for the data sets from Nilsson and Gardiner (1993). As can be seen, both recognition and recall exhibit a fairly wide range of values across the 302 known data sets. It is important to realize that in experiments using the recognition-failure paradigm, recognition hit rates are obtained using a number of different procedures, including free versus forced-choice paradigms and liberal versus stringent scoring criteria. This helps to explain why recognition hit rates in Figure 4 cover such a wide range.

As it turns out, there is, in fact, a significant, positive correlation between $P(\mathrm{Rn})$ and $P(\mathrm{Rc})$ in Figure 4 [$r(300) = .324$, $p < .001$]. Considering the number of unrelated design variables that vary over the 302 conditions, the value of this correlation seems to us to be fairly large. As stated earlier, retrieval-independence theory explains any empirical dependency between these two measures by assuming that the underlying processes of recognition and recall are independent, but that they both access a common memory trace. These cognitive processes are represented in the model by parameters r_1 (recognition), r_2 (recall), and s (storage). Thus, the model would explain the positive correlation in Figure 4 by assuming that $P(\mathrm{Rn})$ and $P(\mathrm{Rc})$ are "spuriously correlated" because the storage

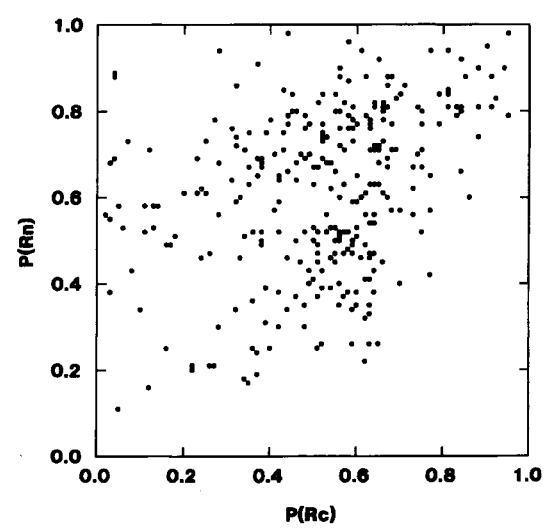

Figure 4. Relationship between probability of recall, $P(\mathrm{Rc})$, and probability of recognition, $P(\mathrm{Rn})$, plotted from Nilsson and Gardiner's (1993) database.

parameter s varies over conditions. Fortunately, r_1 and r_2 are designed to measure the conditional processes of recognition and recall given successful storage, in essence factoring out the common-storage process. Thus, if the multinomial model is a valid representation of retrieval independence, we would expect to see no significant correlation between these two parameters. Figure 5 is a scatterplot of the values of $\hat{r}_1$ and $\hat{r}_2$ from the Appendix for all 302 conditions in Nilsson and Gardiner's (1993) database. As predicted from retrieval-independence theory, there is essentially no correlation across conditions between the two conditional measures of recognition and recall [$r(300) = -.066$, n.s.].

Next, we explore the effect of different independent variables on the values of the model's parameters. The 44 different studies represented in Nilsson and Gardiner's (1993) database examine a wide range of different experimental manipulations, many of which can be expected to have systematic effects on storage, recognition, or recall. The question is whether these experimental variables affect the values of s, r_1, and r_2 in a manner that is psychologically reasonable. For example, certain variables can be expected to have their primary effect on storage, while others should affect retrieval; or some variables might be hypothesized to affect retrieval on recognition, but not retrieval on recall.

We examined this issue by searching the Nilsson and Gardiner (1993) database for studies relevant to this analysis. For each study, we attempted to identify independent variables for which clear hypotheses could be developed, on the basis of logical principles or contemporary memory theory. In particular, we concentrated on four different types of variables: (1) those that should affect storage and maintenance capacity (s); (2) those that should affect retrieval capacity (r_1 and r_2); (3) those that should affect recognition (r_1) but not recall (r_2); and (4) those that should affect recall (r_2) but not recognition (r_1). We were able to identify a sizable set of independent variables falling into these various categories (see Table 2, which also lists the specific studies that examined each variable).

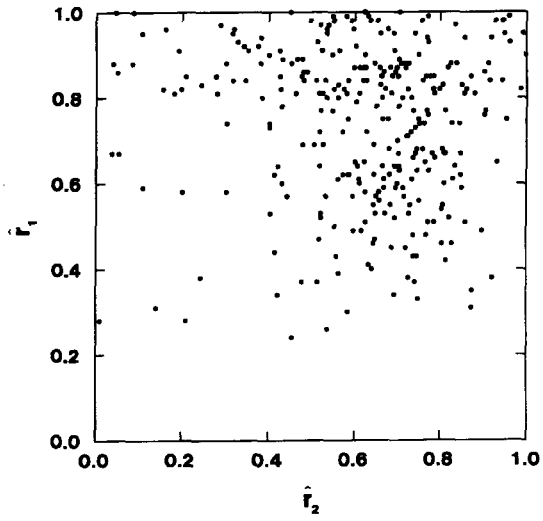

Figure 5. Relationship between $\hat{r}_1$ and $\hat{r}_2$, parameter estimators of the multinomial model that measure recognition retrieval and recall retrieval, respectively.

For all of these variables, we conducted hypothesis tests of the model's parameters, comparing the relevant experimental conditions from each study. Hypothesis testing for general processing-tree models is straightforward, and is discussed in detail in Riefer and Batchelder (1988, 1991b) and Hu and Batchelder (1994). Basically, hypothesis tests can be conducted using the log-likelihood ratio statistic G^2, which is distributed approximately as a χ^2 variable. Except where indicated, all tests used a significance level of $\alpha = .01$. The results of these analyses are presented next. Space limitations prevent us from presenting detailed results for every hypothesis test we conducted; instead, we present a substantial subset of our findings that is representative of the overall pattern of results.

Variables that should affect storage and maintenance. There are a number of experimental manipulations whose primary effect can reasonably be expected to be one of influencing how well items are stored and maintained within memory. A straightforward example would be the depth of encoding given to each item. For example, Bryant (1991) instructed subjects to process items using either structural, phonetic, or semantic encoding (Cases 281, 282, and 283, respectively, in the Appendix). Analysis of Bryant's data by the model reveals that this manipulation did have a significant effect on storage, [$G^2(2) = 21.50$], with deeper processing resulting in larger values of s. Depth of encoding had no significant effect on r_1 [$G^2(2) = 2.18$], but did have a marginal effect on r_2 [$G^2(2) = 7.27, p < .05$].

Other experiments have manipulated depth of processing by examining meaningful versus rote rehearsal. A typical study was conducted by Begg (1979, Experiment 1, Case 119 vs. Case 120 in the Appendix). An analysis of these data shows that meaningful rehearsal resulted in significantly better storage than did rote rehearsal [$G^2(1) = 42.78$]. As with Bryant's (1991) study, this variable had no significant effect on r_1 [$G_2(1) = 2.16$], but did have a significant effect on r_2 [$G_2(1) = 91.18$]. Studies by Nilsson et al. (1988) and Fisher (1979) show similar patterns of results.

Another variable that should have a strong effect on storage capacity is the number of study presentations for each item. This was examined by Arlemalm and Nilsson (1992), who presented items for either one or five presentations (Cases 297, 298, 301, and 302 vs. Cases 295, 296, 299, and 300 in the Appendix). As expected from the model, this had a significant effect on storage, with five presentations resulting in higher values of s [$G^2(4) = 122.98$]. Number of presentations also significantly improved recognition retrieval (r_1), as well as recall retrieval (r_2) [$G^2(4) = 27.23$ and 91.80, respectively].

Storage capacity, as measured by the model, should also be adversely affected if there is a delayed test of memory. This is because the parameter s measures the probability that items remain stored at the time of testing, which should be affected by when the test is administered. This variable was examined by Donnelly (1988), who tested subjects either immediately or after delays of 3 days or 2 weeks

Table 2
Independent Variables Examined to Test the Validity of the Multinomial Model

Variables that should affect storage and maintenance	
Delayed testing	Tulving & Watkins (1977)
	Nilsson & Shaps (1980, Experiment 3)
	Donnelly (1988, Experiments 1, 2)
	Ronnberg, Lyxell, Samuelsson, Erngrund, & Nilsson (1991, Experiments 6,7)
Depth of encoding	Begg (1979, Experiment 3)
	Nilsson, Law, & Tulving (1988)
	Bryant (1991, Experiment 1)
Number of presentations	Sandberg (1990)
	Arlemalm & Nilsson (1992)
Rote versus meaningful instructions	Begg (1979, Experiments 1, 2)
	Fisher (1979, Experiment 2)
Variables that should affect retrieval	
Change in context	Tajika (1978)
	Tajika (1979, Experiments 1, 2)
Young versus elderly	Shaps & Nilsson (1980, Experiments 1, 2)
	Rabinowitz (1984)
Variables that should affect recognition retrieval	
Free versus forced-choice recognition	Watkins & Tulving (1975, Experiments 1–6)
	Wiseman & Tulving (1976, Experiments 3,4)
	Jones & Gardiner (1990, Experiment 1)
Frequency of targets	Reder, Anderson, & Bjork (1974, Experiment 2)
Generated versus nongenerated targets	Watkins & Tulving (1975, Experiment 4)
	Wiseman & Tulving (1975, Experiment 1)
Strict versus lenient scoring	Gardiner (1988, Experiment 2)
Variables that should affect recall retrieval	
Level of cue–target association	Wiseman & Tulving (1976, Experiments 1–4)
	Tajika (1977)
	Tajika (1978)
	Fisher (1979, Experiment 2)
	Vining & Nelson (1979, Experiment 1)
	Shaps & Nilsson (1980, Experiments 1, 2)
	Neely & Payne (1983)
	Bryant (1991, Experiment 2)

(Cases 222, 223, and 224 in the Appendix). Delayed recall did in fact result in significantly poorer storage capacity (s) [$G^2(2) = 30.84$]. Not surprisingly, delayed testing also had an effect on subjects' ability to retrieve items from memory. Longer delays resulted in significantly poorer retrieval for both recognition and recall [$G^2(2) = 124.45$ and 49.85, respectively].

