Psychological Review
1988, Vol. 95, No. 3, 318-339

0033-295X/88/$00.75

Multinomial Modeling and the Measurement of Cognitive Processes

David M. Riefer
California State University,
San Bernardino

William H. Batchelder
School of Social Science
University of California, Irvine

This article presents a detailed discussion and application of a methodology, called multinomial modeling, that can be used to measure and study cognitive processes. Multinomial modeling is a statistically based technique that involves estimating hypothetical parameters that represent the probabilities of unobservable cognitive events. Models in this class provide a statistical methodology that is compatible with computational theories of cognition. Multinomial models are relatively uncomplicated, do not require advanced mathematical techniques, and have certain advantages over other, more traditional methods for studying cognitive processes. The statistical methodology behind multinomial modeling is briefly discussed, including procedures for data collection, model development, parameter estimation, and hypothesis testing. Three substantive examples of multinomial modeling are presented. Each example, taken from a different area within the field of human memory, involves the development of a multinomial model and its application to a specific experiment. It is shown how multinomial models facilitate the interpretation of the experiments. The conclusion discusses the general advantages of multinomial models and their potential application as research tools for the study of cognitive processes.

Theoretical explanations in cognitive psychology often assume the existence of hypothetical mental processes. For example, in the areas of learning and memory, such processes include stimulus discrimination, spreading activation, storage and retrieval, and various interference factors. The behavior that results from these processes, such as reaction time or number of words recalled, is easy to measure and record; however, the underlying mental processes themselves are not directly observable.

This article describes and applies a type of modeling called multinomial modeling that can be used to measure and study cognitive processes. Multinomial modeling is a statistically based technique that involves estimating hypothetical parameters that represent the probabilities of unobservable events. This approach has been used extensively in scientific areas outside of psychology, for example, in statistical genetics (Elandt-Johnson, 1971). In addition, examples of this type of modeling have appeared on a few occasions within the psychological literature (e.g., Batchelder & Riefer, 1980, 1986; Chechile & Meyer, 1976; Humphreys & Bain, 1983; Ross & Bower, 1981).

We feel that multinomial modeling has the potential for many more productive applications in cognitive psychology. Thus, one of our goals in this article is to critically evaluate this methodology in order to make it more readily understood by cognitive psychologists who might benefit from its use. As we show in the next section, multinomial modeling provides a statistical methodology that is compatible with computational theories of cognition. To make this case, we describe the key concepts of multinomial modeling and then show how this technique helps one to understand theoretical issues in three substantively different memory paradigms.

In addition to illustrating the techniques of multinomial modeling, another goal of this article is to outline the advantages that multinomial models have over other, more commonly used methods for studying cognitive processes. Foremost among these advantages is their simplicity. Multinomial models are relatively uncomplicated and do not require advanced mathematical techniques. Basically, the only mathematical prerequisite for working with these models is some familiarity with elementary calculus and mathematical statistics. To apply multinomial modeling to the study of cognitive processes, one needs only to make a set of simple theoretical assumptions concerning the nature of these processes. Often this will involve no more than identifying which cognitive processes are relevant to the situation and specifying how they influence the data. Once these assumptions have been made, the model itself can be expressed as a series of simple equations that relate the cognitive processes to the observable data. Standard statistical theory can be applied to these equations to yield estimates for the probability of each cognitive event as well as to develop hypothesis tests of how these processes vary over experimental conditions.

In addition to their simplicity, multinomial models also overcome certain limitations that are inherent in other, more commonly used methods in cognitive psychology. For example, one standard approach for studying mental processes is the application of general-purpose, off-the-shelf statistical models, such as the general linear model of analysis of variance (ANOVA) or log-linear models. These models may yield insights into how cognitive processes operate by studying differences in empirical data between experimental conditions. However, because this methodology is intended to have general applications, models such as ANOVA are not tailored specifically to the particular theoretical issues being studied. In fact, ANOVA usually does not permit

Correspondence concerning this article should be addressed to William H. Batchelder, School of Social Science, University of California, Irvine, California 92717.

one to measure directly underlying mental variables but instead provides a method for assessing whether cognitive processes act in conjunction to create differences between conditions. In this approach, one's cognitive theory motivates the selection of experimental conditions, but the theory itself is not reflected in the statistical tools used to analyze the experimental data. Multinomial modeling is one simple way theoretical ideas can be represented in data analysis.

Another common approach in cognitive psychology is the development of strong, theoretical models that can explain data in a variety of experimental paradigms. A number of such models have been imported from various scientific fields, such as computer science, electrical engineering, and neurobiology, and examples include the buffer model (Atkinson & Shiffrin, 1968), EPAM (Feigenbaum, 1970), holographic models (Pribram, Newer, & Baron, 1974), ACT (Anderson, 1976), and connectionist models (Rumelhart & McClelland, 1986). Because they are theoretically motivated, these types of models have certain advantages over general-purpose statistical models. Specifically, when data are analyzed using strong theoretical models, the results can be directly interpreted in terms of the underlying constructs being studied. However, one of the goals of this type of modeling is to give close fits to data for as wide a range of cognitive phenomena as possible. Because of this, one drawback to using these models in research is that they are often complicated and require highly technical mathematical or computer analysis. Estimating the parameter values and testing statistical hypotheses can be difficult or even infeasible and, if possible, may require Monte Carlo or other ad hoc techniques.

In our opinion, multinomial models bridge a gap between general-purpose statistical models, on the one hand, and strong theoretical models, on the other. Unlike, say, ANOVAS, multinomial models are theoretically motivated, making certain assumptions about the nature of the underlying cognitive processes. They can therefore be used to measure more directly the separate influence of cognitive events on overt behavior. Of course, strong theoretical models also share this feature. But in contrast to these models, multinomial models are analytically simple enough so that data can be evaluated without relying on complicated ad hoc mathematical analysis. Classical parameter estimation and hypothesis-testing techniques are relatively simple for multinomial models, and it is often the case that closed-form solutions can be found for the parameter estimators.

One class of theoretical models that multinomial models bear a family resemblance to are finite-state Markov models. Markov models have been developed for many paradigms in cognitive psychology (see Greeno & Bjork, 1973, for a review), and the statistical techniques for analyzing them are well represented in the psychological literature (e.g., Atkinson, Bower, & Crothers, 1965; Levine & Burke, 1972; Wickens, 1982). Both multinomial models and Markov models describe discrete data events as arising probabilistically from underlying cognitive processes; however, Markov models go further and attempt to describe trial-to-trial changes in the underlying processing events.

In practice, Markov models usually classify data on any trial into a very small number of categories, such as error or success. Then statistics of the trial-to-trial data protocols provide the information for parameter estimation. In contrast, multinomial models require that data be classified into a large number of categories, typically exceeding the number of parameters of the model. Thus, a multinomial model attempts to achieve a static account of a more detailed representation of data, whereas a Markov model attempts both a static and dynamic account of a less detailed data representation. A Markov model may fail to account for data either because its cognitive assumptions are wrong or because trial-to-trial changes among cognitive states do not occur as postulated. Thus, in paradigms that permit both types of models, a multinomial model is less theoretically committing and may be superior for the purpose of measuring process probabilities.

In the next section, the statistical methodology for multinomial modeling is presented in some detail, along with references for further study. We briefly outline the basic concepts necessary for using multinomial models, including data collection, model development, and parameter estimation. After this, three substantive examples of multinomial modeling developed from the areas of learning and memory are analyzed. The applications are in the areas of proactive inhibition, storage and retrieval processes in free recall, and discrimination learning, and each application reveals facets of how multinomial modeling can shed light on important theoretical issues in cognitive psychology. Finally, in the conclusion of this article we further discuss the advantages and potential application of multinomial modeling as a method for studying cognitive processes.

Multinomial Models

Theoretical Orientation

We start with an assumption that is implied by many explanations and theories in cognitive psychology, basically, that cognitive processing involves a finite set of discrete processing states. In cognitive paradigms involving finitely many behavioral categories, it is often assumed that each discrete behavioral act arises from one and only one of these underlying states. This assumption has been formulated more rigorously by cognitive scientists who view cognitive processing as computations by a formal automaton with finitely many states such as a Turing machine (Minsky, 1967). The computational assumption is quite unrestrictive, and many theoreticians accept it as forming the basis of contemporary cognitive theories (e.g., Pylyshyn, 1986).

We now show that multinomial models follow naturally from this assumption, and thus multinomial modeling can be viewed as providing a statistical methodology compatible with the computational assumption. To see this more clearly, we can recast the assumption in probabilistic terms. For notational purposes let the mutually exclusive behavioral categories be denoted by $C_1, C_2, \ldots, C_J$, and let the cognitive states be denoted by $T_1, T_2, \ldots, T_I$ (where J and I are positive integers). We can also define p_j to be the probability of observing a behavior in category C_j. Then the computational assumption implies the elementary, conditional probability formulae

$$p_j = \sum_{i=1}^{I} P(C_j|T_i)P(T_i) = \sum_{i=1}^{I} a_{ij}b_i, \quad (1)$$

where a_{ij} is the conditional probability of behavior category C_j given cognitive state T_i, and b_i is the unconditional probability of being in state T_i (which of course depends on the history of the automaton).

In actual research it is relatively easy to obtain estimates for the values of p_j from experimental observations, as we demonstrate in upcoming sections. What would be desirable is a rational technique for inferring the values of the cognitive parameters a_{ij} and b_i from these estimates of p_j. If we had such a technique, we could measure these cognitive quantities and study how they vary over different subjects or experimental factors. However, one barrier to developing such a technique is that Equation 1 contains only $J - 1$ independent known quantities (because the p_j sum to one) and a larger number of unknown cognitive parameters ($IJ - 1$ to be exact). Without some restrictions to reduce the number of unknowns, further progress is not possible.

The restriction that we impose on Equation 1 is that each of the a_{ij} and b_i is determined from a set of statistical parameters $\theta_1, \theta_2, \ldots, \theta_S$, where S (the number of parameters) is no larger than $J - 1$. These parameters are assumed to be the probabilities of various cognitive events. Under this restriction, the $J - 1$ independent values of p_j are each a function of the cognitive parameters through Equation 1. With the number of unknown parameters now smaller than the number of known data probabilities ($S < J$), unique estimates for these parameters may be obtained.

Of course, it is possible to make less restrictive assumptions about the number of parameters than the ones we made above (see, e.g., Bamber & van Santen, 1985; Chechile & Meyer, 1976). However, our assumptions allow for the general formulation and classical statistical analysis of multinomial models. We illustrate this in the next few subsections.

Data Representation

In any particular application of multinomial modeling, one needs to consider several steps. First, an experimental paradigm must be selected to generate the appropriate data for the model. The paradigm should be one in which data can be classified into a finite number of discrete, observable categories, $C_1, C_2, \ldots, C_J$. In the most general case, suppose a researcher has N total data observations. We can then define N_j as the number of observations in C_j, and we can define $\mathbf{D} = (N_1, \ldots, N_j, \ldots, N_J)$ to be the data vector of observations for the model. If the data observations are mutually independent and identically distributed with probability p_j that an observation falls into C_j, then the joint distribution of the data $\mathbf{D}$ is given by the general multinomial model

$$P(\mathbf{D}; p_1, \ldots, p_J) = N! \prod_{j=1}^{J} p_j^{N_j}/N_j!, \qquad (2)$$

where $N = \sum_{j=1}^{J} N_j$. The general model can be viewed as having the parameter space

$$\mathcal{G}_J = \{\mathbf{p} = (p_1, \ldots, p_J) | 0 \le p_j \le 1, \sum_{j=1}^{J} p_j = 1\}. \qquad (3)$$

Model Development

A substantive multinomial model can be viewed as a restriction of the general model expressed in Equation 3. Specifically, the substantive model assigns to each cognitive event a parameter value that represents the probability of that event occurring. Then the values of p_j in Equation 3 can be expressed in terms of these postulated parameters.