In general, the model demonstrates that experimental manipulations occurring during the study phase and before the test phase, such as memory instructions, number of presentations, and delayed testing, have a predictable and systematic effect on storage capacity. As it turns out, some of these manipulations also had an effect on retrieval capacities. For example, encoding instructions affected not only storage, as measured by the parameter s, but also retrieval during recall, as measured by the parameter r_2. Considering that r_2 reflects the ability of the A-term to cue the recall of the B-term, encoding manipulations could very well affect r_2 because they strengthen the A-B association. Thus it is not unreasonable to expect that experimental manipulations during study might have some bearing on eventual retrieval processes, in addition to their expected effects on storage.

Variables that should affect retrieval. In contrast to variables that affect storage, it is reasonable to expect that experimental variables exist whose primary effect is on retrieval capacities (r_1 and r_2). A good example comes from an experiment by Tajika (1979), in which memory for paired associates was tested in different or same contexts (Cases 152 and 153 vs. 154 and 155 in the Appendix). The model reveals that testing in different contexts resulted in significantly poorer retrieval, as measured both by r_1 [$G^2(2) = 72.37$] and r_2 [$G^2(2) = 22.05$]. However, this manipulation had no significant effect on storage [$G^2(2) = 3.24$]. From the point of view of the model, this is a reasonable result, considering that in the experiments, the switch in testing context occurred after the material was presented for study.

Retrieval differences should also be observed between young and elderly adults. A number of prior studies using paradigms other than recognition failure (Erdfelder & Bayen, 1991; Riefer & Batchelder, 1991a; Schonfield & Robertson, 1966) have concluded that the elderly exhibit retrieval deficits in memory. Within the recognition-failure paradigm, this variable has been examined in a study by Rabinowitz (1984; Cases 202 and 203 in the Appendix).

Consistent with other research and theory, older adults exhibited poorer retrieval capacity for both r_1 [$G^2(1)$ = 11.68] and r_2 [$G^2(1) = 51.26$]. However, young and elderly adults showed no difference in their storage capacities [$G^2(1) = 0.00$]. A similar pattern of results was observed when we analyzed data from Shaps and Nilsson (1980).

Variables that should affect recognition but not recall. A number of studies in the recognition-failure paradigm have specifically manipulated the nature of the recognition task. It is reasonable to expect that these manipulations should have an effect on the recognition parameter, r_1, but not on the recall parameter, r_2. This prediction was borne out by the model's analysis. For example, a number of researchers (e.g., Gardiner, 1988; Nilsson et al., 1988) have used strict versus lenient scoring when measuring recognition performance. Not surprisingly, an analysis of Gardiner's (1988) data (Cases 230, 231, and 232 vs. Cases 233, 234, and 235 in the Appendix) shows that strict scoring on recognition produced significantly lower estimates of r_1 than did lenient scoring [$G^2(3) = 21.20$]. Also as expected, this manipulation, which simply involves a reanalysis of the same recognition data, had no significant effect on the storage parameter, s [$G^2(3)$ = 5.46] or on the recall parameter, r_2 [$G^2(3) = 1.17$].

Still other researchers (e.g., Jones & Gardiner, 1990; Watkins & Tulving, 1975) have examined free versus forced-choice recognition-testing procedures. A study by Wiseman and Tulving (1976, Experiment 3) provides a typical example (Cases 62 and 63 in the Appendix). In their study, targets and distractors were arranged on response sheets in rows, with each row containing one target and two distractors. The subjects were first allowed to circle any items that they recognized as old (free recognition). After this, they were required to go back over their response sheets and circle an item from each row that they had left blank (forced-choice recognition). As expected, this forced-choice procedure significantly elevated the estimates of r_1 [$G^2(1) = 18.77$], but had no significant effect on s or r_2 [$G^2(1) = 1.69$ and 0.68, respectively].

Variables that should affect recall but not recognition. Just as certain manipulations can be expected to affect recognition, so others should have their primary effect on cued recall (the parameter r_2). Because r_2 measures the retrieval of the B-terms given the A-terms as cues, one variable that should have a demonstrable effect on this process is the strength of association between A and B. This has been manipulated in a number of studies, including one by Wiseman and Tulving (1976, Experiment 4) in which related and unrelated pairs were compared (Cases 64 and 65 in the Appendix). Consistent with the predictions of the model, level of association had a significant effect on r_2 [$G^2(1) = 105.78$], but not on s or r_1 [$G^2(1)$ = 1.53 and 0.54, respectively]. We also examined other studies that manipulated level of association (e.g., Bryant, 1991; Neely & Payne, 1983; Shaps & Nilsson, 1980; Vining & Nelson, 1979), and they all revealed sizable effects on r_1. Some of these studies also exhibited significant effects of A-B association on storage capacity (the parameter s). As we indicated earlier, variables that strengthen the A-B association may improve the ability to store the paired associates as well as the ability to retrieve B given A as a cue.

On the whole, the preceding analyses indicate that the multinomial model does a credible job of interpreting and accounting for a wide range of experimental variables. In particular, when an independent variable in Table 2 was argued on psychological grounds either to affect or to not affect a specific parameter of the model, the expected result always occurred when the corresponding experiment was analyzed with the model. For some analyses, in addition to the expected result, a given variable affected other parameters as well. While the model does not give as clear-cut a story for the data in some of the latter cases, we think it has reasonable validity as a mathematical representation of retrieval-independence theory.

THE TULVING-WISEMAN FUNCTION REVISITED

With the multinomial model in place, progress can now be made in exploring the implications that retrieval independence has for the variables related by the Tulving-Wiseman function. To begin with, Equation 15b shows that $P(\text{Rn}|\text{Rc}) = \hat{r}_1$. We can expand this equation to produce

$$P(\text{Rn}|\text{Rc}) = \hat{s}\,\hat{r}_1 + (1 - \hat{s})\,\hat{r}_1. \tag{16}$$

But from Equations 15a and 15b, with $g = 0$, $P(\text{Rn}) = \hat{s}\,\hat{r}_1$. Substituting this into Equation 16 produces

$$P(\text{Rn}|\text{Rc}) = P(\text{Rn}) + (1 - \hat{s})\,\hat{r}_1. \tag{17}$$

Thus, the multinomial model yields an equation in the form of the Tulving-Wiseman function, except that in place of the quadratic term in Equation 1, there is now an expression in terms of the parameter estimators for storage and retrieval.

It is clear from Equation 17 that certain combinations of $\hat{s}$ and $\hat{r}_1$ will yield points tending to conform to the function, while other values of $\hat{s}$ and $\hat{r}_1$ will generate violations of the function. To examine this, we took the 302 data observations from Nilsson and Gardiner (1993) and used Equation 15 to compute the point estimates of s and r_1 (see Appendix). Two of these data points were excluded because the inequality constraint associated with Equation 15 was not met. Figure 6 presents the values of $\hat{s}$ plotted against the value of $\hat{r}_1$ for the remaining points. Concerning Figure 6, it should be noted that the values of $\hat{s}$ and $\hat{r}_1$ cover a fairly wide range, which is to be expected, considering that the data points come from a wide variety of experiments with different stimulus materials, experimental manipulations, and so forth. However, the large majority of the data points lie in the upper right quadrant of the parameter space, where $\hat{s}$ and $\hat{r}_1$ are both greater than .5.

As stated earlier, there are three degrees of freedom inherent in the data structure of the recognition-failure paradigm. The Tulving-Wiseman function only accounts for two of these degrees of freedom, because it deals with only two quantities, $P(\text{Rn})$ and $P(\text{Rn}|\text{Rc})$, in the two-dimensional

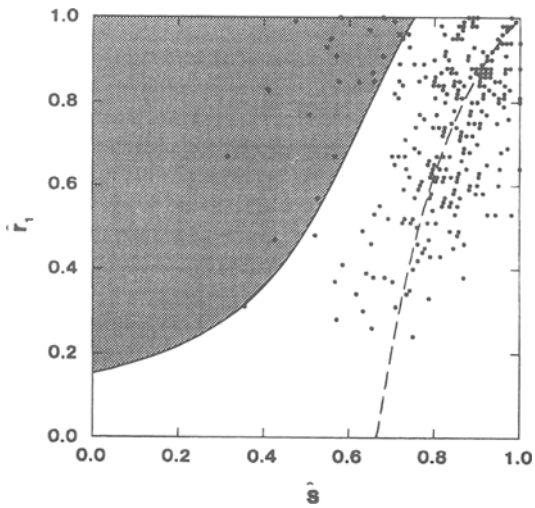

Figure 6. Relationship between $\hat{s}$ and $\hat{r}_1$, parameter estimators of the multinomial model that measure storage and recognition retrieval, respectively. The shaded region represents those combinations of $\hat{s}$ and $\hat{r}_1$ that violate the Tulving-Wiseman function. The dashed line represents the combinations of $\hat{s}$ and $\hat{r}_1$ that perfectly fit the function.

plot in Figure 1. However, notice from Equation 15 that the estimates of s and r_1 similarly depend only on the values of P(Rn) and P(Rn|Rc). Thus, Figure 6 basically represents a reparameterization of the Tulving-Wiseman data, translating the empirical data space in Figure 1 into a storage-retrieval parameter space. To make this clearer, Figure 6 also indicates which values of $\hat{s}$ and $\hat{r}_1$ yield points that perfectly conform to the Tulving-Wiseman function. This is represented by the dashed line, which can be thought of as the "shadow" of the function as it is projected into the new parameter space.

Violations of the Tulving-Wiseman Function

Next, we explore the question of which combinations of $\hat{s}$ and $\hat{r}_1$ lead to violations of the Tulving-Wiseman function. In particular, we will focus on encoding and retrieval exceptions to the function, as defined by Nilsson and Gardiner (1991, 1993). Earlier, we showed that exceptions to the function can be determined on a case-by-case basis using a simple χ^2 test of the data in the form of Table 1. However, Nilsson and Gardiner (1993) have proposed that exceptions can be defined by a critical ratio for each condition, which is the difference between the observed value of P(Rn|Rc) and the value predicted by the function, divided by the standard deviation of the differences for all 302 conditions. All critical ratios with absolute values larger than 1.96 represent points that are outside the 95% confidence interval, thus indicating a sizable deviation from the function. Nilsson and Gardiner's criterion has the advantage of treating all 302 data points as a total set; however, as we showed, their approach does not take into account the entire pattern of data observations in Table 1.

We can use the multinomial model to evaluate Nilsson and Gardiner's (1993) theoretical distinction between encoding and retrieval exceptions. To do this, we will adopt their criterion in the following analysis as we explore exceptions to the Tulving-Wiseman function. For the multinomial model, Equation 17 can be used to determine which values of $\hat{s}$ and $\hat{r}_1$ will lead to critical ratios exceeding 1.96. The boundary of these values is indicated by the solid curve in Figure 6, which represents a sort of borderline between two regions in the parameter space, one that leads to violations of the function and one that leads to nonviolations. In particular, exceptions to the Tulving-Wiseman function occur in the shaded region to the upper left of the curve.

One point to note about these two regions is that they differ in size; in fact, the violation region is approximately half the size of the nonviolation region. Moreover, the violation region covers only about 25% of the upper right quadrant, where most of the points actually occur. Most importantly, Figure 6 reveals that, according to the Nilsson-Gardiner criterion, it is impossible to violate the function if the estimate of the storage parameter s exceeds .75. As it turns out, a full 80% of all data points yield a value of $\hat{s}$ at .75 or higher. This means that, according to the model, the large majority of studies contributing to the Tulving-Wiseman function are incapable of violating the function, regardless of their level of retrieval.

Figure 6 also specifies the conditions under which violations of the function do occur. As a general rule, exceptions occur when weak storage (low value of $\hat{s}$) is coupled with strong retrieval (high value of $\hat{r}_1$). This finding helps to shed some light on the two categories of exceptions identified by Nilsson and Gardiner (1991, 1993). As mentioned earlier, encoding exceptions occur under conditions of weak encoding, which should make the probability of storing items fairly low. In addition, retrieval exceptions occur when redundant information between target and cue inflates the probability of retrieval. To examine this more thoroughly, Table 3 presents the estimates of the model's parameters from Equation 15 for all of the encoding and retrieval exceptions specifically identified as such by Nilsson and Gardiner (1993). Each entry is coded by its case number from the Appendix. Except perhaps for Case 301, all violations result in relatively low values of $\hat{s}$ and high values of $\hat{r}_1$.

The next question to consider from the point of view of the multinomial model concerns what differences, if any, exist between Nilsson and Gardiner's (1993) encoding and retrieval exceptions. One might expect that encoding exceptions, because they are engineered to produce weak encoding of stimuli, might exhibit particularly low estimates of the parameter s when compared with retrieval exceptions. In contrast, retrieval exceptions, because they result from enhanced retrieval cues for stimuli, might exhibit comparatively large estimates of the parameter r_1. However, an examination of Table 3 reveals that both types of exceptions result in comparable levels of both $\hat{s}$ and $\hat{r}_1$. The average values of $\hat{s}$ and $\hat{r}_1$ are .59 and .85, respectively, for the 16 encoding exceptions, and .53 and .80 for the 6 retrieval exceptions. This implies that encoding and retrieval exceptions not only lead to violations of the Tulving-Wiseman function, but they do so in the same way in terms of their effects on $\hat{s}$ and $\hat{r}_1$.

Table 3
Parameter Estimates for Encoding and Retrieval Exceptions to the Tulving-Wiseman Function

Case Number	s	r_1	r_2
	Encoding exceptions		
129	.58	1.00	.09
130	.68	1.00	.04
142	.68	.91	.35
160	.55	.93	.33
164	.62	.85	.21
269	.57	.91	.19
271	.57	.67	.05
280	.72	.96	.32
281	.66	.85	.43
282	.74	.93	.51
284	.56	.95	.11
285	.65	.97	.54
297	.53	.57	.53
298	.58	.85	.28
301	.42	.47	.52
302	.41	.83	.24
	Retrieval exceptions		
9	.52	.48	.77
168	.71	.99	.58
201	.47	.99	.55
249	.51	.77	.77
252	.31	.67	.70
254	.66	.87	.63

Note—s, probability of storage; r_1, probability of retrieval during recognition; r_2, probability of retrieval during recall.

This finding has important implications for how exceptions to the function are interpreted. For example, encoding exceptions may force the storage of stimuli to be weak, but they produce violations of the function only when the retrieval of those stimuli during recognition remains strong. It is important to remember that the parameter r_1 in the model is a conditional probability—it is the probability of successful retrieval during recognition given successful storage. As Batchelder and Riefer (1980) have pointed out, even under conditions of poor encoding, conditional retrieval can still be high if, say, the few stimuli that do get encoded have relatively high retrieval probabilities. This can be accomplished by manipulating independent variables that primarily affect storage without having a major effect on retrieval. As it turns out, encoding exceptions have historically been found by using A-B pairs that are unrelated (as opposed to weakly related), and by using shallow levels of processing (Begg, 1979; Bryant, 1991; Fisher, 1979; Gardiner & Tulving, 1980). We pointed out in the previous section that this manipulation does, in fact, have a strong effect on storage but not on retrieval during recognition. Thus, there is evidence that those studies producing encoding violations have done so by engineering conditions that weaken storage capacity (s) but not retrieval capacity (r_1).

Because encoding and retrieval exceptions both have the same underlying effect on the parameters s and r_1 in the model, it may be asked whether there is even a theoretical need to make a distinction between these two types of exceptions. The answer to this question can be found by examining the value of $\hat{r}_2$ for each type of exception. Although encoding and retrieval exceptions do not differ substantially on the values of $\hat{s}$ and $\hat{r}_1$, they show very systematic differences on the value of $\hat{r}_2$. Table 3 shows that retrieval exceptions exhibit much larger values of $\hat{r}_2$ than do encoding exceptions. In fact, the smallest value of $\hat{r}_2$ for the retrieval exceptions is higher than the largest value of $\hat{r}_2$ for the encoding exceptions. This makes sense theoretically, because retrieval exceptions are produced when the A-terms in the stimulus pairs contain redundant information for the B-terms. This should certainly elevate the value of $\hat{r}_2$, which measures the retrievability of B given A as a cue, in ways that are not accomplished in encoding exceptions. We can conclude, then, that even though encoding and retrieval exceptions both result from weak storage coupled with strong retrieval, they achieve this result in the model in categorically different ways. This supports the psychological validity of Nilsson and Gardiner's (1991, 1993) distinction, as well as the validity of the model.

Hintzman (1992,1993) has identified certain empirical patterns in P(Rn) and P(Rc) that he argues are likely to result in exceptions to the Tulving-Wiseman function. These include P(Rn) not close to one, P(Rc) relatively low, and P(Rn) much larger than P(Rc). It may be informative to explore the relationship between these empirical conditions and the pattern of parameter values for the multinomial model that leads to exceptions. First, remember that $p(\text{Rn}) = sr_1$ and $p(\text{Rc}) = sr_2$. Relatively low values of P(Rn) and P(Rc) can be achieved when the estimate of the parameter s itself is small. This is precisely the finding revealed in Figure 6, which showed that exceptions can occur only when the value of $\hat{s}$ is below .75. In addition, the model implies that P(Rn) will be much larger than P(Rc) only when $\hat{r}_1$ is much larger than $\hat{r}_2$. As we showed, this is true for encoding exceptions, but does not characterize retrieval exceptions. More specifically, the mean of $\hat{r}_1 - \hat{r}_2$ is .56 for the 16 encoding exceptions in Table 3, but is only .13 for the 6 retrieval exceptions. Correspondingly, the mean of P(Rn) − P(Rc) is .34 and .08 for the encoding and retrieval exceptions, respectively. Hintzman did not discuss the distinction between encoding and retrieval exceptions, but our analysis suggests that his set of empirical conditions applies primarily to encoding exceptions, and not to retrieval exceptions. We make this conclusion with caution, however, because it is based on only a small set of retrieval exceptions.

Nonviolations of the Tulving-Wiseman Function

The fact that the nonviolation region in Figure 6 is larger than the violation region gives some insight into why more points tend to conform to the function than to violate it, on the basis of Nilsson and Gardiner's (1993) criterion. But even if we discount the many additional exceptions to the function revealed earlier by our chi-square analysis, the question of why the majority of observations from so many different studies seem to lie so closely to the function specified by Tulving and Wiseman (1975) still remains to be explained. As stated earlier, Hintzman (1992, 1993) claims that no special psychological explanation is

needed for the function—that mathematical constraints can completely account for the apparent regularity of the data when $P(\text{Rn}|\text{Rc})$ is plotted against $P(\text{Rn})$. Most theorists would probably now agree that certain constraints are probably an important factor in the apparent fit of the Tulving-Wiseman function—a viewpoint that is also supported by some of our analyses. However, it is not generally accepted that mathematical or experimental constraints provide the entire explanation of the function. Tulving and Flexser (1992, 1993) in particular have disputed Hintzman's claims and have argued that the Tulving-Wiseman function is still of sufficient theoretical interest to warrant a psychological explanation.

Given the possibility that memory factors may contribute to the Tulving-Wiseman function, it can be asked what these factors might be. For example, we can explore what additional assumptions might need to be added to retrieval-independence theory in order to predict a precise relationship between $P(\text{Rn})$ and $P(\text{Rn}|\text{Rc})$ similar to the Tulving-Wiseman function. Progress on this question can be made if we make a straightforward, simplifying assumption about the relationship between storage and retrieval in the recognition-failure paradigm. According to Equation 17, the multinomial model will approximate the Tulving-Wiseman function if the value of $(1 - \hat{s})\hat{r}_1$ approximates the quadratic component in Equation 1—namely, $cP(\text{Rn})[1 - P(\text{Rn})]$. A necessary condition for this to happen is for $\hat{s}$ and $\hat{r}_1$ to be positively correlated across experimental conditions; that is, it will only happen if storage and retrieval capacities both tend to be either small or large for each experimental condition. Positive correlations often exist across experimental conditions between performance measures, a fact that is easily explained by variations from experiment to experiment in variables such as list length, study time, and A-B associative strength.

If we ignore the violation region in Figure 6, the parameter estimates $\hat{s}$ and $\hat{r}_1$ do exhibit a fairly strong positive correlation, with the largest values of $\hat{s}$ and $\hat{r}_1$ converging on the upper right corner of the parameter space, where $\hat{s}$ and $\hat{r}_1$ are both equal to one. The simplest way to represent this correlation is to postulate an underlying, straight-line relationship through the parameter space in Figure 6. Of course, many possible lines can be drawn through this space, and to specify a particular one requires only the identification of its slope and intercept. The most natural simplifying assumption is to leave the slope as a free parameter, and to assume that the line intercepts the point ($\hat{s} = 1$, $\hat{r}_1 = 1$). In other words, the line reaches its maximum level at the same place where the data points converge on their maximum values. Therefore the "best-fitting" line through $\hat{s}$ and $\hat{r}_1$ will have the equation

$$c(1 - r_1) = (1 - s), \tag{18}$$

where the slope of the line is denoted by c, with $c > 0$.