To put this more formally, the substantive model specifies one or more functionally independent parameters $\theta_1, \ldots, \theta_S$, $1 \le S < J$, where each θ_s lies within some interval of real numbers I_s, such as the [0, 1] interval. The parameter space for the model is then given by

$$\Omega = \{\theta = (\theta_1, \ldots, \theta_s, \ldots, \theta_S) | \theta_s \in I_s, \quad s = 1, \ldots, S\}. \qquad (4)$$

The next step in developing a model is to generate a series of equations that expresses the probabilities of the data events as a function of the model's parameters. This can be accomplished by specifying a continuous function $\mathbf{p}$ from Ω into $\mathcal{G}_J$ that gives the probability distribution over the J categories. Thus, $\mathbf{p}(\theta) = (p_1(\theta), \ldots, p_J(\theta))$ for each θ in Ω.

If the range of $\mathbf{p}$ is

$$\Omega^* = \{\mathbf{p}(\theta) | \theta \in \Omega\}, \qquad (5)$$

then Ω^* is a proper subset of $\mathcal{G}_J$, usually an S-dimensional subset. A substantive model is said to be globally identifiable if $\mathbf{p}$ is a one-to-one function from Ω onto Ω^*. In other words; $\theta \neq \theta'$ implies $\mathbf{p}(\theta) \neq \mathbf{p}(\theta')$. Global identifiability is a useful property if the researcher's goal is to measure a unique value of θ from the data $\mathbf{D}$.

Once the data events and the substantive model have been identified, there are three classes of statistical problems that can be addressed. The first involves using the empirical data to obtain estimates of the model's parameters. The second problem involves testing the goodness of fit of the substantive model relative to the general multinomial model. Finally, the third problem involves testing various hypotheses concerning how the parameters may vary over the conditions of an experiment. In the next three subsections we consider these problems briefly and provide references for those readers interested in exploring these issues in more detail.

Parameter Estimation

In this section we discuss parameter estimation from the point of view of maximum likelihood methods. We have taken this approach because maximum likelihood estimation is fairly standard, has many desirable properties, and is particularly useful for multinomial modeling. The discussion that follows explores parameter estimation for three situations: (a) the general multinomial model, (b) substantive models with one parameter, and (c) substantive models with more than one parameter. Excellent discussions of maximum likelihood estimation in

multinomial modeling can be found in Bishop, Fienberg, and Holland (1975, chapter 14), Cox and Hinkley (1974), and Elandt-Johnson (1971, chapters 11 and 12). For a comprehensive discussion of other approaches to estimation, see Lehmann (1983).

Estimation for the General Multinomial Model

To obtain maximum likelihood estimators (MLEs) for any model, it is first necessary to generate the likelihood function for that model. This is an equation that expresses the probability of the data as a function of the parameter values. Equation 2 gives the likelihood function for the general multinomial model, provided that the equation is viewed as a function of $\mathbf{p} = (p_1, \ldots, p_J)$ given the data $\mathbf{D}$.

The MLEs, $\hat{p}_1, \ldots, \hat{p}_J$ are the values that maximize the likelihood function in Equation 2 for any given data set $\mathbf{D}$. It is straightforward to show that these MLEs are unique and given by

$$\hat{p}_j(\mathbf{D}) = N_j/N, \tag{6}$$

for $j = 1, 2, \ldots, J$. Furthermore, the estimators are unbiased $E(\hat{p}_j) = p_j$, have variance

$$\mathrm{Var}(\hat{p}_j) = \frac{p_j(1-p_j)}{N},$$

and covariance

$$\mathrm{Cov}(\hat{p}_j, \hat{p}_\ell) = \frac{-p_j p_\ell}{N},$$

for $1 \leq j \neq \ell \leq J$.

Estimation for a Substantive Model With One Parameter

In this subsection we will consider parameter estimation for the simplest situation: when there is only a single parameter θ to be estimated. When this is the case, the likelihood function for the substantive model is a straightforward extension of Equation 2:

$$L(\mathbf{D}; p_1, \ldots, p_J) = N! \prod_{j=1}^{J} \frac{[p_j(\theta)]^{N_j}}{N_j!}. \tag{7}$$

As before, the MLEs of θ are the values of θ in Ω that maximize this function. From this point on we are assuming certain standard regularity conditions that are almost always satisfied in practice (see Bishop et al., 1975, chapter 14.8.1). Among other things, the regularity conditions assure global identifiability and the existence of the quantities we compute from the likelihood function.

To maximize Equation 7 for a given data set $\mathbf{D}$, it is convenient to work with the natural log of the likelihood function. Log L is a monotonically increasing function of L, and thus any θ that maximizes L will maximize log L as well. This function is given by

$$\log L(\mathbf{D}; \mathbf{p}(\theta)) = \log N! + \sum_{j=1}^{J} [N_j \log p_j(\theta) - \log N_j!]. \tag{8}$$

Equation 8 has local maxima for all θ' that satisfy the equations

$$U(\theta) = \frac{d \log L}{d\theta} = \sum_{j=1}^{J} [N_j/p_j(\theta)] \frac{dp_j(\theta)}{d\theta} = 0 \tag{9}$$

and

$$\left.\frac{d^2 \log L}{d\theta^2}\right|_{\theta=\theta'} < 0,$$

unless the maximum is achieved on the boundary of Ω.

Equation 9 may yield several local maxima θ' within Ω. However, these maxima can be screened to obtain the MLE that is a global maximum in Ω. In practice, under the assumed regularity conditions, there will usually be only a single MLE $\hat{\theta}$ in Ω. Sometimes Equation 9 can be solved analytically, yielding closed-form solutions for the parameter estimates. However, if Equation 9 is complicated, then iterative search methods may have to be used to find the MLEs. These methods are available in a number of computer packages, such as that from the International Mathematical and Statistical Libraries (IMSL; 1982). Elandt-Johnson (1971, chapter 11.7.2) and Lawley and Maxwell (1971, Appendix II) provide useful discussions of iterative methods for maximum likelihood estimation.

If the substantive model is true, and N is sufficiently large, then the regularity conditions assure that the MLE derived from the aforementioned methods will have a number of useful asymptotic properties. For one, it is asymptotically unbiased, $E(\hat{\theta}) = \theta$. Also, $\hat{\theta}$ is efficient in the sense that its variance is no larger than that of any other asymptotically unbiased estimator. Furthermore, $\hat{\theta}$ (suitably standardized) has an asymptotic normal distribution.

The value of $\hat{\theta}$ provides a useful point estimate for the measurement of θ. But with any type of measurement, one is also concerned with its reliability. Fortunately, confidence intervals can be computed for θ as well. To do this, one must compute the asymptotic variance of $\hat{\theta}$, which is given by

$$\mathrm{Var}(\hat{\theta}) = 1/I(\theta), \tag{10}$$

where $I(\theta)$ is the expected "total amount of information" about θ obtainable from the data through $\hat{\theta}$ (see Elandt-Johnson, 1971, chapter 12). The exact value of $I(\theta)$ is given by

$$I(\theta) = E\left[\frac{dU(\theta)}{d\theta}\right] = N \sum_{j=1}^{J} [1/p_j(\theta)]\left[\frac{dp_j(\theta)}{d\theta}\right]^2, \tag{11}$$

where $U(\theta)$ is defined from Equation 9. In practice, however, Equation 11 usually does not yield a numerical value because the true value of θ is unknown. When this is the case, the value of $I(\theta)$ can be approximated by computing $I(\hat{\theta})$, the estimated total amount of information. This is given by

$$I(\hat{\theta}) = \sum_{j=1}^{J} N_j [1/p_j(\hat{\theta})]\left[\frac{dp_j(\theta)}{d\theta}\right]^2_{\hat{\theta}=\theta}. \tag{12}$$

Once $\hat{\theta}$ and $\mathrm{Var}(\hat{\theta})$ have been computed, then the quantity $z = (\hat{\theta} - \theta)/\sigma(\hat{\theta})$ is asymptotically distributed as a standard normal

variate, $N(0, 1)$. Using this information, one can obtain a two-tailed $1 - \alpha$ confidence interval for θ given by

$$[\hat{\theta} - z_\alpha \sigma(\hat{\theta})] \leq \theta \leq [\hat{\theta} + z_\alpha \sigma(\hat{\theta})], \quad (13)$$

where z_α is the standard normal deviate that satisfies $Pr(Z \leq z_\alpha) = 1 - (\alpha/2)$, and $\sigma(\hat{\theta})$ is obtained from Equations 10 and 12.

Example

At this point, it may be helpful to illustrate the techniques presented so far with a simple, concrete example. Suppose one has N independent data observations, where each observation consists of two trials of a Bernoulli process with an unknown probability of success θ, $0 \leq \theta \leq 1$. Suppose also that the observations are classified into one of three categories: (a) C_1—no successes; (b) C_2—one and only one success; and (c) C_3—two successes. Let $\mathbf{D} = \langle N_1, N_2, N_3 \rangle$ be the data and p_j be the probability that an observation falls into $C_j, j = 1, 2, 3$.

Under these assumptions, the parameter space for the general multinomial model from Equation 3 becomes

$$\mathcal{G}_3 = \{\mathbf{p} = (p_1, p_2, p_3) | 0 \leq p_j \leq 1, \sum_{j=1}^{3} p_j = 1\}.$$

From Equation 6, the MLEs for the general model are given by $\hat{p}_j = N_j/N$.

The next step is to specify the substantive multinomial model. This model has a parameter space given by Equation 4: $\Omega = \{\theta | 0 \leq \theta \leq 1\}$. The substantive model can be defined by the continuous function $\mathbf{p}(\theta) = (p_1, p_2, p_3)$, which expresses the probabilities of the data events as a function of the parameter θ. It is easy to see that for this example $p_1 = (1 - \theta)^2$, $p_2 = 2\theta(1 - \theta)$, $p_3 = \theta^2$. Also Ω^* from Equation 5 given by

$$\Omega^* = \{[(1 - \theta)^2, 2\theta(1 - \theta), \theta^2] | 0 \leq \theta \leq 1\}. \quad (14)$$

It is a simple matter to determine that the model is globally identified. For example, note that if $\theta \neq \theta'$, then $p_1(\theta) = (1 - \theta)^2 \neq (1 - \theta')^2 = p_1(\theta')$, and so $\mathbf{p}(\theta) \neq \mathbf{p}(\theta')$.

Once the substantive model has been specified, one can obtain the MLE of θ. First, the likelihood function for the model is expressed as follows:

$$L(\mathbf{D}; p_1, p_2, p_3) = \frac{N!}{N_1! N_2! N_3!} [(1 - \theta)^2]^{N_1} \times [2\theta(1 - \theta)]^{N_2} [\theta^2]^{N_3}. \quad (15)$$

Then, the function $U(\theta)$ from Equation 9 can be computed by differentiating $\log L$ with respect to θ:

$$U(\theta) = \frac{d \log L}{d\theta} = \frac{-2N_1}{(1 - \theta)} + \frac{2N_2(1 - 2\theta)}{2\theta(1 - \theta)} + \frac{2N_3}{\theta}.$$

Setting $U(\theta) = 0$ and solving algebraically yields a unique MLE given by

$$\hat{\theta} = (N_2 + 2N_3)/2N. \quad (16)$$

Finally, if one wishes to obtain a confidence interval for θ, this can be accomplished by computing $\mathrm{Var}(\theta)$ from Equations 10 and 12. Working with these equations for this example yields

$$\mathrm{Var}(\hat{\theta}) = \frac{(N_2 + 2N_3)^2 (2N_1 + N_2)^2}{16N^2[N_1(N_2 + 2N_3)^2 + N_2(N_1 - N_3)^2 + N_3(2N_1 + N_2)^2]}. \quad (17)$$

Once $\hat{\theta}$ and $\mathrm{Var}(\hat{\theta})$ have been computed, they can be inserted into Equation 13 to derive the confidence interval.

To give a numerical illustration of this example, suppose that a researcher has $N = 100$ observations, resulting in the data vector $\mathbf{D} = (15, 35, 50)$. From Equation 16, the MLE $\hat{\theta}$ can be computed to be .675. Also, $\mathrm{Var}(\hat{\theta}) = .001$ and $\sigma(\hat{\theta}) = .033$ from Equation 17. From this a 95% confidence interval for θ is given from Equation 13 to be $.61 \leq \theta \leq .74$.