We can now explore at a theoretical level what specific predictions are made by the multinomial model, given this additional assumption. If we substitute the value of $(1 - s)$ from Equation 18 into Equation 17 (replacing proportions and estimators by probabilities and parameters), we obtain

$$p(\text{Rn}|\text{Rc}) = p(\text{Rn}) + c(1 - r_1)\, r_1. \tag{19}$$

However, according to the model, $r_1 = p(\text{Rn}|\text{Rc})$. Thus, at the level of data proportions, we obtain

$$P(\text{Rn}|\text{Rc}) = P(\text{Rn}) + c\, P(\text{Rn}|\text{Rc})\, [1 - P(\text{Rn}|\text{Rc})]. \tag{20}$$

Interestingly, Equation 20 yields an alternative function that is very similar to the Tulving-Wiseman function, except that $P(\text{Rn}|\text{Rc})$ is substituted for $P(\text{Rn})$ in the quadratic part of the equation. Given the strong linear component to the relation between $P(\text{Rn}|\text{Rc})$ and $P(\text{Rn})$ already established in Figure 1, one can expect that this new function will yield a comparable fit to the data. In fact, Equations 1 and 20 can be directly compared on their ability to account for the data, because each contains one free parameter c. To make this comparison, we computed the best-fitting value of c for each function, determined by a least-squares fit to the data.[4] We excluded from this analysis the 25 points identified by Nilsson and Gardiner (1993) as violations of the function, because Equations 1 and 20 are not really designed to account for them. The values of the parameter c were estimated at .508 for the Tulving-Wiseman function (very close to the traditional value of .500), and at .544 for Equation 20.

Figure 7 presents the same data points as Figure 1, along with the best-fitting curve from Equation 20. Visually, the fit of this new function appears almost indistinguishable from the Tulving-Wiseman function in Figure 1. A more precise comparison can be obtained by examining the proportion of variance explained by each function. The two functions are virtually identical on this measure as well ($r^2 = .933$ vs. .922 for Equations 1 and 20, respectively). Thus, Equation 20 provides a very credible description for the relationship between $P(\text{Rn})$ and $P(\text{Rn}|\text{Rc})$, and one that is derived from a psychologically explicit model, unlike the Tulving-Wiseman function.

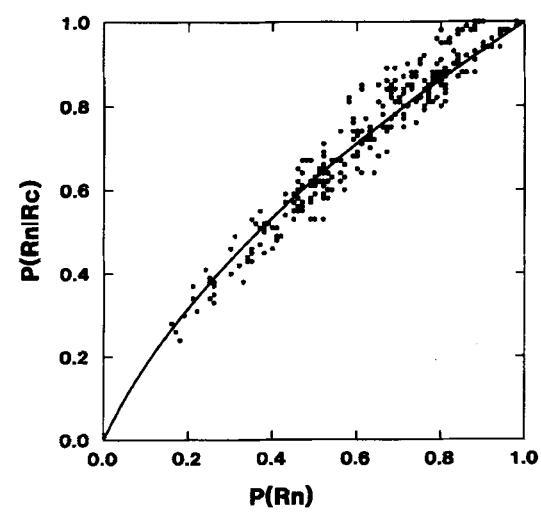

Figure 7. Relationship between probability of recognition, $P(\text{Rn})$, and probability of recognition given recall, $P(\text{Rn}|\text{Rc})$, plotted from Nilsson and Gardiner's (1993) database. The solid line through the points is the best-fitting curve from Equation 20, with a value of $c = .544$.

Earlier, we saw that despite the apparent good fit of the Tulving-Wiseman function, it could be rejected as a model for the recognition-failure paradigm by a χ^2 test that looked at the entire data structure in Table 1 on a case-by-case basis. A similar test can be constructed for the alternative relationship in Equation 20. First, note that Equation 20 can be written as

$$P(\text{Rn \& Rc}) = P(\text{Rn})\,P(\text{Rc}) + \frac{c\,P(\text{Rn \& Rc})\,[P(\text{Rc}) - P(\text{Rn \& Rc})]}{P(\text{Rc})}, \tag{21}$$

which implicitly relates $P(\text{Rn \& Rc})$ to the two marginals, $P(\text{Rn})$ and $P(\text{Rc})$. Equation 21 can be explicitly solved for $P(\text{Rn \& Rc})$, and the result is

$$P(\text{Rn \& Rc}) = \left[\frac{(1-c)P(\text{Rc})}{2c}\right]\left[-1+\sqrt{1+\frac{4c\,P(\text{Rn})}{(1-c)^2}}\right], \tag{22}$$

where $0 < c < 1$. Equation 22 can function exactly like Equation 4 in constructing a χ^2 test of the two-parameter restriction of the multinomial model implied by Equation 18. The marginals are estimated, as before, by Equations 9 and 10. Then, Equation 22 is inserted into Equation 11 in place of Equation 4, and the four expected frequencies, F_{ij}, for $i,j = 1,2$, are calculated. Finally, the χ^2 in Equation 12 can be computed for each condition.

As in the earlier χ^2 test, we analyzed all the conditions in Nilsson and Gardiner's (1993) database that were classified by the authors as nonexceptions and that had expected frequencies for this new analysis of five of more. There were 251 cases that met these restrictions, and the mean χ^2 for these cases was 5.90—well above the expected value of 1.00. Furthermore, 37% and 25% of these cases had a $\chi^2(1)$ value that was rejected at the $\alpha = .05$ and $\alpha = .01$ levels, respectively.[5] Thus, as with the Tulving-Wiseman function itself, the alternative function in Equation 20 can be rejected as a model for the recognition-failure paradigm. Both Equations 1 and 20 appear visually to account for the relationship between $P(\text{Rn}|\text{Rc})$ and $P(\text{Rn})$; however, neither equation does an adequate statistical job of accounting for the relationship across conditions with a single value of c, even when the so-called exceptions to the Tulving-Wiseman function are excluded.

IMPACT OF SIMPLIFYING ASSUMPTIONS

Our analysis of the recognition-failure paradigm was done using a very simple model of retrieval independence. The model is surely wrong in detail, but we regard it as a principled approximation to a more complete and valid theory of memory. To support this argument, it is necessary to consider in more detail several places where simplifying assumptions were made.

Recognition Guessing

The model presented in Equation 13 incorporates the possibility of a recognition hit through guessing alone. However, our main analysis assumed that $g = 0$ in order to analyze the Nilsson-Gardiner (1993) database. In this section, we consider the impact of our analysis if $g > 0$. First, Figure 4 exhibits the positive correlation between $P(\text{Rn})$ and $P(\text{Rc})$. If $g > 0$, then theoretically, $p(\text{Rn}) = s\,r_1\,(1 - g) + g$, which is a linear function in $s\,r_1$. Thus, on the reasonable assumption that g reflects experimental design factors that are approximately independent of storage and retrieval across different experiments, the predicted correlation between $P(\text{Rn})$ and $P(\text{Rc})$ should be relatively unaffected by the $g = 0$ assumption.

Second, Figure 5 shows that the correlation between $\hat{r}_1$ and $\hat{r}_2$ is essentially zero over experimental conditions, where $\hat{r}_1$ and $\hat{r}_2$ are given by Equation 15. If $g > 0$ is a known value for any experimental condition, it is easy to solve Equation 13 to yield new estimators of r_1 and r_2 given by

$$\hat{r}_1 = \frac{P(\text{Rn}|\text{Rc}) - g}{1 - g} \tag{23}$$

and

$$\hat{r}_2 = \frac{P(\text{Rc})\,[P(\text{Rn}|\text{Rc}) - g]}{P(\text{Rn}) - g}. \tag{24}$$

It is not straightforward to assess the exact impact of the variable g on the correlation between $\hat{r}_1$ and $\hat{r}_2$ in Equations 23 and 24. Nevertheless, it can be shown that the value of $\hat{r}_2$ is relatively insensitive to changes in g.[6] Thus, if the variation in g is approximately independent of storage and retrieval factors, as assumed earlier, it is possible to show that the correlation between $\hat{r}_1$ and $\hat{r}_2$ across conditions is largely unaffected by g.

Our analysis of the Tulving-Wiseman function started with Equation 16, and with $g > 0$, it becomes

$$p(\text{Rn}|\text{Rc}) = p(\text{Rn}) + r_1\,(1 - s)\,(1 - g). \tag{25}$$

It is easy to see from Equation 25 that introducing $g > 0$ does not alter the basic conclusions drawn from Equation 16, as long as g is approximately independent of r_1 and s. In the case of the alternative to the Tulving-Wiseman function in Equation 20, when Equation 17 is inserted into Equation 25 along with $r_1 = p(\text{Rn}|\text{Rc}) - (1 - r_1)\,g$, the result is

$$P(\text{Rn}|\text{Rc}) = P(\text{Rn}) + c\,P(\text{Rn}|\text{Rc})\,[1 - P(\text{Rn}|\text{Rc})] + cg\,(1 - g)\,(1 - \hat{r}_1)^2. \tag{26}$$

Equations 20 and 26 are practically indistinguishable, with a maximum discrepancy of $c\,(1 - \hat{r}_1)^2 / 4$, due to the fact that $g\,(1 - g) \leq .25$.

In conclusion, as long as design factors that influence the value of g across different experiments are approximately unrelated to storage and retrieval, we argue that our analyses of the multinomial model are robust under the introduction of $g > 0$ into the model.

Individual Differences Within Conditions

The database in Nilsson and Gardiner (1993) reports data from each condition in the aggregate form of Table 1. Thus, possible individual subject- or item-parameter differences

within a condition are not taken into account in our analysis across conditions. There are several reasons why we think our analysis is not severely affected by this assumption.