Estimation for a Substantive Model With Many Parameters

In actual practice, most multinomial models will usually contain more than one parameter. This will be true of the three examples we present in the next section. Fortunately, the basic estimation procedures for this situation are straightforward extensions of the one-parameter case, and thus the key ideas in this subsection parallel the previous subsection. The presentation follows closely to that of Elandt-Johnson (1971, chapter 12.12). We continue to assume the standard regularity conditions, guaranteeing a unique MLE in the interior of Ω.

If we assume that $S > 1$ in Equation 4, then the likelihood function for the substantive model is still given by Equation 7, except with $p_j(\theta)$ becoming $p_j(\theta) = p_j(\theta_1, \ldots, \theta_S)$. Again the MLEs of θ are those values $\theta_1, \ldots, \theta_S$ in Ω that maximize the natural log of the likelihood function. Under our assumptions, this amounts to simultaneously solving for the θ' that satisfies the equations

$$U_s(\theta) = \frac{\partial \log L(\mathbf{D}; \theta)}{\partial \theta_s} = \sum_{j=1}^{J} [N_j/p_j(\theta)] \left[\frac{\partial p_j(\theta)}{\partial \theta_s}\right] = 0, \quad (18)$$

for all $s = 1, 2, \ldots, S$. As can be seen, Equation 18 is an extension of Equation 9 for the multiple-parameter case. The values of θ' that satisfy this equation and also satisfy the usual constraint for a local maximum (see, e.g., Cox & Hinkley, 1974, p. 283) can be screened to obtain the MLE $\hat{\theta}$ guaranteed by the regularity conditions.

Although closed-form solutions can sometimes be derived for the parameter estimates, it is often the case that iterative search methods must be used to obtain $\hat{\theta}$. If iterative methods are required, it is recommended that the search method be guided by the Fisher information matrix, $I(\hat{\theta})$ (see Lawley & Maxwell, 1971, Appendix II). Such methods provide an estimate of $I(\theta)$ at the end of the search, and knowledge of this allows one to estimate the variance–covariance matrix for $\hat{\theta}$. As seen next, the variance–covariance matrix is an important component for computing confidence intervals and confidence regions for the parameters.

If N is sufficiently large, and if the model is true, then $\hat{\theta}$ is

asymptotically unbiased. Furthermore, $\hat{\theta}$, suitably standardized, is asymptotically distributed as a multivariate normal distribution with mean vector θ (the true parameter vector) and asymptotic variance–covariance matrix given by

$$\mathbf{V}(\hat{\theta}) = \mathbf{I}^{-1}(\theta). \tag{19}$$

In this equation, $\mathbf{I}(\theta)$ is the $S \times S$ Fisher information matrix whose stth term is given by

$$\mathbf{I}_{st}(\theta) = N \sum_{j=1}^{J} [1/p_j(\theta)] \left(\frac{\partial p_j(\theta)}{\partial \theta_s}\right) \cdot \left(\frac{\partial p_j(\theta)}{\partial \theta_t}\right). \tag{20}$$

As with the one-parameter case, a useful approximation to Equation 19 is to substitute the estimated information matrix $\mathbf{I}(\hat{\theta})$. The general term for this matrix is given by

$$\mathbf{I}_{st}(\hat{\theta}) = \sum_{j=1}^{J} N_j [1/p_j(\hat{\theta})] \left(\frac{\partial p_j(\theta)}{\partial \theta_s}\right) \cdot \left(\frac{\partial p_j(\theta)}{\partial \theta_t}\right)_{\theta = \hat{\theta}}, \tag{21}$$

$s, t = 1, 2, \ldots, S$.

Once the variance–covariance matrix has been computed, confidence regions for $\hat{\theta}$ can be obtained using standard methods that we will not detail here (see Johnson & Kotz, 1972, chapter 35). It is important to realize that confidence regions for any subset of parameters in $\hat{\theta}$ can be computed. For example, confidence regions for a pair of parameters will consist of ellipses within a two-dimensional space. For more detail on this, the case of two-dimensional regions is discussed in Batchelder and Riefer (1986) and also in Johnson and Kotz (1972, chapter 36).

The case of confidence intervals for individual parameters can be handled by computing $\mathbf{V}(\hat{\theta})$ from Equation 19 and inserting the diagonal terms of this matrix into Equation 13. However, a disadvantage of this method is that one must be able to compute the estimated information matrix $\mathbf{I}(\hat{\theta})$ and then invert it to obtain $\mathbf{V}(\hat{\theta})$. For situations where these computations are unwieldy, Elandt-Johnson (1971, Equation 12.55) provides an approximate formula for the variance of each parameter:

$$\mathrm{Var}(\hat{\theta}_i) \doteq \sum_{j=1}^{J} \left(\frac{\partial \theta_i}{\partial p_j}\right)^2 \frac{p_j(1 - p_j)}{N} - 2 \sum\sum_{\ell < m} \left(\frac{\partial \theta_i}{\partial p_\ell}\right)\left(\frac{\partial \theta_i}{\partial p_m}\right) \frac{p_\ell p_m}{N}. \tag{22}$$

To use this formula, set $p_j = N_j/N$ and view the MLE $\hat{\theta}_i$ as a function of the p_js, that is,

$$\frac{\partial \theta_i}{\partial p_j} = \left(\frac{\partial \hat{\theta}_i}{\partial p_j}\right)\Big|_{\hat{\mathbf{p}} = \mathbf{p}}, \tag{23}$$

where $\hat{\mathbf{p}} = (N_1/N, \ldots, N_J/N)$.

This method is illustrated in the first empirical example, and the method based on the estimated information matrix is illustrated in the third empirical example (and also in Batchelder & Riefer, 1986).

Goodness of Fit

Once parameter estimates have been computed for the model, a separate issue is whether the model adequately fits the data. Of course, the general multinomial model supposes no constraints on the true $\mathbf{p}$ in $\mathcal{S}_J$ and thus always fits the data perfectly. However, the substantive multinomial model will often contain fewer parameters than the general model ($S < J - 1$). More formally, the substantive model requires that $\mathbf{p}$ be contained within Ω^*, which is an S-dimensional subset of $\mathcal{S}_J$. The substantive model, with its fewer parameters, can be said to fit the data if its description of the data is comparable to that achieved by the general model.

Bishop et al. (1975, chapter 14) treated this goodness-of-fit problem for a variety of estimation methods. They showed that the MLE $\hat{\theta}$ is the member of Ω that minimizes the "distance" between the MLE $\hat{\mathbf{p}}$ of the general model given by Equation 6 and the "fitted" probability distribution $\mathbf{p}(\theta)$ in Ω^*. This distance is given by

$$G^2[\hat{\mathbf{p}}, \mathbf{p}(\theta)] = 2 \sum_{j=1}^{J} N_j \log[N_j/Np_j(\theta)]. \tag{24}$$

Under the assumed regularity conditions, Equation 24 is minimized by the vector $\mathbf{p}(\hat{\theta})$, where $\hat{\theta}$ is the MLE of θ in Ω. If the model is true, then the asymptotic distribution of $G^2[\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta})]$, called the log-likelihood ratio statistic, is a chi-square with $J - S - 1$ degrees of freedom. Thus, if N is sufficiently large, one can test the hypothesis that the model fits the data by computing G^2 and seeing if it is sufficiently small.

Fortunately, the computation of G^2 is simplified by noting that the goodness-of-fit test based on G^2 is actually a likelihood ratio test (Hogg & Craig, 1978, chapter 7). The likelihood ratio test is based on the statistic

$$\lambda = \frac{L[\mathbf{D}; \mathbf{p}(\hat{\theta})]}{L(\mathbf{D}; \hat{\mathbf{p}})} \leq 1, \tag{25}$$

which contains expressions originally defined in Equations 7 and 8. The term λ is simply the maximized value of the likelihood function for the substantive model, divided by the maximized value of the likelihood function for the general model. These values are easy to compute once the MLEs for each function have been obtained. It is straightforward to show that

$$-2 \log \lambda = G^2[\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta})], \tag{26}$$

and thus G^2 is a function of the likelihood ratio statistic.

Hypothesis Testing

In this subsection we briefly discuss how to test various hypotheses concerning the parameters of a multinomial model. Hypotheses about parameter values are often of critical concern in research situations. For example, research data will often come from several independent experimental conditions, representing the various levels of an independent variable. One might therefore wish to test the hypothesis that certain parameters have a constant value across the conditions. Hypotheses for a model's parameters can take on many forms, but basically any hypothesis will constitute some restriction of the dimensionality of the parameter space Ω. The restrictions considered in this article are of two types: (a) a set of parameters are constrained

to be equal to each other, and (b) one or more parameters are set equal to specific values.

Suppose we have a substantive multinomial model M_1, with parameter space Ω_1 given by Equation 4. If we put a restriction on one or more of the model's S parameters, this in essence creates a restricted version of M_1, which we can denote as M_2. Suppose there are a total of R such restrictions on the parameters, $R < S$. Then the restricted model M_2 will have a new parameter space Ω_2, which will be a nested subset contained in Ω_1. In addition, Ω_2^* will be a proper subset of Ω_1^*, usually an $S - R$ dimensional subset.

The main question of interest is whether the restricted version of the model is able to capture the true parameter vector of the model. If not, then the restriction placed on the model's parameters is not a valid one and can be rejected. If θ is the true parameter vector, then we are basically testing the null hypothesis that θ lies in the restricted parameter space Ω_2 versus the alternative hypothesis that θ is not in Ω_2.

It is straightforward to develop a likelihood ratio test of these hypotheses. The asymptotic test is based on the statistic

$$G^2[\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta})] - G^2[\hat{\mathbf{p}}, \mathbf{p}(\hat{\hat{\theta}})] = -2N \sum_{j=1}^{J} (N_j/N) \log[p_j(\hat{\theta})/p_j(\hat{\hat{\theta}})]. \quad (27)$$

In this equation, $\hat{\theta}$ is the vector of MLEs within Ω_1 for the full version of the model M_1. Also, $\hat{\hat{\theta}}$ is the vector of new MLEs that maximize the likelihood function for the restricted model M_2. The value of $\hat{\hat{\theta}}$ may differ from $\hat{\theta}$ because it is constrained to be within the restricted parameter space Ω_2.

If N is sufficiently large, then the statistic in Equation 27 is approximately chi-square with S − R degrees of freedom. As with the goodness-of-fit test, computation of this statistic is simplified by the fact that Equation 27 can be rewritten as a function of the likelihood ratio:

$$G^2[\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta})] - G^2[\hat{\mathbf{p}}, \mathbf{p}(\hat{\hat{\theta}})] = -2 \log[L(\mathbf{D}; \hat{\hat{\theta}})/L(\mathbf{D}; \hat{\theta})]. \quad (28)$$

Joint Multinomial Models

In some of the examples that we present in the next section, the models are not technically multinomial models as developed so far. Rather, they are more accurately termed *joint multinomial models.* This type of model is needed when the experimental observations are separated into two or more disjoint data sets based on the experimental design. When there is more than one set of data, a possibly different multinomial model applies to each set independently. Then the likelihood function is given by the product of the separate multinomial likelihood functions. Familiar examples of this are models for yes–no signal detection (Green & Swets, 1966), where data are separated into signal and noise trials. Although they are not usually thought of as multinomial models, models of signal detection can in fact be classified as joint multinomial models because the signal and noise data are each analyzed by separate binomial distributions.

More formally, suppose there are K separate category systems for the data, each possibly containing a different number of categories. If J_k is the number of categories for the kth system, then $C_{1k}, C_{2k}, \ldots, C_{J_k k}$ are the mutually exclusive and exhaustive categories for that system. The data observations are then partitioned into K separate data sets based on the experimental design, with $\mathbf{D}_k = \langle N_{1k}, \ldots, N_{J_k k} \rangle$ being the data set for each separate system of categories.