First, Riefer and Batchelder (1988, p. 328) considered the individual-difference problem for the multinomial model with $g = 0$ and, using a simulation analysis, concluded that the results of the model were robust under modest variation in individual parameters. Second, the considerable range of P(Rn) and P(Rc) in Nilsson and Gardiner's (1993) database suggests that the parameter variation between conditions is much greater than the variation within conditions. Finally, in the section on model validity, we documented many cases in which experimental variables had a psychologically expected effect on parameter estimates, even though the estimates were obtained from aggregate data. Nevertheless, the effect of individual subject or item differences within a condition of the recognition-failure paradigm remains an open issue deserving more empirical and modeling effort.

Discrete-State Representation

It is a basic property of multinomial processing-tree models that they use discrete-state representations of psychological processes. For example, in the current model, either successful storage completely occurs or it fails to occur at all. We regard such assumptions as approximations to more realistic but less tractable continuous-state representations. From the extensive work on models for signal detection (Macmillan & Creelman, 1991, chapter 4) and with Markov models of learning and memory (e.g., Wickens, 1982), as well as in other modeling areas, it is well known that discrete-state approximations can be quite accurate in fitting data in specific situations (Batchelder, 1993).

There has been some discussion recently about high-threshold assumptions in multinomial processing-tree models for source monitoring (Batchelder, Riefer, & Hu, 1994; Kinchla, 1994). It is important to note that even though the model we present here incorporates a discrete-state representation of recognition memory, this in itself does not make it a high-threshold model. Numerous models exist that contain discrete-state accounts of recognition without making high-threshold assumptions (see Macmillan & Creelman, 1991, chapter 4). Our multinomial model in Equation 13 only concerns behavior on old A-B pairs, and in order for the model to be complete, it would have to specify a processing tree to handle performance on distractor items. This could be done in a number of different ways, including using a high-threshold representation (e.g., setting the false-alarm probability equal to g). But other modifications of the model to handle distractor data could be made that would maintain its discrete-state assumptions without making it a high-threshold model. We did not attempt to account for distractors in the current model because Nilsson and Gardiner (1993) did not include information on distractors in their database. Moreover, many published studies on the Tulving-Wiseman function do not include a complete account of false-alarm data in the presentation of their results.

Substituting Estimates for True Parameters

In the scatterplots in Figures 5 and 6, we exhibited the relationship between parameters across conditions by plotting estimates rather than true values. In multinomial modeling, parameter estimators are often correlated within any condition, and these correlations can be estimated using the asymptotic variance-covariance matrix of the parameters (Riefer & Batchelder, 1988). We used the methods described in Hu and Batchelder (1994) to obtain estimates of the within-condition correlation between $\hat{r}_1$ and $\hat{r}_2$, and between $\hat{r}_1$ and $\hat{s}$, for all 302 conditions in Nilsson and Gardiner (1993). The results were consistent across conditions; namely, the correlations between $\hat{r}_1$ and $\hat{r}_2$ were all positive, while the correlations between $\hat{r}_1$ and $\hat{s}$ were all negative.

It is important to consider the impact of these consistent within-condition correlations between parameter estimators on the conclusions reached from Figures 5 and 6. For example, we concluded from Figure 5 that the estimators $\hat{r}_1$ and $\hat{r}_2$ were uncorrelated, but this result was obtained across conditions rather than within a condition. The positive correlation between $\hat{r}_1$ and $\hat{r}_2$ that occurs within a condition means that each observation in Figure 5 can be viewed as a single observation from its own bivariate sampling distribution with a positive correlation. Nevertheless, the correlation between $\hat{r}_1$ and $\hat{r}_2$ across conditions is reflected in the pattern of the bivariate means of these distributions, and not in their correlations. Thus we conclude that the impact of the correlation between estimators within a condition does not affect the conclusions drawn in the comparison between conditions. The same holds for the observed positive correlation between $\hat{r}_1$ and $\hat{s}$ across conditions in Figure 6.

CONCLUSION

A great deal of experimental, theoretical, and modeling effort has been devoted to the recognition-failure paradigm over the last two decades. The multinomial model presented here, and the resultant analysis of the model, has enabled us to examine this paradigm from a different perspective. We have taken the basic assumptions behind retrieval-independence theory and, through the model, have expressed them in simple mathematical form. The multinomial model postulated separate parameters for storage and retrieval, and it represented recognition and cued-recall retrieval processes as conditionally dependent on storage. The data from each of 302 conditions covering 44 separate studies were analyzed with the model, and it did a capable job of understanding the impact of various experimental manipulations on the parameter values. In addition, the model provided an understanding of the exceptions to the Tulving-Wiseman function, as noted by Nilsson and Gardiner (1993), in terms of a psychologically reasonable relationship among the parameters. Because of these results, we believe that this model is capable of taking aggregate data from a standard condition of the recognition-failure paradigm and generating parameter estimates that

separately measure storage and retrieval processes at a group aggregate level.

A large number of earlier analyses of the recognition-failure paradigm have focused on the Tulving-Wiseman function. We have shown that reformulations of the Tulving-Wiseman function that are probabilistically equivalent may produce larger errors and poorer goodness-of-fit measures than have been reported for the original function. We have further shown that the Tulving-Wiseman function can be viewed as a particular statistical model for the dependence among recognition and cued-recall performance statistics. When this model was tested with a χ^2 goodness-of-fit test, it did a poor job of explaining many of the conditions that are currently catalogued as nonexceptions to the function. Researchers should be wary of focusing attention on a particular feature of a data set, especially when analyzing that feature with a standard correlational approach. The lesson here is that models should be tested on the entire data array instead of just on an aspect of it.

Using the multinomial model as a representation of retrieval-independence theory, we were able to explore what additional assumptions are necessary to produce a precise equation, in a form similar to the Tulving-Wiseman function, to describe the relationship between $P(\mathrm{Rn}|\mathrm{Rc})$ and $P(\mathrm{Rn})$ in Figure 1. The necessary assumption turned out to be a positive correlation between storage and retrieval capacities across experimental conditions. We represented this assumption as a perfect correlation in the model's parameters across conditions, and despite this strong assumption, proceeded to derive a theoretical relationship between $P(\mathrm{Rn})$ and $P(\mathrm{Rn}|\mathrm{Rc})$ that accounts for the observed data essentially as well as the original Tulving-Wiseman function. However, just as with the Tulving-Wiseman function, the strong formulation could be rejected using a valid goodness-of-fit test.

Actually, the idea that storage and retrieval are correlated across experiments may reveal more about research trends than it does about human memory.[7] The points in Figure 1 represent separate experiments or experimental conditions, and as such, these points were certainly not determined by random selection. Instead, it is reasonable to assume that most experimenters choose stimulus materials, select subjects, and manipulate variables in such a manner that overall levels of both storage and retrieval are roughly comparable. Moreover, many experiments on the Tulving-Wiseman function are replications or slight variations of the experimental procedure developed originally by Tulving and Thomson (1973) to reveal the recognition-failure phenomenon. It is not surprising that points from similar experiments tend to cluster around a relatively narrow region in the data space. In fact, it wasn't until experimenters began to change basic elements of the standard paradigm (e.g., Muter's [1984] extension to semantic memory) that exceptions to the Tulving-Wiseman function began to occur with some regularity. These and other experimental factors could account for any correlation between storage and retrieval, which was the critical assumption needed by retrieval-independence theory to precisely account for the data in Figure 1. Thus it is reasonable to believe that the apparent good fit of the Tulving-Wiseman function across conditions may be as much a reflection of external design factors as it is of basic psychological and statistical processes.

The analysis of our model also provides a very logical reason why so many more data points tend to conform to the function than to violate it. The model reveals that only under a very special set of circumstances—namely, weak storage coupled with strong retrieval—do exceptions to the function, in Nilsson and Gardiner's (1993) sense, occur. Moreover, the model indicates that if the level of storage is high enough, which it is for the majority of studies contributing to the Tulving-Wiseman function, large deviations from the function are impossible, on the basis of Nilsson and Gardiner's criterion. This conclusion is somewhat similar to that of Hintzman (1991, 1992, 1993), in that there are certain constraints that tend to force points to conform roughly to the Tulving-Wiseman function. Our analysis can be seen as extending Hintzman's criticism in two important ways: First, we identify the possible psychological constraints behind the function dealing with the effective range of storage and retrieval processes across experiments; and second, we show how the natural constraints among the empirical statistics can be represented in a goodness-of-fit measure that examines all the data and avoids the pitfalls of earlier analyses of the Tulving-Wiseman function.

The analytic results that we obtained were possible only because we represented retrieval independence within a particularly simple model. Because of this, it was necessary to argue, as we did in the previous section, that the main conclusions of our analyses were unaffected by several simplifying assumptions. We also wish to stress that it is important in the analysis of simple, multinomial memory models to establish a formal link between the simplified model and more complete and realistic models. In the current case, it was observed earlier that the multinomial model (with $g = 0$) is equivalent to the simplest, nontrivial case of Flexser and Tulving's (1978) model that represents memory traces as feature vectors. In turn, the Flexser-Tulving model was one of the forerunners to the more recent composite trace models such as Metcalfe's (1992) CHARM and Murdock's (1993) TODAM2. Our analysis reveals consequences of this approach to modeling that are not evident in the analyses of the more complex models. In general, we think that detailed analysis of simple multinomial models of memory is a useful strategy to supplement the simulation analysis of more complex memory models.

REFERENCES

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

Arlemalm, T., & Nilsson, L.-G. (1992). Recognition failure of recallable words: Exception due to poor integration. *Scandinavian Journal of Psychology*, **33**, 266-276.

Bahrick, H. P. (1970). A two-phase model for prompted recall. *Psychological Review*, **77**, 215-222.

Batchelder, W. H. (1975). Individual differences and the all-or-none vs. incremental learning controversy. *Journal of Mathematical Psychology*, **12**, 53-73.

Batchelder, W. H. (1993, November). *Parameter variation in retrieval-independence theories of memory*. Paper presented at the meeting of the Psychonomic Society, Washington, DC.

BATCHELDER, W. H., & RIEFER, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review*, **87**, 375-397.

BATCHELDER, W. H., RIEFER, D. M., & HU, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, **101**, 172-176.

BEGG, I. (1979). Trace loss and the recognition failure of unrecalled words. *Memory & Cognition*, **7**, 113-123.

BISHOP, Y. M. M., FIENBERG, S. E., & HOLLAND, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.

BRYANT, D. J. (1991). Exceptions to recognition failure as a function of the encoded association between cue and target. *Memory & Cognition*, **19**, 210-219.