The parameter space for this joint multinomial model is given by Equation 4 as before, and the likelihood function for the kth data set is given by

$$L_k(\mathbf{D}_k; \theta) = N_k! \prod_{j=1}^{J_k} [p_{jk}(\theta)]^{N_{jk}}/N_{jk}!, \quad (29)$$

where

$$N_k = \sum_{j=1}^{J_k} N_{jk}.$$

The likelihood function for the joint multinomial model is simply the product of the K individual likelihood functions. If the entire collection of data sets is denoted by $\mathbf{D}$, then this likelihood function is

$$L(\mathbf{D}; \theta) = \prod_{k=1}^{K} L_k(\mathbf{D}_k; \theta), \quad (30)$$

and, of course,

$$\log L = \sum_{k=1}^{K} \log L_k.$$

The results of the previous subsections apply to joint multinomial models, with the obvious notational changes. In particular, the basic techniques for parameter estimation, goodness-of-fit, and hypothesis testing still apply to the joint likelihood function in Equation 30. The details will be illustrated in the empirical examples to follow.

Special Considerations

In this subsection, we briefly mention several practical problems that researchers may face with multinomial modeling. Usually these problems are ignored in applications, and some of them are appropriate for further research.

The first problem concerns the data collection procedure for testing a model. Any substantive model as well as the general multinomial model requires the sampling assumptions that the observations be independently and identically distributed over the J categories. Often, in practice, data are collected from several subjects, and several observations are obtained from each subject. This practice is typical in learning and memory modeling. Such an approach to data collection runs the risk of violating both the identically distributed and the independence assumptions. In particular, possible individual differences and within-subject dependence are risks in such a practice and may lead to estimated confidence intervals that are too narrow. Informally, it appears that many multinomial models are fairly robust under small violations in the sampling assumptions. However, if this is of concern for a particular model, it is usually

easy to introduce violations in the model and study their effects in Monte Carlo studies. This approach is followed in the first example in the next section.

A second problem often occurs in obtaining the MLE. Sometimes analytic solutions lead to estimates outside Ω, and unconstrained iterative search methods may lead to the same problem. If the model is true, such problems are likely to occur if the true θ lies near or on the boundary of Ω. If the model is false, then data can easily exist where standard methods fail to obtain an MLE in Ω. In cases like these, iterative search methods that impose boundary constraints on the search are required (e.g., IMSL, 1982; for a general discussion see Lawley & Maxwell, 1971, p. 143). Even if, say, θ_s has an MLE $\hat{\theta}_s$ on the boundary of I_s, the other parameters may have MLEs that are interior points. Such parameters can be identified adequately and usually have the standard maximum likelihood asymptotic properties.

A third problem deals with the minimum number of data observations needed for reliable estimation of the parameters. As we stated earlier, if a model is true, and N is large enough, the asymptotic theory of maximum likelihood estimation enables one to get confidence intervals for the model's parameters. In practice, this raises the question of how large an N is required. As with any mathematical approximation, one can always use it if the approximation is accurate enough for the investigator's purposes. For a given level of accuracy, the requisite level of N will depend on the particular model. Therefore, if accurate confidence intervals are an important issue, it may be useful to conduct Monte Carlo studies for various parameter values and N to examine the usefulness of the asymptotic theory. In practice, most researchers collect at least $N = 100$ observations, and in our opinion this is a useful rule of thumb if Monte Carlo studies are unavailable.

A fourth problem is that the goodness-of-fit test in Equation 26 is frequently rejected if N is very large. This is because most substantive models are only approximations to reality and therefore are technically false. Because the power of the goodness-of-fit test increases with N, rejection of a good, but technically incorrect, model becomes very likely with very large N.

If the model is incorrect (misspecified), Bishop et al. (1975) remarked that the asymptotic properties of $\hat{\theta}$ and $G^2[\hat{\mathbf{p}}, \mathbf{p}(\theta)]$ are not as developed here. Some work has been done on the asymptotic properties of these likelihood statistics for general models (e.g., see White, 1982, for results and further references). In particular, quasi maximum likelihood estimators can be defined with desirable properties, and robust inference procedures can sometimes be obtained based on the information matrix. This is an interesting subject for further research in the context of multinomial modeling. Until some of this work becomes standard in the psychological modeling area, it seems reasonable to continue to analyze data with technically incorrect models if they explain much of the variance in the data. On the other hand, goodness-of-fit statistics should be routinely computed and presented in such cases.

Finally, the property of global identifiability is a useful and necessary one for most of the apparatus discussed in this section. It is sometimes possible to check global identifiability directly by showing that if $\theta \neq \theta'$, then $\mathbf{p}(\theta) \neq \mathbf{p}(\theta')$, for example, by obtaining the inverse $\mathbf{p}^{-1}$ from Ω^* into Ω. Actually the asymptotic results require a stronger form of identifiability, namely, that $\mathbf{p}^{-1}$ is continuous at all θ in the interior of Ω. Most substantive models will be identifiable in this sense; however, this issue should be addressed with either analytic or numerical work in any application of multinomial modeling where identifiability is not obvious.

Empirical Examples

In this section we use multinomial modeling to explore several theoretical issues dealing with cognitive processes. Three multinomial models are presented, each one designed to address a specific issue in a different area within the field of human memory. The examples also provide a concrete illustration of the general techniques for multinomial modeling outlined in the previous section, including model development, parameter estimation, goodness-of-fit, and hypothesis testing.

The first example is a model originally presented by Greeno, James, DaPolito, and Polson (1978) to investigate a controversial finding in the area of interference theory known as the *independent retrieval phenomenon.* We present here a detailed analysis of their model that goes beyond what was presented in Greeno et al. (1978). For example, we present techniques of parameter estimation and hypothesis testing that were not included in their original discussion of the model. The second example is a model for examining storage and retrieval processes in human memory that was originally developed by Batchelder and Riefer (1980; see also Riefer, 1982). Although this model has been extensively analyzed elsewhere (Batchelder & Riefer, 1986), a new application of the model is reported that explores the role storage and retrieval play in retroactive inhibition. The third model deals with discrimination learning and is a new model that has been developed specifically for this article. The model is capable of taking the recall of discriminable stimuli and separating that recall into learnability and confusability factors. The usefulness and validity of this model is demonstrated using a standard experiment on discrimination learning.

Proactive Inhibition

Theoretical Background

Research in interference theory has a long history in the field of memory (see Underwood, 1983, for a review). Proactive inhibition is one form of interference in which the recall of material is interfered with by information learned at an earlier date. In the standard experiment dealing with proactive inhibition, a list of paired associates is presented for learning (A–B), followed by a second list of associates where the same stimuli are paired with new responses (A–C). Recall of the C responses is typically poorer after the A–B list is learned, in comparison with a control group that does not receive the A–B list.

A controversial finding in the area of interference theory is the independent retrieval phenomenon (Martin, 1971). A number of research studies (e.g., DaPolito, 1966; Eich, 1982; Wichawut & Martin, 1971) have found the recall of B and C to be stochastically independent, even when strong interference effects are evident. The controversy centers around the interpre-

tation of this finding. Martin (1971, 1981) claimed that independent retrieval presents compelling evidence against associative interference theories (Postman & Underwood, 1973). According to these theories, the strength of the B associations should have a direct effect on the C responses. Specifically, if B competes or interferes with the recall of C, then stronger B associations should inhibit C more. From this line of reasoning, interference theory predicts a negative correlation between B and C that is simply not evident in the data. However, Martin's viewpoint has come under strong criticism from Hintzman (1972, 1980) on methodological grounds. The crux of Hintzman's argument is that the finding of independent retrieval is based on correlational evidence, from which conclusions about cause and effect are not necessarily justified. Specifically, positive correlations between B and C may exist (due to subject or item differences) that would tend to obscure the negative correlation predicted by interference theory.

In our opinion, a strong test of the independent retrieval phenomenon has been provided in an experiment by DaPolito (reported in Greeno et al., 1978, chapter 7). In this experiment, DaPolito directly manipulated the strength of the A–B associations by presenting the A–B list for one, two, or three study trials. Even though the recall of B increased with trials, and even though there was strong proactive inhibition in the experiment, the decrement of A–C attributable to proactive inhibition was unaffected by the strength of B. As Greeno et al. pointed out, this study provides strong, noncorrelational evidence that the recall of the C responses is independent of the strength of B.

As part of their discussion of this research, Greeno et al. (1978) developed a simple process model capable of analyzing data from proactive inhibition experiments. The model was used to provide a quantitative illustration of some of their findings and conclusions about the independent retrieval phenomenon. However, Greeno et al. stopped short of formulating the model statistically as a multinomial model or applying the model in any detail to their experimental data. Thus, as our first example of multinomial modeling, we will present a complete analysis of Greeno et al.'s model for proactive inhibition. We begin first by developing the model in detail, followed by an application of the model to DaPolito's experimental data.

Model Development

The experimental paradigm used by DaPolito follows closely to the standard experiment for studying proactive inhibition. Subjects are presented with a list of paired associates, followed by a standard recall test. Within the list, a subset of the paired associates are presented only once and therefore have only one correct response. These are the control stimuli. For the remaining paired associates, the stimulus member occurs twice within the list, each time with a different response (i.e., A–B followed by A–C). These are the experimental stimuli. If the recall of the C responses is poorer than the controls, then proactive inhibition can be assumed to have occurred.

In this experimental task, some stimuli have two correct responses, whereas other stimuli have only one response. To avoid biasing the subjects' recall by decisions about whether an item has one or two responses, on test trials subjects are always asked to give two responses to each stimulus. (This is a variation of a recall procedure known as modified modified free recall.) Of course, for those stimuli with two responses this is no problem. For stimuli with only one response, subjects are instructed to give that response plus any other.

For the analysis of the model to be described next, only the recall data from the A–B and A–C stimuli will be considered here. It should be easy to see that there are four distinct recall events for this situation: (a) E_1, both the B and C responses are correctly recalled (BC); (b) E_2, only B is recalled ($B\bar{C}$); (c) E_3, only C is recalled ($\bar{B}C$); and (d) E_4, neither response is recalled ($\bar{B}\bar{C}$). Let N_i be the frequency of E_i in an experiment, with

$$N = \sum_{i=1}^{4} N_i.$$

Greeno et al.'s (1978) model is based on the idea that stimuli are encoded in memory as a series of features. Recognition and retrieval of stimuli are then determined by a retrieval network that allows discrimination between different stimuli on the basis of these features. With these assumptions, the model attempts to describe the recall of the paired associates as a consequence of three hypothetical, dichotomous events.

Accessing the retrieval network. Stimuli that share different responses (such as A–B and A–C) are assumed to be stored within the same retrieval network. During a test trial, this retrieval network may or may not be successfully accessed. Define p to be the probability of accessing the retrieval network for an item, $0 \leq p \leq 1$.

Retrieving the A–B association. Even though A–B and A–C may share the same stimulus network, it is assumed that each item occupies a different terminal node of that network. Even if the retrieval network is successfully accessed, the A–B item may or may not be retrieved. Define q to be the probability of retrieving the A–B item given that the retrieval network is accessed, $0 \leq q \leq 1$. In this event, B is given as a correct response to A.

Retrieving the A–C association. If the retrieval network is successfully accessed, it is assumed that the A–C item is retrieved with probability r, $0 \leq r \leq 1$. In this event, C will also be given as a correct response to A.

In multinomial models such as this one, in which the data statistics are assumed to be a function of dichotomous events, it is often convenient to express the model in tree diagram form. Figure 1 presents the tree diagram for Greeno et al.'s (1978) model. From this tree structure, it is easy to write expressions for the probabilities of each data event:

$$P(\mathrm{BC}) = pqr \tag{31a}$$

$$P(\mathrm{B}\bar{\mathrm{C}}) = pq(1 - r) \tag{31b}$$

$$P(\bar{\mathrm{B}}\mathrm{C}) = p(1 - q)r \tag{31c}$$

$$P(\bar{\mathrm{B}}\bar{\mathrm{C}}) = (1 - p) + p(1 - q)(1 - r). \tag{31d}$$

In summary, the current model consists of the data vector $\mathbf{D} = (N_1, N_2, N_3, N_4)$, where N_1 corresponds to recall event BC, N_2 to $B\bar{C}$, N_3 to $\bar{B}C$, and N_4 to $\bar{B}\bar{C}$. The parameter space can be represented as $\Omega = \{\theta = (p, q, r) | 0 \leq p, q, r \leq 1\}$.