DONNELLY, R. E. (1988). Priming effects in successive episodic tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 256-265.

ERDFELDER, E., & BAYEN, U. J. (1991). Episodisches gedachtnis im Alter: Methodologische und empirische Argumente für einen Zugang über mathematische Modelle [Episodic memory in old age: Methodological and empirical arguments for a mathematical modeling approach]. In D. Frey (Ed.), *Bericht über den 37. Kongress der Deutschen Gesellschaft für Psychologie in Kiel 1990* [Proceedings of the 37th Conference of the German Society for Psychology in Kiel 1990] (Vol. 2, pp. 172-180). Göttingen, Germany: Hogrefe.

FISHER, R. P. (1979). Retrieval operations in cued recall and recognition. *Memory & Cognition*, **7**, 224-231.

FLEXSER, A. J., & TULVING, E. (1978). Retrieval independence in recognition and recall. *Psychological Review*, **85**, 153-171.

GARDINER, J. M. (1988). Recognition failures and free-recall failures: Implications for the relation between recall and recognition. *Memory & Cognition*, **16**, 446-451.

GARDINER, J. M., & TULVING, E. (1980). Exceptions to recognition failure of recallable words. *Journal of Verbal Learning & Verbal Behavior*, **19**, 194-209.

GREENO, J. G., JAMES, C. T., DAPOLITO, F., & POLSON, P. G. (1978). *Associate learning: A cognitive analysis*. Englewood Cliffs, NJ: Prentice-Hall.

HINTZMAN, D. L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review*, **87**, 398-410.

HINTZMAN, D. L. (1987). Recognition and recall in MINERVA 2: Analysis of the "recognition-failure" paradigm. In P. Morris (Ed.), *Modelling cognition* (pp. 215-229). London: Wiley.

HINTZMAN, D. L. (1991). Contingency analysis, hypotheses, and artifacts: Reply to Flexser and to Gardiner. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 341-345.

HINTZMAN, D. L. (1992). Mathematical constraints and the "Tulving-Wiseman Law." *Psychological Review*, **99**, 536-542.

HINTZMAN, D. L. (1993). On variability, Simpson's paradox, and the relation between recognition and recall: Reply to Tulving and Flexser. *Psychological Review*, **100**, 143-148.

HU, X., & BATCHELDER, W. H. (1994). Statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, **59**, 21-47.

HUMPHREYS, M. S., & BOWYER, P. A. (1980). Sequential testing effects and the relationship between recognition and recognition failure. *Memory & Cognition*, **8**, 271-277.

JONES, G. V. (1978). Recognition failure and dual mechanisms in recall. *Psychological Review*, **85**, 464-469.

JONES, G. V., & GARDINER, J. M. (1990). Recognition failure when recognition targets and recall cues are identical. *Bulletin of the Psychonomic Society*, **28**, 105-108.

KINCHLA, R. A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review*, **101**, 166-171.

KINTSCH, W. (1978). More on recognition failure of recallable words: Implications for generation-recognition models. *Psychological Review*, **85**, 270-281.

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

MARTIN, E. (1971). Verbal learning theory and independent retrieval phenomena. *Psychological Review*, **75**, 421-441.

METCALFE, J. (1992). Recognition failure and the composite memory trace in CHARM. *Psychological Review*, **98**, 529-553.

MURDOCK, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, **100**, 183-203.

MUTER, P. (1984). Recognition and recall of words with a single meaning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 198-202.

NEELY, J. H., & PAYNE, D. G. (1983). A direct comparison of recognition failure rates for recallable names in episodic and semantic memory tests. *Memory & Cognition*, **11**, 161-171.

NILSSON, L.-G., & GARDINER, J. M. (1991). Memory theory and the boundary conditions of the Tulving-Wiseman law. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory* (pp. 57-74). Hillsdale, NJ: Erlbaum.

NILSSON, L.-G., & GARDINER, J. M. (1993). Identifying exceptions in a database of recognition failure studies from 1973 to 1992. *Memory & Cognition*, **21**, 397-410.

NILSSON, L.-G., LAW, J., & TULVING, E. (1988). Recognition failure of recallable unique names: Evidence for an empirical law of memory and learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 266-277.

NILSSON, L.-G., & SHAPS, L. P. (1980). A functional view of memory. In F. Klix & J. Hoffmann (Eds.), *Cognition and memory: Interdisciplinary research of human memory activities* (pp. 40-46). Berlin: Deutscher Verlag der Wissenschaften.

RAAIJMAKERS, J. G., & SHIFFRIN, R. M. (1981). Search of associative memory. *Psychological Review*, **88**, 93-134.

RABINOWITZ, J. C. (1984). Aging and recognition failure. *Journal of Gerontology*, **39**, 65-71.

RATCLIFF, R., & MCKOON, G. (1989). Memory models, text-processing, and cue-dependent retrieval. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 73-92). Hillsdale, NJ: Erlbaum.

REDER, L. M., ANDERSON, J. R., & BJORK, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, **102**, 648-656.

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

RIEFER, D. M., & BATCHELDER, W. H. (1991a). Age differences in storage and retrieval: A multinomial modeling analysis. *Bulletin of the Psychonomic Society*, **29**, 415-418.

RIEFER, D. M., & BATCHELDER, W. H. (1991b). Statistical inference for multinomial processing tree models. In J.-P. Doignon & J.-C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 213-336). Berlin: Springer-Verlag.

RONNBERG, J., LYXELL, B., SAMUELSSON, S., ERNGRUND, K., & NILSSON, L.-G. (1991). Recognition failure of prose-embedded words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 288-301.

SANDBERG, K. (1990). Integration and cue overlap in recognition failure of recallable words. *Scandinavian Journal of Psychology*, **31**, 302-314.

SCHONFIELD, D., & ROBERTSON, B. A. (1966). Memory storage and aging. *Canadian Journal of Psychology*, **20**, 228-236.

SHAPS, L. P., & NILSSON, L.-G. (1980). Encoding and retrieval operations in relation to age. *Developmental Psychology*, **16**, 636-643.

TAJIKA, H. (1977). Features of recognition tasks in encoding specificity: Types of frequency associates in extralist cue words and types of recognition tasks. *Psychologia*, **20**, 151-158.

TAJIKA, H. (1978). Features of recognition tasks in encoding specificity: The function of context in recognition tasks. *Japanese Psychological Research*, **20**, 93-100.

TAJIKA, H. (1979). Memory processes in recall and recognition. *Psychologia*, **22**, 146-154.

TULVING, E., & FLEXSER, A. J. (1992). On the nature of the Tulving-Wiseman function. *Psychological Review*, **99**, 543-546.

TULVING, E., & FLEXSER, A. J. (1993). Recognition-failure constraints and the average maximum. *Psychological Review*, **100**, 149-153.

TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.

TULVING, E., & WATKINS, O. C. (1977). Recognition failure of words with a single meaning. *Memory & Cognition*, **5**, 513-522.

TULVING, E., & WISEMAN, S. (1975). Relation between recognition and

recognition failure of recallable words. *Bulletin of the Psychonomic Society*, **6**, 79-82.

VINING, S. K., & NELSON, T. O. (1979). Some constraints on the generality and interpretation of the recognition failure of recallable words. *American Journal of Psychology*, **92**, 257-276.

WATKINS, M. J., & TULVING, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, **104**, 5-29.

WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: W. H. Freeman.

WISEMAN, S., & TULVING, E. (1975). A test of confusion theory of encoding specificity. *Journal of Verbal Learning & Verbal Behavior*, **14**, 370-381.

WISEMAN, S., & TULVING, E. (1976). Encoding specificity: Relation between recall superiority and recognition failure. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 349-361.

NOTES

1. Tulving and Wiseman (1975) originally expressed their function in the form $P(\mathrm{Rn}|\mathrm{Rc}) = P(\mathrm{Rn}) + c\,[P(\mathrm{Rn}) - P(\mathrm{Rn})^2]$, which is algebraically equivalent to our Equation 1.

2. Models of human memory typically postulate cognitive processes that may occur differently for different items; however, they rarely incorporate any intrinsic subject or item differences directly into the framework of the model. Because some amount of parameter variation over subjects and items surely occurs, it is important in any paradigm to assess the robustness of a model to such variation. Riefer and Batchelder (1988, 1991b) discuss these issues and show how to conduct such analyses, and Batchelder (1975, 1993) discusses how subject–item parameter variation can be incorporated into models directly.

3. For these χ^2 tests, we used a value of $c = .508$, which was computed by minimizing the least-square difference between data points and the Tulving-Wiseman function in Figure 1. However, it is possible that this value of c, based on a least-square fit, is not the best value for the χ^2 tests. To explore the possibility that a different value of c might produce better fits, we used Equation 4 to compute the exact value of c for each of the 253 cases. We then took the mean of these values, which was .527, and redid the χ^2 tests with this new value of c. However, the performance of the Tulving-Wiseman function did not improve by using this new value.

4. In Equation 20, $P(\mathrm{Rn}|\mathrm{Rc})$ appears in both the left and right sides of the equation, unlike the Tulving-Wiseman function, which expresses $P(\mathrm{Rn}|\mathrm{Rc})$ explicitly as a function of $P(\mathrm{Rn})$. This discrepancy is easily fixed by rewriting Equation 20 as a quadratic in $P(\mathrm{Rn}|\mathrm{Rc})$ and solving to obtain

$$P(\mathrm{Rn}|\mathrm{Rc}) = (1/2c)\left[\sqrt{(1-c)^2 + 4c\,P(\mathrm{Rn})} - (1-c)\right].$$

This equation is algebraically equivalent to Equation 20, and expresses $P(\mathrm{Rn}|\mathrm{Rc})$ explicitly in terms of $P(\mathrm{Rn})$. We used this equation in all comparisons to the Tulving-Wiseman function.

5. In the reported analysis, we used $c = .544$, which was the value obtained in the fit reported in Figure 6. We found other values of c that worked better, but we were not able to get the mean χ^2 value below 5.00. Thus the two-parameter version of the multinomial model does not fit well for any value of c held constant over all conditions. However, on the whole, it fits just as well as the Tulving-Wiseman function.