As described in the previous section, MLEs for the model's

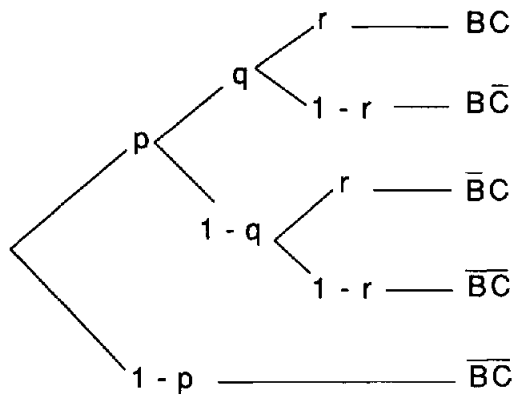

Figure 1. Tree diagram for Greeno, James, DaPolito, and Polson's (1978) model for proactive inhibition. (BC = both responses correctly recalled; $B\bar{C}$ = only B recalled; $\bar{B}C$ = only C recalled; and $\bar{B}\bar{C}$ = neither response recalled.

parameters can be found by determining the values of the parameters that maximize the likelihood function. The likelihood function for this model is given by

$$L = (N!/N_1!N_2!N_3!N_4!)(pqr)^{N_1}[pq(1-r)]^{N_2} \times [p(1-q)r]^{N_3}[1-p+p(1-q)(1-r)]^{N_4}, \quad (32)$$

where L has previously been defined in Equation 7.

Fortunately, the calculus methods in Equation 18 can be used to derive closed-form solutions for the MLEs of the model. The results are

$$\hat{p} = \frac{(N_1+N_2)(N_1+N_3)}{(N_1)(N)} = \frac{(P_1+P_2)(P_1+P_3)}{P_1} \quad (33a)$$

$$\hat{q} = \frac{N_1}{N_1+N_3} = \frac{P_1}{P_1+P_3}, \quad (33b)$$

$$\hat{r} = \frac{N_1}{N_1+N_2} = \frac{P_1}{P_1+P_2}, \quad (33c)$$

where $P_i = N_i/N$.

DaPolito's Experiment

We will apply the model to the DaPolito experiment discussed earlier (Experiment 2 in Greeno et al., 1978, chapter 7). The main experimental variable in this study was the number of A–B presentations. Specifically, subjects received one, two, or three A–B presentations for each item before studying A–C. They also received two test trials for this material.

According to Greeno et al. (1978), increasing the number of A–B presentations should function to strengthen the B response. This should in turn be reflected in an increase in the value of parameter q, which measures the strength of the A–B associations. Moreover, associative interference theory predicts that the value of r (which measures the recall of C) should decrease as q increases, because of the interference that the A–B associations have on C. But Greeno et al. argued that the recall of C is independent of B. Thus, although they agreed that q should increase with extra A–B presentations, they predicted that r should be unaffected by this manipulation.

The multinomial model can now be used to test these predictions. Table 1 presents the original data from the DaPolito study. The recall frequencies for events BC, $B\bar{C}$, $\bar{B}C$, and $\bar{B}\bar{C}$ are given for each of the three A–B presentation conditions and for both test trials. (Because they are not central to the model's analysis, the recall data for the control stimuli are not presented here; however, they reveal that there were strong proactive inhibition effects in the experiment.)

Table 1 also presents the estimates for p, q, and r from Equation 33. Two of the data sets yielded an estimate of p from Equation 33a that slightly exceeded the upper limit of one. As it turns out, it is possible for Equation 33a to yield impossible values for the parameter p, that is, values that are outside the [0, 1] interval. This could occur even if the model is true (especially if p is close to one), and in fact it did occur for two of these data sets. For these cases, instead of using Equation 33, we used the computer subroutine ZXMIN from IMSL (1982) to search for the best fitting set of parameters, under the constraint that the solutions be within the [0, 1] interval.

As can be seen from Table 1, the value of q does increase as the number of presentations increases. This is true for both the first and second test. However, the value of parameter r does not show any corresponding decrease as a function of the number of A–B presentations.

The MLEs in Table 1 provide a measure for the probabilities of the hypothesized cognitive processes of the model. But with any type of measurement, one should also be interested in the reliability of the measurements. Constructing confidence intervals for each of the parameters is a way of assessing this reliability. As described in the previous section, one way to compute these confidence intervals is to use the technique in Equation 22 to obtain the variances of the MLEs. To use Equation 22, it is first necessary to obtain the partial derivatives for each parameter, $\partial\hat{p}/\partial P_i$, $\partial\hat{q}/\partial P_i$, and $\partial\hat{r}/\partial P_i$. This is done by differentiating Equation 33 with respect to each P_i, where, as before,

Table 1
Recall Frequencies and Parameter Estimates for DaPolito's (1966) Experiment on Proactive Inhibition

No. of A–B presentations	BC	$B\bar{C}$	$\bar{B}C$	$\bar{B}\bar{C}$	$\hat{p}$	$\hat{q}$	$\hat{r}$
Test 1							
1	24	65	30	61	1.00[a]	.49	.30
2	40	91	12	37	.95	.77	.31
3	50	98	7	25	.94	.88	.34
Test 2							
1	25	58	29	68	.99	.46	.30
2	28	97	18	37	1.00[a]	.69	.26
3	52	97	10	21	.99	.84	.35

Note. BC = both responses correctly recalled; $B\bar{C}$ = only B recalled; $\bar{B}C$ = only C recalled; $\bar{B}\bar{C}$ = neither response recalled. Column headings are estimates of parameters from Equation 33.

[a] The estimate for p slightly exceeded the upper limit of 1.00 for this data set. In this case, we used iterative search methods to obtain the parameter estimates.

Table 2
95% Confidence Intervals for Parameters q and r From the Proactive-Inhibition Model

No. of A–B presentations	Parameter	
	q	*r*
Test 1		
1	(.36, .63)	(.21, .39)
2	(.66, .88)	(.23, .38)
3	(.79, .96)	(.26, .41)
Test 2		
1	(.33, .60)	(.20, .40)
2	(.55, .84)	(.18, .33)
3	(.75, .93)	(.27, .43)

$P_i = N_i/N$, $i = 1, 2, 3$. These quantities can then be inserted into Equation 22 along with the P_i. The results are

$$\mathrm{Var}(\hat{p}) \doteq \frac{(P_1^2 - P_2P_3)^2 + P_1P_2(P_1 + P_3)^2 + P_1P_3(P_1 + P_2)^2 - P_1(P_1 + P_2)^2(P_1 + P_3)^2}{NP_1^3}, \tag{34a}$$

$$\mathrm{Var}(\hat{q}) = \frac{P_1P_3}{N(P_1 + P_3)^3}, \tag{34b}$$

$$\mathrm{Var}(\hat{r}) = \frac{P_1P_2}{N(P_1 + P_2)^3}. \tag{34c}$$

Once these variances have been computed, the standard deviations can be inserted into Equation 13 to obtain approximate confidence intervals for the parameters. We have computed the 95% confidence intervals for the parameters q and r, which are presented in Table 2. Unfortunately, it was not feasible to construct confidence intervals for p because most of the estimates of p are at or close to the upper boundary of one.

We emphasize that the parameter estimates and asymptotic confidence intervals in Tables 1 and 2 are only approximate, because they are based on the assumptions that N is sufficiently large and that the data observations are identically distributed. As we mentioned earlier, small sample sizes or individual differences in the parameter values may lead to violations of these assumptions, and an important question is what happens to the parameter estimates when this happens. For example, these violations may introduce systematic bias in the parameter estimates or result in confidence intervals that are too narrow. To explore this, we have conducted Monte Carlo simulations of Greeno et al.'s (1978) model under the assumption that $p = 1$, and q and r vary independently across observations. We chose the beta distribution to represent the variation in q and r because it is a convenient and flexible distribution for representing individual differences within the unit internal.

We first conducted a simulation of the model using very small sample sizes (30 data observations) and large individual differences in the model's parameters (standard deviations exceeding .3). As expected, substantial bias in the parameter estimates and noticeable underestimation of the confidence intervals were introduced under these extreme conditions. However, these deviations were negligible (less than 5% error) when the number of data observations exceeded 120, even for extreme variances. Moreover, when the variability of the parameters was only moderate or small (a standard deviation as high as .15), the deviations were negligible even for small samples as low as 50 data observations. Thus, the estimation procedures for the model seem fairly robust, and the 180 data observations obtained for DaPolito's (1966) experiment appear sufficient for reliable parameter estimation.

Hypothesis Testing

In Greeno et al.'s (1978) multinomial model, the number of parameters of the model (three) is equal to the number of independent data events. Thus, there are no degrees of freedom with which to test the goodness of fit of the model. However, it is possible to conduct hypothesis testing of the model's parameters. Although DaPolito's (1966) experiment revealed differences in the parameters across the experimental conditions, a separate question concerns whether these differences are statistically reliable. In other words, it would be useful to be able to test the null hypothesis that the parameters are a constant value across conditions, versus the alternate hypothesis that they are significantly different.

These hypotheses can be assessed using the likelihood ratio test for multinomial models described earlier. If there are J sets of data (e.g., the J conditions in an experiment), then the hypothesis tests are based on the joint likelihood function for all of the J data sets combined. This likelihood function is a straightforward extension of Equation 32 and is

$$L = \prod_{j=1}^{J} \frac{N_j!}{N_{1j}!N_{2j}!N_{3j}!N_{4j}!} [p_jq_jr_j]^{N_{1j}} \times [p_jq_j(1 - r_j)]^{N_{2j}}[p_j(1 - q_j)r_j]^{N_{3j}} \times [(1 - p_j) + p_j(1 - q_j)(1 - r_j)]^{N_{4j}}. \tag{35}$$

For the purposes of this example, we will use the likelihood ratio test to test two separate hypotheses: (a) q_j is a constant across J, and (b) r_j is a constant across J.

Hypothesis 1: q_j is a constant. To test the null hypothesis that q_j is a constant across J, it is first necessary to rewrite the likelihood function in Equation 35 with q_j as a constant. Then it is a relatively simple matter to use the calculus methods in Equation 18 to derive new MLEs for this function. Again, closed-form solutions for these MLEs can be found, which are

$$\hat{q} = \frac{\sum_j N_{1j}}{\sum_j (N_{1j} + N_{3j})} \tag{36a}$$

$$\hat{r}_j = \frac{(N_{1j} + N_{3j})\hat{q}}{(N_{1j} + N_{3j})\hat{q} + N_{2j}} \tag{36b}$$

$$\hat{p}_j = \frac{(N_{1j} + N_{3j})\hat{q} + N_{2j}}{N_j\hat{q}}. \tag{36c}$$

The MLEs from Equation 36 represent the parameter estimates for the model under the restriction that q is a constant across J. Once these have been computed from the data, they can be used in conjunction with the MLEs for the unrestricted model (from Equation 33) to complete the likelihood ratio test. By inserting both sets of MLEs into Equation 28, the likelihood ratio statistic G^2 can be computed, which in this case approximates a χ^2 statistic with $J - 1$ degrees of freedom.

The likelihood ratio test can now be conducted on the data in Table 1 to test the hypothesis that q is a constant across the three A–B presentations. We have performed this test across both test trials, yielding a G^2 statistic with 4 degrees of freedom (2 degrees of freedom for each test trial). The test reveals that the null hypothesis of constant q can be rejected, $G^2(4) = 45.69$, $p < .01$. Thus, the value of q significantly increases as the number of A–B presentations increases.