6. Using calculus, we see that

$$\frac{\delta r_2}{\delta g} = \frac{\mathrm{cov}\,(\mathrm{Rn}, \mathrm{Rc})}{[P(\mathrm{Rn}) - g]^2},$$

and since the correlation between $P(\mathrm{Rn})$ and $P(\mathrm{Rc})$ is small, $\hat{r}_2$ is relatively unaffected by variations in g.

7. We thank Douglas Hintzman for suggesting and detailing this possibility.

APPENDIX
Point Estimates and 95% Confidence Intervals (CI) for the Parameters of the Model

Case	s CI	r_1 CI	r_2 CI
1	.68 ± .08	.38 ± .09	.92 ± .08
2	.71 ± .10	.31 ± .09	.87 ± .11

APPENDIX (Continued)

Case	s CI	r_1 CI	r_2 CI
3	.76 ± .07	.46 ± .07	.83 ± .07
4	.75 ± .06	.51 ± .07	.86 ± .07
5	.95 ± .04	.85 ± .05	.78 ± .06
6	.97 ± .03	.88 ± .04	.84 ± .05
7	.70 ± .13	.37 ± .11	.74 ± .14
8	.67 ± .11	.55 ± .12	.68 ± .13
9	.52 ± .10	.48 ± .13	.77 ± .14
10	.84 ± .16	.49 ± .13	.60 ± .14
11	.83 ± .12	.57 ± .12	.72 ± .13
12	.64 ± .21	.39 ± .17	.56 ± .20
13	.82 ± .06	.61 ± .06	.61 ± .06
14	.77 ± .08	.60 ± .07	.43 ± .06
15	.77 ± .05	.77 ± .05	.55 ± .05
16	.94 ± .09	.53 ± .07	.40 ± .06
17	.79 ± .10	.58 ± .08	.30 ± .05
18	.80 ± .06	.74 ± .06	.40 ± .05
19	.78 ± .08	.51 ± .07	.63 ± .08
20	.73 ± .06	.64 ± .07	.69 ± .07
21	.76 ± .06	.42 ± .06	.81 ± .07
22	.78 ± .07	.45 ± .07	.72 ± .08
23	.93 ± .07	.53 ± .06	.67 ± .07
24	.90 ± .07	.58 ± .07	.63 ± .07
25	.87 ± .08	.38 ± .06	.73 ± .08
26	.85 ± .07	.48 ± .06	.74 ± .07
27	.84 ± .06	.56 ± .06	.75 ± .06
28	.81 ± .05	.67 ± .06	.78 ± .06
29	.91 ± .05	.74 ± .07	.83 ± .06
30	.81 ± .06	.77 ± .07	.91 ± .05
31	.92 ± .05	.77 ± .06	.81 ± .06
32	.80 ± .06	.84 ± .06	.92 ± .04
33	.84 ± .09	.45 ± .07	.69 ± .09
34	.86 ± .08	.55 ± .08	.64 ± .08
35	.75 ± .10	.40 ± .08	.64 ± .10
36	.79 ± .06	.66 ± .07	.74 ± .07
37	.83 ± .07	.64 ± .07	.66 ± .07
38	.66 ± .08	.53 ± .08	.73 ± .09
39	.87 ± .08	.46 ± .06	.64 ± .07
40	.79 ± .08	.47 ± .07	.65 ± .07
41	.79 ± .06	.63 ± .06	.71 ± .06
42	.76 ± .06	.59 ± .06	.67 ± .07
43	.84 ± .07	.55 ± .07	.73 ± .08
44	.80 ± .07	.51 ± .07	.77 ± .08
45	.79 ± .08	.43 ± .07	.75 ± .09
46	.73 ± .07	.51 ± .08	.79 ± .08
47	.76 ± .06	.74 ± .06	.85 ± .06
48	.73 ± .06	.67 ± .07	.85 ± .06
49	.77 ± .06	.61 ± .07	.83 ± .06
50	.70 ± .06	.67 ± .07	.81 ± .07
51	.92 ± .03	.77 ± .04	.74 ± .04
52	.91 ± .07	.63 ± .08	.77 ± .08
53	.82 ± .08	.72 ± .09	.73 ± .08
54	.85 ± .07	.81 ± .08	.61 ± .09
55	.85 ± .08	.84 ± .08	.51 ± .09
56	.82 ± .07	.83 ± .08	.66 ± .08
57	.75 ± .08	.89 ± .08	.47 ± .09
58	.80 ± .08	.81 ± .09	.52 ± .09
59	.85 ± .07	.78 ± .08	.65 ± .09
60	.79 ± .07	.33 ± .05	.75 ± .07
61	.75 ± .14	.24 ± .06	.45 ± .10
62	.75 ± .05	.69 ± .06	.50 ± .05
63	.80 ± .04	.85 ± .04	.48 ± .05
64	.86 ± .04	.63 ± .04	.75 ± .04

APPENDIX (Continued)

Case	s CI	r_1 CI	r_2 CI
65	.82 ± .06	.62 ± .06	.41 ± .05
66	.86 ± .08	.62 ± .08	.59 ± .08
67	.84 ± .06	.62 ± .06	.57 ± .06
68	.88 ± .08	.64 ± .08	.61 ± .07
69	.82 ± .06	.61 ± .06	.56 ± .06
70	.91 ± .03	.78 ± .03	.70 ± .04
71	.89 ± .04	.80 ± .05	.38 ± .04
72	.88 ± .07	.72 ± .07	.57 ± .07
73	.91 ± .04	.84 ± .04	.53 ± .05
74	.91 ± .06	.81 ± .06	.58 ± .07
75	.88 ± .05	.77 ± .05	.52 ± .05
76	.95 ± .06	.64 ± .07	.69 ± .07
77	.85 ± .06	.82 ± .06	.55 ± .07
78	.83 ± .06	.92 ± .06	.38 ± .06
79	.82 ± .05	.96 ± .03	.54 ± .06
80	.89 ± .07	.98 ± .04	.65 ± .10
81	.91 ± .07	.85 ± .08	.87 ± .08
82	.79 ± .09	.62 ± .10	.65 ± .10
83	.80 ± .08	.61 ± .09	.73 ± .09
84	.73 ± .07	.71 ± .08	.72 ± .08
85	.88 ± .06	.64 ± .07	.83 ± .07
86	.84 ± .07	.75 ± .08	.68 ± .08
87	.80 ± .08	.64 ± .09	.70 ± .09
88	.88 ± .05	.68 ± .06	.76 ± .05
89	.89 ± .04	.75 ± .05	.75 ± .05
90	.85 ± .02	.85 ± .02	.77 ± .02
91	.91 ± .04	.88 ± .04	.83 ± .05
92	.65 ± .15	.26 ± .08	.54 ± .14
93	.59 ± .10	.41 ± .09	.63 ± .11
94	.98 ± .02	1.00 ± .00	.45 ± .06
95	.98 ± .02	.98 ± .02	.59 ± .06
96	.93 ± .05	.83 ± .06	.56 ± .06
97	1.00 ± .05	.80 ± .06	.56 ± .06
98	.99 ± .01	.99 ± .01	.96 ± .02
99	.79 ± .04	.65 ± .05	.77 ± .05
100	.90 ± .03	.87 ± .03	.66 ± .04
101	.65 ± .05	.46 ± .05	.64 ± .06
102	.95 ± .03	.93 ± .03	.96 ± .03
103	1.05 ± .16	.37 ± .08	.51 ± .10
104	.95 ± .03	.95 ± .03	.99 ± .01
105	.74 ± .08	.35 ± .07	.88 ± .08
106	.96 ± .02	.82 ± .05	.99 ± .02
107	.91 ± .07	.75 ± .07	.59 ± .08
108	.92 ± .03	.90 ± .04	1.00 ± .01
109	.90 ± .09	.62 ± .08	.59 ± .08
110	.79 ± .11	.43 ± .08	.56 ± .09
111	.89 ± .07	.72 ± .07	.52 ± .07
112	.63 ± .09	.30 ± .06	.58 ± .09
113	.81 ± .09	.64 ± .09	.52 ± .08
114	.57 ± .12	.37 ± .10	.48 ± .12
115	.75 ± .08	.69 ± .08	.48 ± .07
116	.87 ± .05	.68 ± .06	.68 ± .06
117	.72 ± .05	.67 ± .06	.81 ± .06
118	.81 ± .05	.74 ± .06	.75 ± .06
119	.96 ± .03	.79 ± .04	.60 ± .04
120	.73 ± .05	.84 ± .05	.32 ± .04
121	.91 ± .08	.87 ± .09	.68 ± .11
122	.97 ± .07	.94 ± .08	.38 ± .10
123	.89 ± .12	.78 ± .12	.43 ± .11
124	.94 ± .08	.92 ± .09	.34 ± .09
125	.95 ± .09	.81 ± .10	.52 ± .10
126	.97 ± .06	.97 ± .06	.29 ± .08
127	.73 ± .11	.95 ± .09	.32 ± .12

APPENDIX (Continued)