Hypothesis 2: r_j is a constant. Although the significant increase in q is certainly to be expected, a more crucial issue is the effect that presentations of A–B have on the parameter r. Although the values of r do change somewhat over conditions, the question can be asked if these changes are large enough to be statistically significant. This would involve testing the null hypothesis that r is a constant across J.

The likelihood ratio test can be used to test this hypothesis, in a manner analogous to the test for constant q. New MLEs can be derived for the model's parameters, under the restriction that r is now a constant. These MLEs are

$$\hat{r} = \frac{\sum_j N_{1j}}{\sum_j (N_{1j} + N_{2j})} \tag{37a}$$

$$\hat{q}_j = \frac{(N_{1j} + N_{2j})\hat{r}}{(N_{1j} + N_{2j})\hat{r} + N_{3j}} \tag{37b}$$

$$\hat{p}_j = \frac{(N_{1j} + N_{2j})\hat{r} + N_{3j}}{N_j \hat{r}}. \tag{37c}$$

The MLEs from Equation 37, along with the MLEs for the unrestricted model from Equation 33, can now be used to complete the test. When this is done, the test reveals that the values of parameter r do not significantly differ across conditions, $G^2(4) = 6.47$, $p > .10$.

Conclusions

The results of the multinomial model support Martin's (1981) and Greeno et al.'s (1978) interpretation of the independent retrieval phenomenon. As predicted, the value of q significantly increases as a function of the number of A–B presentations, but without a corresponding decrease in the value of r.

One important thing about these results is that nothing in the multinomial model itself constrains the results to come out as they do. For example, even though $P(C)$ is unaffected by the number of A–B presentations, it does not necessarily follow from this that the value of parameter r must be constant. As Greeno et al. (1978) pointed out, the model makes the straightforward prediction that $P(C) = pr$. Thus, one possible pattern of results might be for p to increase and r to decrease as a function of A–B strength in such a way that pr and $P(C)$ appear to be constant. This would be a pattern of results, quite possible for certain patterns of the data statistics, that would be damaging to Martin's and Greeno et al.'s theory. But instead, the model's parameters behave as predicted. Therefore, the results of the model, coupled with the fact that retrieval independence is evident even after a direct experimental manipulation of A–B strength, question the validity of associative theories of proactive interference.

Storage and Retrieval

The second example consists of a model, originally presented in Batchelder and Riefer (1980), designed to measure, separately, storage and retrieval processes in human memory. The difference between storage and retrieval plays an important part in many theories of human memory (see Batchelder & Riefer, 1986, for a partial review). Because of this, a multinomial model capable of examining these processes separately has many potential uses for memory research and theory. As an illustration of this, and in an attempt to relate this example to the previous one, we will conduct a storage–retrieval analysis of a common phenomenon related to interference theory—retroactive inhibition.

Theoretical Background

In contrast to proactive inhibition, retroactive inhibition is a form of interference in which recall of material is inhibited by interpolated material learned at a later time. One theoretical issue in this area is whether this recall decrement is due to a storage loss or a retrieval failure (or both). Storage explanations of retroactive inhibition center on the "unlearning" of item associations (Barnes & Underwood, 1959; Melton & Irwin, 1940). Presumably, this unlearning results from a weakening or loss of memory traces caused by the interfering material (Earhard, 1976; Reynolds, 1977). In contrast, retrieval-based theories of interference state that interfered items are still fully available in memory but have trouble being retrieved (Miller, Kasprow, & Schactman, 1986; Newton & Wickens, 1956; Postman & Stark, 1969). This idea is supported by the finding that retroactive inhibition disappears when recognition memory is used, because recognition presumably eliminates retrieval problems.

Another piece of supporting evidence for a retrieval explanation of retroactive inhibition comes from an experiment by Tulving and Psotka (1971). In their study, subjects learned a series of free-recall lists containing words related by categories. Subjects memorized up to six of these lists and then attempted to recall all of them in a final free recall. The main variable of interest was the recall of the first list as a function of the number of interpolated lists after it. Tulving and Psotka found that the number of words and the number of categories recalled from the first list decreased as a function of the number of interpolated lists, reflecting the effects of retroactive inhibition. However, the number of words recalled per category remained relatively constant. From this, Tulving and Psotka concluded that retroactive inhibition influences the accessibility of categories, but not the availability of information within each category.

The main conclusion to be drawn from their study is that retroactive inhibition affects the retrieval of items but not necessarily their storage. However, this conclusion is not universally accepted. In particular, some researchers (Nelson & Brooks, 1974; Sowder, 1976) have attempted to replicate Tulving and Psotka's results and have found that the number of words recalled per category is inhibited somewhat by retroactive inhibition. Because of this, these researchers have concluded that retroactive inhibition may have some effects on storage, contrary to Tulving and Psotka's conclusions.

Fortunately, Tulving and Psotka's claim that retroactive inhibition affects retrieval and not storage is a straightforward prediction that can easily be tested using Batchelder and Riefer's storage–retrieval model. We will begin with a description of this model, followed by an application of the model to a new experiment on retroactive inhibition.

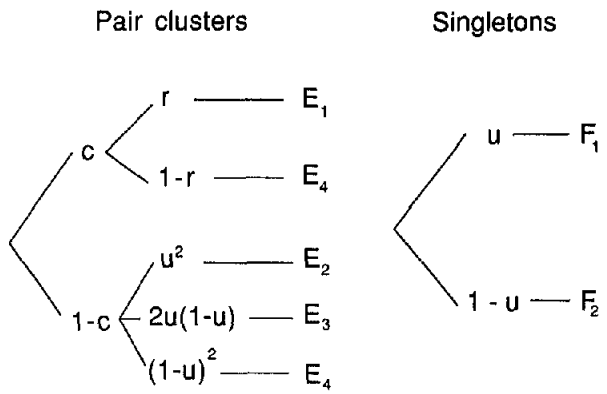

Figure 2. Tree diagram for Batchelder and Riefer's (1986) multinomial model for storage and retrieval.

Model Development

The experimental paradigm involves a free-recall task in which subjects are presented with a list of words that are related by categories (see Murphy & Puff, 1982). In particular, the words consist of several category pairs (e.g., oxygen and hydrogen, doctor and lawyer), plus a number of singleton words. Subjects are presented with this list of words, one word at a time, and are then required to recall the words in any order.

The data consist of the recall events for both the pair clusters and the singletons. Each category pair is scored in one of four mutually exclusive categories:

E_1—both words recalled, adjacently

E_2—both words recalled, nonadjacently

E_3—only one word in the pair recalled

E_4—neither word in the pair recalled.

The recall of the singletons is scored into two categories:

F_1—recalled

F_2—not recalled.

Let N_i be the frequency of occurrence for recall event E_i, and let M_i be the frequency of recall for event F_i. Because there are two distinct sets of data, the model developed next will be a joint multinomial model.

The following multinomial model describes the storage and retrieval of the category pairs as a function of three hypothetical processes.

Storage of clusters. After study, each pair of items either is or is not stored as a cluster. Define c to be the probability that a pair is stored as a cluster, $0 \leq c \leq 1$.

Retrieval of clusters. If a pair is stored as a cluster, it either is or is not retrieved as a cluster during recall. Define r to be the conditional probability that a pair is recalled as a cluster, given that it is stored as a cluster, $0 \leq r \leq 1$. It is assumed that retrieval of a cluster results in adjacent recall of both category members (E_1 event) and that nonretrieved clusters lead to the nonrecall of both items (E_4 event).

Retrieval of nonclustered items. If a pair is not stored as a cluster (with probability $(1 - c)$), then it is assumed that each item either is or is not stored and retrieved independently. Define u to be the probability that a nonclustered item is recalled, $0 \leq u \leq 1$. It is assumed that if both nonclustered items are recalled, they are recalled nonadjacently (an E_2 event). In addition to the nonclustered items, singleton words are also assumed to be stored and retrieved independently. For the version of the model here, assume that singletons are recalled with probability u (F_1 event) and not recalled with probability $1 - u$ (F_2 event).

As with the two previous models, the parameters of the storage–retrieval model are simply the probabilities of dichotomous processes. Thus, it is possible to represent the model in the form of a tree diagram, which is shown in Figure 2. Because this is a joint multinomial model, two separate tree structures are presented, one for the pair clusters and one for the unique stimuli.

It should be easy to see from the tree diagram that the probabilities for the E_i and F_i events are as follows:

$$P(E_1) = cr;$$

$$P(E_2) = (1 - c)u^2;$$

$$P(E_3) = (1 - c)2u(1 - u);$$

$$P(E_4) = c(1 - r) + (1 - c)(1 - u)^2;$$

$$P(F_1) = u; \text{ and}$$

$$P(F_2) = (1 - u).$$

In summary, the storage–retrieval model consists of the data vector $\mathbf{D} = (N_1, N_2, N_3, N_4; M_1, M_2)$ and parameter space $\Omega = \{\theta = (c, r, u) | 0 \leq c, r, u \leq 1\}$.

The likelihood function for the model is

$$L = \frac{N!}{N_1!N_2!N_3!N_4!}[cr]^{N_1} \times [(1 - c)u^2]^{N_2}[(1 - c)2u(1 - u)]^{N_3}$$

$$\times [c(1-r)+(1-c)(1-u)^2]^{N_4} \times \frac{M!}{M_1!M_2!}u^{M_1}(1-u)^{M_2}. \quad (38)$$

Batchelder and Riefer (1986) showed that if certain restrictions are satisfied by the observed frequencies, then closed-form solutions can be found for the parameter estimators that maximize Equation 38. These estimators are

$$\hat{u} = \frac{(2N_2+N_3+2M+M_1) - \sqrt{(2N_2+N_3+2M+M_1)^2 - 8M(N_2+M_1)}}{2M} \quad (39a)$$

$$\hat{c} = \frac{N\hat{u}(2-\hat{u})-(N_2+N_3)}{N\hat{u}(2-\hat{u})} \quad (39b)$$

$$\hat{r} = \frac{N_1\hat{u}(2-\hat{u})}{N\hat{u}(2-\hat{u})-(N_2+N_3)}. \quad (39c)$$

Table 3
Recall Frequencies for the Retroactive-Inhibition Experiment

No. of interpolated lists	N_1	N_2	N_3	N_4	M_1	M_2	P(C)	P(Cat)	*IPC*
0	97	5	9	39	38	37	.71	.74	.96
1	71	2	6	71	24	51	.51	.53	.96
2	55	3	10	82	25	50	.42	.45	.93
3	51	2	9	88	20	55	.38	.41	.93
4	54	2	9	85	22	53	.40	.43	.93

Note. Each data set is based on $N = 150$ total observations (15 subjects per condition, 10 categories per subject). N_1 = number of times a pair is recalled adjacently; N_2 = number of times a pair is recalled nonadjacently; N_3 = number of times only one word in a pair is recalled; N_4 = number of times neither word in a pair is recalled; M_1 = recall of a singleton; M_2 = nonrecall of a singleton; P(C) = proportion of items recalled; P(Cat) = proportion of categories represented in recall; *IPC* = items per category.

Experiment 1

The following experiment is basically a replication of Tulving and Psotka's (1971) study. One major difference between their experiment and this new one is that categories in this experiment consisted of two items, as compared with four items per category used by Tulving and Psotka. It was necessary, of course, to use category pairs in this study in order to provide data in the proper format for the multinomial model.

If retrieval-based theories of interference are correct, and if the model performs as expected, then retroactive inhibition should have a large effect on the retrieval parameter r. More specifically, the value of r should decrease as the number of interfering lists increases, without a corresponding decrease in the value of c.

Method. Seventy-five undergraduates from the University of California at Irvine participated in the experiment for class credit. Each subject was tested individually and was presented with either one, two, three, four, or five successive lists of words. These words were shown on a computer screen, one word at a time, at an exposure rate of 5 s per word. Each list contained 25 words, consisting of 10 categories (with 2 high-associate words per category) and five singletons. Most of the categories were taken from Battig and Montague (1969), although some categories not found in Battig and Montague were also included in the lists. Selection of categories and singletons appearing in each list was determined randomly. Presentation order for the words was also random, under the constraint that items from the same category were presented with no intervening items between them. Subjects were given 1½ min to recall in writing the 25 words from each individual list. After all of the lists had been presented, a final free-recall test was given in which subjects attempted to recall the words from all of the previous lists. Subjects were given up to 5 min for this final written recall.

Results. As in the Tulving and Psotka (1971) study, the focus here is on the recall of the first-list words during the final recall task. But instead of concentrating directly on empirical data as Tulving and Psotka did, our analysis will be based on the storage–retrieval model. The basic recall statistics from the experiment are presented in Table 3, which shows the recall of the first-list words as a function of the number of interpolated lists. In addition to presenting the N_i and M_i data frequencies, Table 3 also includes the three empirical data statistics examined by Tulving and Psotka: (a) P(C), the proportion of items correctly recalled; (b) P(Cat), the proportion of categories represented in recall; and (c) *IPC*, the proportion of items recalled per category. Only the pair-cluster data and not the singletons were used to compute these statistics. As can be seen, the basic empirical results of Tulving and Psotka's study are replicted here. P(C) and P(Cat) both decrease as the number of interpolated lists increases, whereas *IPC* remains relatively constant.

As mentioned before, it was this pattern of results that led Tulving and Psotka (1971) to conclude that retroactive inhibition affects the retrieval of items but not their storage. Their conclusion, however, was not based on any substantive model. The multinomial model for storage and retrieval now allows us to conduct a more formal analysis of these data. The values of N_i from Table 3 were inserted into Equation 39 to derive estimates of the storage and retrieval parameters for each condition, and these estimates are plotted in Figure 3. The results show that retroactive inhibition has a strong effect on the retrievability of items, because of the decreasing value of parameter r. But the storage of items, represented by parameter c, is basically unaffected by the number of interpolated lists.

Goodness of Fit

Once the parameters of the model have been estimated, the next step is to determine if the model adequately fits the data. As described in the previous section, this involves a likelihood ratio test in which the value of the likelihood function with the estimated parameters is compared with the value of the function under the general multinomial model. The storage–retrieval model has three parameters and describes four independent data events, so a goodness-of-fit test with one degree of freedom can be conducted. Specifically, Equations 27 and 28 can be used to compute the value of G^2 for each of the five data sets in the experiment. These values are presented in Table 4. As can be seen, the fit of the model is excellent, with none of the data sets leading to rejection of the model. In fact, it is interesting to note that all of the values are extremely small, even

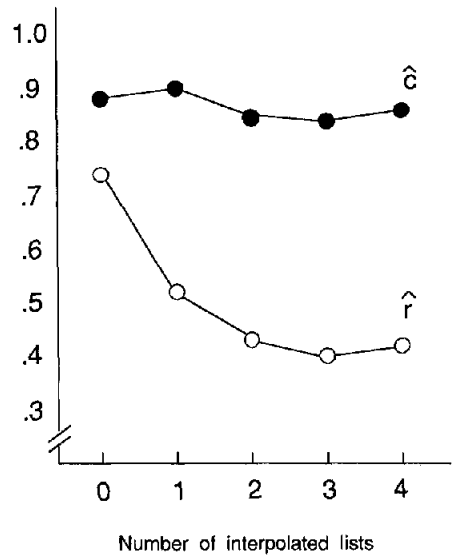

Figure 3. Estimates of the parameters c and r from the storage–retrieval model, plotted as a function of the number of interpolated lists. (C = correct response; CE = confusion error; and NCE = nonconfusion error.)

less than what might be expected by chance. It is possible that some aspect of the experiment (perhaps the fact that category pairs were presented adjacently) may be in part responsible for this because goodness-of-fit values as small as these are not typical of those found in other applications of the model (cf. Batchelder & Riefer, 1980, 1986).

Hypothesis Testing

To complete the analysis of this experiment, it is necessary to determine if the differences in the parameter values across experimental conditions are statistically significant. This involves using the likelihood ratio test to evaluate two null hypotheses: (a) c_j is a constant across J, and (b) r_j is a constant across J. These tests proceed similarly to the ones conducted for the Greeno et al. (1978) model, and a detailed discussion of them can be found in Batchelder and Riefer (1986).

The results of the likelihood ratio test show that the decreasing value of parameter r is a statistically reliable effect, $G^2(4) = 35.81$, $p < .01$. In contrast, the values of parameter c do not significantly differ across conditions, $G^2(4) = 1.43$, *ns*.

Conclusions

The results of the multinomial model support Tulving and Psotka's (1971) conclusion that retroactive inhibition is essentially a retrieval phenomenon. Moreover, the fact that the storage parameter c was unaffected by the number of interfering lists would seem to be inconsistent with theories that imply that retroactive inhibition results in a weakening or a loss of memory traces. Instead, these results are in accordance with theories of interference that postulate retrieval explanations for retroactive inhibition, such as response set suppression theory (Postman & Stark, 1969).

This example also illustrates an important point concerning multinomial modeling. Earlier, we argued that multinomial models have certain advantages over methods of data analysis based solely on ad hoc statistics like *IPC* or $P(\text{Cat})$. This advantage can now be demonstrated very clearly using our earlier example. If the storage–retrieval model is true, then it is possible to derive an expression for IPC over the experimental conditions as a function of the model's parameters. This expression is

$$\text{IPC}_j = \frac{c_j(r_j - u_j) + u_j}{c_j[r_j - u_j(2 - u_j)] + u_j(2 - u_j)}, \qquad (40)$$

where j is an index for each experimental condition. In an actual experiment, if the value of c_j is a constant across j, it is possible for the values of IPC_j also to be constant. According to Tulving and Psotka (1971), this pattern of results would indicate that the storage of items is unaffected by the experimental manipulation.

But the actual values of IPC_j in Equation 40 need not be a constant value over j. For example, suppose c_j is a constant and u_j and r_j have the simple linear relation $u_j = kr_j$ (where k is a constant). Then it can be shown that IPC_j is a decreasing function of j. It is therefore possible for IPC to decrease across experimental conditions even when the storage parameter is totally unchanged, which might lead someone not using the model to the wrong general conclusions about storage. This empirical result has in fact been obtained by other researchers (Nelson & Brooks, 1974; Sowder, 1976), who concluded that retroactive inhibition does result in a certain amount of storage loss. The results of the analysis conducted here suggest that this conclusion may be incorrect.

Discrimination Learning

In the previous two examples, we demonstrated how multinomial modeling can be used to explore and answer questions about cognitive processes. In addition, the examples managed to shed light on some unresolved theoretical issues concerning interference theory. However, both examples involved well-established models that have appeared previously in the psychological literature. For this third example, we would like to introduce a new multinomial model that we have developed specifically for this article. The model has been designed to analyze data from discrimination-learning experiments and is capable of separating memory for discriminable stimuli into learning

Table 4
Goodness-of-Fit Statistics for the Retroactive-Inhibition Experiment

No. of interpolated lists	$\chi^2(1)$
0	0.02
1	0.16
2	0.07
3	0.06
4	0.01

Note. $df = 1$ for each test.

and discrimination parameters. Because this is a new model, our main goal here is not to explore any general theoretical issues, as we did in the previous examples. Instead, we will test the validity and usefulness of the model using a standard discrimination-learning experiment. The model will also enable us to illustrate some of the measurement advantages that multinomial modeling has over traditional methods of data analysis.

Theoretical Background

In a task involving discrimination learning, subjects are presented with similar, potentially confusable stimuli. These stimuli are paired with different responses (e.g., "new" or "old" in a recognition paradigm), and therefore subjects must learn to discriminate between the similar stimuli in order to make the proper responses. This task is analogous in some respects to signal detection (Green & Swets, 1966), in which subjects must discriminate between signal and noise (as opposed to discriminating between multiple stimuli). In the field of human memory, models of signal detection have often been applied to recognition paradigms. This is an area in which multinomial methods of measurement have been used quite successfully to supplement traditional statistical analyses of data. By separating responses into false alarms, correct hits, and so forth, signal detection models can be used to derive parameter estimates for the detectability of stimuli. By now all researchers would probably agree that these parameters reveal more about the underlying cognitive processes than the empirical statistics alone, and for this reason signal detection models have become quite popular as measurement tools in recognition memory.

However, we know of no comparable multinomial analysis that exists for measuring the underlying factors in discrimination learning. It is true, of course, that a number of Markov models have been proposed that can be used to measure these processes separately. For example, Polson, Restle, and Polson (1965) and Brainerd, Howe, and Desrochers (1980) have proposed such models, and Atkinson and Estes's (1963) mixed model could also be applied to this situation. However, because Markov models postulate both cognitive states and the transitions between them, they make stronger theoretical assumptions than multinomial models do. Because they are stronger models, they are more likely to fail to account for the data and thus may not be totally satisfactory as measurement tools.

Instead of modeling approaches to measurement, data analysis in the area of human discrimination learning has traditionally involved the use of ad hoc statistics such as probability correct, $P(C)$, or mean number of errors. But as we have stated previously, global measures such as these also may not be totally satisfactory, because they may fail to reflect the individual contributions of different cognitive factors. Relevant to this idea, Yang (1985) has pointed out that research on the interfering effects of similar stimuli is usually totally confounded by learning and similarity factors. Thus, there is a need for some type of simple multinomial approach in discrimination learning that is capable of separately measuring these processes.

Model Development

Although there exist many paradigms for studying discrimination learning (see Houston, 1981, for a review), human discrimination learning is often studied within the framework of paired-associate learning (e.g., Craig, Humphreys, Rocklin, & Revelle, 1979; Kitahama, 1982; Runquist & Runquist, 1978). In this standard experiment, discriminability is manipulated by having the stimulus terms of the paired associates related along some dimension, and this is the basic paradigm to be used for the model presented here. Specifically, the model is based on a theory of paired-associate learning that states that paired associates are learned in three distinct stages: (a) response learning, (b) stimulus encoding, and (c) stimulus–response learning (see Houston, 1981, chapter 5). In the research to follow, the response-learning component is minimized because all responses are familiar words that have previously been memorized by the subjects. For the other two stages, the model attempts to separately measure the stimulus encoding and stimulus–response learning components of paired-associate learning. Presumably, the process of discriminating between stimuli occurs during the stimulus encoding stage. Restle (1964) and Polson et al. (1965) have theorized that confusable stimuli will require more stimulus encoding in order to be sufficiently stored in memory, and it is this process that the multinomial model attempts to capture.

The experimental paradigm used here for discrimination learning is similar in certain respects to the one described earlier for proactive inhibition. Subjects are presented with a list of paired associates. Within this list, some of the associates have unique stimulus terms, in the sense that the stimuli are distinct from one another. Another subset of the associates consists of pairs of similar stimuli coupled with different responses, which can be represented as A–B, A′–C. For example, if the paired associates consist of word–digit pairs, the critical stimuli might be synonyms paired with different responses (e.g., taxi–2 and cab–7). In the experiment analyzed here, an anticipation procedure is used in which subjects attempt to give the response to each paired associate before the correct stimulus–response pairing is revealed. Over repeated trials, subjects learn to discriminate between the related stimuli in order to make the correct response to each.

In terms of data representation, we will consider separately the recall of the unique items versus the paired, similar items. For the unique items, it is natural to define two recall events:

F_1—recalled (R) and

F_2—not recalled ($\bar{R}$).

When one of the paired items is tested for memory, three distinct recall events are possible:

E_1—correct response (C),

E_2—confusion error (CE), and

E_3—nonconfusion error (NCE).

These three events can be illustrated using a concrete example. Suppose the two related paired associates are taxi–2 and cab–7. If taxi–? is presented for recall, the correct response (E_1) would be 2, a confusion error (E_2) would be 7, and a nonconfusion error (E_3) would be any other incorrect response. This separation of errors into confusion and nonconfusion errors is one of the key features of the model.