Case	s CI	r_1 CI	r_2 CI
128	.89 ± .06	1.00 ± .00	.05 ± .04
129	.58 ± .09	1.00 ± .00	.09 ± .07
130	.68 ± .08	1.00 ± .00	.04 ± .04
131	.72 ± .17	.81 ± .19	.18 ± .09
132	.83 ± .21	.88 ± .22	.08 ± .06
133	2.00 ± 1.94	.28 ± .27	.01 ± .01
134	.73 ± .18	.59 ± .15	.11 ± .04
135	.80 ± .14	.86 ± .15	.05 ± .02
136	.85 ± .14	.58 ± .10	.20 ± .05
137	1.00 ± .15	.88 ± .14	.04 ± .02
138	.96 ± .05	.81 ± .05	.28 ± .03
139	1.00 ± .08	.64 ± .06	.42 ± .05
140	.92 ± .02	.77 ± .03	.74 ± .03
141	.86 ± .03	.73 ± .04	.74 ± .04
142	.68 ± .09	.91 ± .09	.35 ± .10
143	.72 ± .09	.89 ± .09	.43 ± .10
144	.74 ± .08	.62 ± .10	.85 ± .09
145	.85 ± .08	.74 ± .09	.76 ± .09
146	.91 ± .06	.86 ± .07	.70 ± .08
147	.96 ± .04	.96 ± .04	.68 ± .08
148	.92 ± .05	.65 ± .08	.93 ± .05
149	.92 ± .05	.88 ± .06	.91 ± .05
150	.80 ± .05	.84 ± .05	.63 ± .06
151	.86 ± .07	.95 ± .05	.74 ± .09
152	.98 ± .10	.53 ± .08	.52 ± .08
153	.97 ± .07	.69 ± .07	.53 ± .07
154	.88 ± .05	.86 ± .05	.67 ± .06
155	.97 ± .03	.91 ± .04	.69 ± .06
156	.96 ± .02	.84 ± .03	.94 ± .02
157	.93 ± .03	.90 ± .03	.75 ± .04
158	.92 ± .03	.91 ± .03	.56 ± .05
159	.71 ± .10	.82 ± .11	.20 ± .06
160	.55 ± .06	.93 ± .07	.33 ± .07
161	.81 ± .07	.84 ± .08	.35 ± .06
162	.78 ± .06	.89 ± .06	.48 ± .07
163	.71 ± .11	.82 ± .13	.16 ± .05
164	.62 ± .09	.85 ± .11	.21 ± .06
165	.82 ± .06	.84 ± .06	.58 ± .06
166	.85 ± .05	.87 ± .05	.61 ± .06
167	.89 ± .03	.99 ± .01	.63 ± .05
168	.71 ± .04	.99 ± .01	.58 ± .05
169	.97 ± .02	.97 ± .02	.84 ± .03
170	.85 ± .03	.95 ± .02	.95 ± .02
171	.89 ± .04	.93 ± .04	.77 ± .06
172	.87 ± .05	.98 ± .03	.50 ± .07
173	.96 ± .03	.94 ± .03	.92 ± .04
174	.97 ± .03	.87 ± .05	.82 ± .05
175	.92 ± .04	.88 ± .05	.72 ± .06
176	.82 ± .06	.91 ± .05	.52 ± .07
177	.88 ± .04	.75 ± .06	.96 ± .03
178	.98 ± .03	.83 ± .05	.85 ± .05
179	.90 ± .03	.98 ± .01	.76 ± .04
180	.96 ± .02	.98 ± .01	.64 ± .04
181	.87 ± .03	.99 ± .01	.77 ± .04
182	.88 ± .03	1.00 ± .00	.71 ± .04
183	.90 ± .03	1.00 ± .00	.62 ± .05
184	.86 ± .06	.98 ± .02	.95 ± .04
185	.90 ± .05	.98 ± .02	.95 ± .04
186	.85 ± .03	.81 ± .04	.79 ± .04
187	.73 ± .05	.59 ± .06	.71 ± .06
188	.91 ± .03	.85 ± .04	.66 ± .04
189	.74 ± .05	.96 ± .05	.16 ± .04
190	.93 ± .03	.87 ± .03	.72 ± .04

APPENDIX (Continued)

Case	s CI	r_1 CI	r_2 CI
191	.90 ± .04	.88 ± .04	.65 ± .05
192	.83 ± .04	.90 ± .03	.62 ± .05
193	.94 ± .03	.85 ± .04	.70 ± .04
194	.88 ± .03	.90 ± .03	.55 ± .04
195	.89 ± .05	.88 ± .05	.45 ± .05
196	.80 ± .04	.90 ± .04	.40 ± .05
197	.92 ± .05	.82 ± .05	.43 ± .05
198	.82 ± .04	.87 ± .04	.70 ± .05
199	.96 ± .03	.81 ± .04	.69 ± .04
200	.71 ± .05	.65 ± .06	.62 ± .06
201	.48 ± .04	.99 ± .02	.55 ± .06
202	.86 ± .04	.71 ± .05	.72 ± .05
203	.86 ± .08	.57 ± .07	.44 ± .06
204	.94 ± .05	.82 ± .07	.72 ± .07
205	.92 ± .06	.87 ± .07	.60 ± .08
206	.95 ± .07	.75 ± .08	.67 ± .08
207	.92 ± .07	.86 ± .08	.48 ± .08
208	.86 ± .06	.85 ± .07	.69 ± .08
209	.92 ± .07	.84 ± .08	.48 ± .08
210	.93 ± .05	.86 ± .06	.90 ± .05
211	.93 ± .06	.88 ± .06	.71 ± .08
212	.86 ± .06	.81 ± .07	.86 ± .07
213	.85 ± .06	.95 ± .05	.69 ± .08
214	.89 ± .07	.80 ± .08	.73 ± .09
215	.84 ± .07	.87 ± .08	.62 ± .09
216	.86 ± .07	.83 ± .07	.81 ± .08
217	.85 ± .06	.96 ± .04	.60 ± .09
218	.85 ± .08	.82 ± .09	.57 ± .10
219	.82 ± .08	.91 ± .08	.43 ± .09
220	.83 ± .09	.80 ± .10	.53 ± .10
221	.83 ± .10	.88 ± .11	.30 ± .09
222	.97 ± .02	.98 ± .02	.93 ± .03
223	.75 ± .09	.57 ± .09	.65 ± .10
224	.63 ± .11	.49 ± .11	.62 ± .12
225	.98 ± .01	.94 ± .02	.86 ± .03
226	.76 ± .07	.50 ± .06	.55 ± .07
227	.77 ± .10	.44 ± .07	.41 ± .07
228	.87 ± .04	.84 ± .04	.76 ± .05
229	.86 ± .07	.72 ± .08	.63 ± .08
230	.77 ± .06	.68 ± .07	.75 ± .07
231	.84 ± .06	.67 ± .07	.74 ± .07
232	.82 ± .10	.56 ± .09	.66 ± .10
233	.83 ± .05	.82 ± .05	.78 ± .06
234	.90 ± .05	.80 ± .06	.71 ± .06
235	.87 ± .08	.69 ± .09	.64 ± .09
236	.66 ± .18	.38 ± .12	.24 ± .08
237	.69 ± .07	.52 ± .07	.52 ± .07
238	.85 ± .05	.53 ± .05	.65 ± .05
239	.36 ± .15	.31 ± .14	.14 ± .07
240	.57 ± .21	.28 ± .11	.21 ± .09
241	.62 ± .14	.34 ± .10	.42 ± .11
242	.82 ± .07	.73 ± .07	.40 ± .05
243	.80 ± .05	.81 ± .05	.46 ± .05
244	.82 ± .08	.74 ± .07	.30 ± .05
245	1.00 ± .01	.94 ± .03	.77 ± .05
246	.87 ± .04	.97 ± .03	.52 ± .06
247	.90 ± .08	.60 ± .09	.70 ± .09
248	.85 ± .09	.59 ± .09	.70 ± .09
249	.51 ± .08	.77 ± .09	.77 ± .09
250	.81 ± .12	.62 ± .13	.68 ± .13

APPENDIX (Continued)

Case	s CI	r_1 CI	r_2 CI
251	.73 ± .12	.52 ± .13	.82 ± .12
252	.31 ± .11	.67 ± .20	.70 ± .20
253	.75 ± .14	.52 ± .14	.69 ± .15
254	.66 ± .11	.87 ± .10	.63 ± .13
255	.98 ± .04	.88 ± .05	.73 ± .07
256	.94 ± .06	.82 ± .07	.67 ± .08
257	.85 ± .06	.85 ± .07	.66 ± .08
258	.91 ± .07	.89 ± .08	.70 ± .10
259	.95 ± .08	.80 ± .10	.66 ± .11
260	.88 ± .14	.60 ± .13	.61 ± .13
261	.91 ± .09	.84 ± .10	.62 ± .11
262	.86 ± .07	.95 ± .06	.66 ± .10
263	.94 ± .04	.82 ± .05	.80 ± .05
264	.88 ± .05	.59 ± .07	.85 ± .06
265	.79 ± .10	.57 ± .10	.60 ± .10
266	.96 ± .04	.82 ± .06	.86 ± .05
267	.85 ± .09	.61 ± .09	.67 ± .09
268	.96 ± .05	.68 ± .07	.81 ± .07
269	.57 ± .10	.91 ± .12	.19 ± .08
270	.82 ± .06	.98 ± .03	.55 ± .08
271	.57 ± .32	.67 ± .38	.05 ± .05
272	.84 ± .08	.88 ± .08	.38 ± .08
273	.91 ± .09	.58 ± .09	.66 ± .09
274	1.00 ± .02	.81 ± .06	.88 ± .05
275	.90 ± .07	.70 ± .08	.70 ± .08
276	.97 ± .03	.76 ± .06	.90 ± .05
277	.72 ± .10	.85 ± .11	.28 ± .08
278	.84 ± .06	.95 ± .04	.53 ± .08
279	.82 ± .46	.67 ± .38	.04 ± .03
280	.72 ± .07	.96 ± .06	.32 ± .08
281	.66 ± .09	.85 ± .10	.43 ± .10
282	.74 ± .07	.93 ± .06	.51 ± .09
283	.91 ± .05	.92 ± .05	.60 ± .08
284	.56 ± .07	.95 ± .09	.11 ± .04
285	.65 ± .05	.97 ± .03	.54 ± .06
286	.77 ± .05	.85 ± .05	.79 ± .05
287	.86 ± .06	.66 ± .07	.79 ± .07
288	.82 ± .07	.55 ± .08	.81 ± .07
289	.80 ± .07	.54 ± .08	.80 ± .07
290	.95 ± .05	.60 ± .07	.81 ± .06
291	.87 ± .05	.46 ± .05	.81 ± .06
292	.74 ± .08	.34 ± .06	.69 ± .08
293	.86 ± .04	.49 ± .05	.90 ± .04
294	.81 ± .07	.43 ± .06	.74 ± .07
295	.78 ± .04	.76 ± .05	.77 ± .05
296	.88 ± .04	.91 ± .04	.52 ± .05
297	.53 ± .07	.57 ± .08	.53 ± .08
298	.58 ± .06	.85 ± .08	.28 ± .06
299	.80 ± .06	.65 ± .06	.60 ± .06
300	.78 ± .05	.86 ± .05	.49 ± .05
301	.43 ± .07	.47 ± .09	.52 ± .10
302	.41 ± .06	.83 ± .11	.24 ± .07

Note—Case numbers for each data set correspond to those used by Nilsson and Gardiner (1993). CI, approximate 95% confidence interval computed from Equation 34 in Batchelder and Riefer (1988); s, probability of storage; r_1, probability of retrieval during recognition; r_2, probability of retrieval during recall.

(Manuscript received February 22, 1994;
accepted for publication August 9, 1994.)