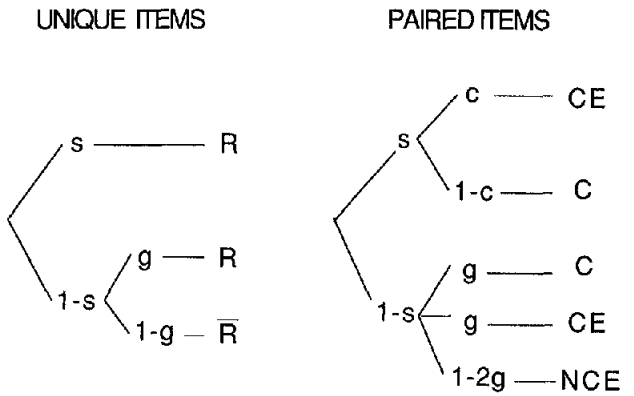

Figure 4. Tree diagram for the multinomial model for discrimination learning.

For notational purposes, let M_i denote the frequency of occurrence for event F_i, and let N_i denote the frequency of E_i. Also, let

$$M = \sum_{i=1}^{2} M_i \quad \text{and} \quad N = \sum_{i=1}^{3} N_i.$$

The experimental design just described generates two distant sets of data, one for the unique items and one for the paired items. The following model is therefore a joint multinomial model of the type described in the previous section.

The model attempts to describe the learning and recall of related paired associates as a consequence of three hypothetical events: stimulus–response learning, stimulus confusion, and guessing.

Stimulus–response learning. Learning occurs when an adequate representation of the paired associate is stored in memory, that is, with some response associated to a representation of the stimulus. For any given test trial, a paired associate either has or has not been learned for that trial. Define s to be the probability that it has been learned, $0 \leq s \leq 1$.

Stimulus confusion. If a paired associate is successfully stored in memory, it is possible that the representation of the stimulus term is confused with a related stimulus. This would occur, for example, when two similar stimuli have not been adequately discriminated. On any given test trial, a learned stimulus either is or is not confused with a related stimulus. Define c to be the probability that this occurs, $0 \leq c \leq 1$. In this event, a confusion error will be given as a response. It is assumed that unique stimuli have no related stimuli to be confused with and, hence, are confused with other stimuli with probability zero.

Guessing. If a paired associate remains unlearned on a given test trial, then the subject guesses from the set of possible responses with an equal probability of guessing each response alternative. Define g to be this guessing probability, $0 \leq g \leq 1$.

As with the model for proactive inhibition, the parameters of this model represent the probabilities of dichotomous events. Thus, a tree diagram can be used to represent the model, which is illustrated in Figure 4. Because this is a joint multinomial model, a separate tree structure has been drawn for the unique stimuli and the similar stimuli.

With the assumptions behind the model just outlined, it is possible to write expressions for the probabilities of the data events:

$$P(\mathrm{R}) = s + (1 - s)g \tag{41a}$$

$$P(\bar{\mathrm{R}}) = (1 - s)(1 - g) \tag{41b}$$

$$P(\mathrm{C}) = s(1 - c) + (1 - s)g \tag{41c}$$

$$P(\mathrm{CE}) = sc + (1 - s)g \tag{41d}$$

$$P(\mathrm{NCE}) = (1 - s)(1 - 2g). \tag{41e}$$

In summary, this model consists of the data vector $\mathbf{D} = (M_1, M_2; N_1, N_2, N_3)$ and parameter space $\Omega = \{\theta = (s, c, g) | 0 \leq s, c, g, \leq 1\}$.

The likelihood function for the model is

$$L = \left(\frac{M!}{M_1!M_2!}\right)[s + (1 - s)g]^{M_1}[(1 - s)(1 - g)]^{M_2} \times \left(\frac{N!}{N_1!N_2!N_3!}\right)[s(1 - c) + (1 - s)g]^{N_1} \cdot \times [sc + (1 - s)g]^{N_2}[(1 - s)(1 - 2g)]^{N_3}. \tag{42}$$

By using the calculus methods in Equation 18, closed-form solutions can be derived for the parameter values that maximize this function. These MLEs are

$$\hat{s} = \frac{N(M_1 - M_2) + MN_3}{\mathrm{NM}} \tag{43a}$$

$$\hat{c} = \frac{M_1N - MN_1}{N(M_1 - M_2) + MN_3} \tag{43b}$$

$$\hat{g} = \frac{NM_2 - MN_3}{2NM_2 - MN_3}. \tag{43c}$$

Equation 43 can thus be used to compute estimates of the model's parameters. But in addition to determining parameter estimates, it is also desirable to be able to test the goodness of fit for the model. Unfortunately, the model for discrimination learning contains three parameters and attempts to describe three independent recall events. Thus, like Greeno et al.'s (1978) model for proactive inhibition, there are no degrees of freedom with which to test the model. Because of this, it would be desirable to develop some restriction that would free up at least one degree of freedom for testing if the model adequately fits the data.

One possibility along these lines is to note that the guessing parameter g should depend on the number of response alternatives in the experiment. Specifically, if there are n such response alternatives, then a reasonable restriction on the model is to set g equal to $1/n$. If we assume that $g = 1/n$ in the likelihood function in Equation 42, then new estimates for s and c can be computed that will maximize this function. In this manner, parameter estimates for the model can be found that leave one degree of freedom (the one gained by estimating $g = 1/n$) for testing the model.

Unfortunately, we were unable to derive closed-form solu-

Table 5
Recall Frequencies for the Discrimination-Learning Experiment

	A–B items					A′–C items				
Trial	R	$\bar{R}$	C	CE	NCE	R	$\bar{R}$	C	CE	NCE
1	82	119	164	74	164	62	139	95	122	185
2	118	83	201	68	133	93	108	152	89	161
3	136	65	242	58	102	126	75	219	67	116
4	156	45	288	46	68	144	57	256	58	88
5	167	34	319	31	52	167	34	309	41	52
6	175	26	324	18	60	175	26	328	23	51
7	181	20	343	26	33	180	21	335	28	39
8	181	20	351	15	36	181	20	345	21	36
9	191	10	365	14	23	182	19	365	16	21

Note. R = recalled; $\bar{R}$ = not recalled; C = correct response; CE = confusion error; NCE = nonconfusion error.

tions for the parameters using this estimation procedure, but this is not a serious problem. When closed-form solutions cannot be found, computer search methods discussed earlier can be used to find the parameter values that maximize the likelihood function. This is the method that is used to compute the parameter estimates in the example that follows.

Experiment 2

To test the validity and usefulness of this multinomial model, we will analyze the data from a multitrial experiment on discrimination learning. This experiment is similar in design to one conducted by Batchelder (1971), except that the paired stimuli in Batchelder's experiment had the same responses (A–B, A′–B) as opposed to different responses in the current experiment (A–B, A′–C).

Method. Subjects consisted of 67 student volunteers from introductory psychology courses. Each subject was presented with a list containing 18 paired associates. The stimulus terms consisted of Gibson-form stimuli (Gibson, 1941), and the set of possible responses consisted of six previously memorized English words. The paired associates were presented one at a time using a slide projector. The first 9 paired associates in the list were six A–B items and three unique items. The second half of the list contained six A′–C items (paired associates with stimulus terms similar to ones in the first half of the list) and three more unique paired associates. Presentation was random within each half of the list. Recall was by an anticipation procedure in which subjects were presented with the stimulus term, given up to 6 s to make a push-button response, and then given a 2-s presentation of the correct stimulus–response pair. After the first presentation of the list (during which the subjects' anticipation responses were guesses), the list was presented for nine trials.

Results. The recall frequencies for each of the nine anticipation trials are presented in Table 5. Using subroutine ZXMIN from IMSL (1982), the values of s and c were found that maximize the likelihood function in Equation 42, under the constraint that $g = 1/6$. Figure 5 presents these MLEs for each of the nine trials in the experiment, plotted separately for the A–B and the A′–C items. As can be seen from the figure, the confusability of the A′–C items (as measured by c) is high for the early trials. In contrast, confusability is not as high for the A–B items, which have the advantage of being presented first. Moreover, confusability of both stimuli decreases across study trials. This, of course, represents the discrimination component of discrimination learning.

Looking at the storage parameter s, one can see that the confusability of the stimuli also results in a storage deficit for the A′–C items (as compared with A–B). This difference in storage disappears, however, as both items are learned over trials, and it is completely gone by Trial 5. Also, the storage parameter for both items monotonically increases across trials. This, of course, represents the learning component of discrimination learning.

In addition to the point estimates for the parameters already given, confidence regions can also be derived for the parameters. These regions take the form of an ellipse given by the formula

$$\left(\frac{s-\mu_s}{\sigma_s}\right)^2 - 2\rho\left(\frac{s-\mu_s}{\sigma_s}\right)\left(\frac{c-\mu_c}{\sigma_c}\right) + \left(\frac{c-\mu_c}{\sigma_c}\right)^2 = 2(1-\rho^2)\ln(1-\alpha). \quad (44)$$

The above ellipse gives the $100\alpha\%$ confidence region for the parameters s and c. In Equation 44, μ_s and μ_c are the values of the parameters that maximize the likelihood function. (We are again assuming that $g = 1/6$.) The values of σ_s, σ_c, and ρ are obtained by computing the estimated information matrix from Equation 21 and then inverting this matrix to obtain the variance–covariance matrix for s and c.

Because the differences between the parameters are greatest for the first trial, we have computed the confidence ellipses for the first-trial data for both the A–B and A′–C items. These confidence ellipses are presented in Figure 6. As can be seen, the

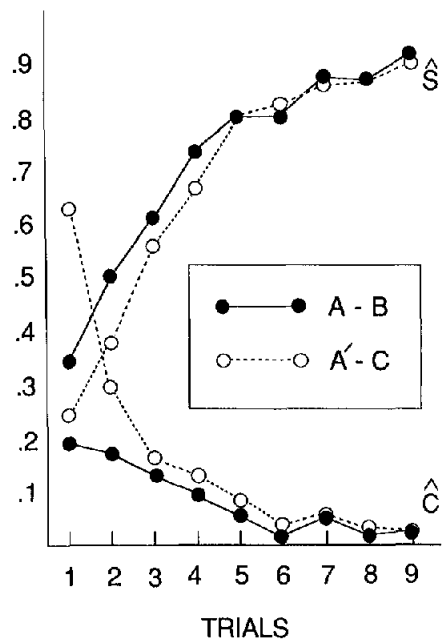

Figure 5. Estimates of the parameters s and c from the discrimination-learning model. The estimates are plotted as a function of trials separately for the A–B and A′–C items.

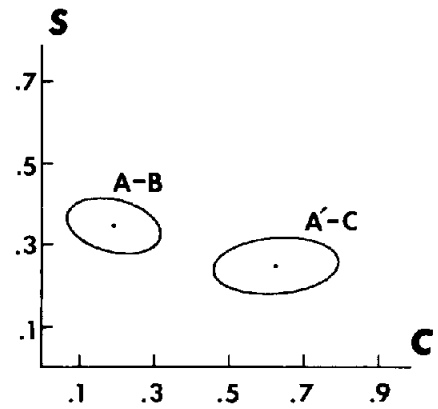

Figure 6. Confidence ellipses for the values of parameters c and s from the discrimination-learning model, plotted separately for the A–B and A'–C items.

major difference between the two stimuli is in their confusability, reflected by the displacement of the ellipses along the c dimension. They differ to a lesser degree in their learnability, as evidenced by a smaller shift in the ellipses along the s dimension.

Goodness of Fit

Once the parameters of the model have been estimated, the likelihood ratio test can be used to determine whether the model adequately fits the data. To conduct this test for the current model, the parameter estimates in Figure 5 were inserted into Equation 42 to determine the value of the likelihood function for the substantive model. Next, the probabilities of the empirical data events were inserted into Equation 1 to determine the value of the function for the general model. These values were then placed into Equation 28 to compute the value of G^2 for this test. In this case, G^2 is an approximate chi-square value with one degree of freedom.

The data in Table 5 were used to compute this goodness-of-fit statistic for each condition in the experiment. The results are presented in Table 6. In general, the fit of the model is quite satisfactory. Only 1 of the 18 tests exceeds the limits for an acceptable fit, a result not much different from what would be expected by chance.

Conclusions

The multinomial model for discrimination learning appears to do a satisfactory job of separating the recall of similar stimuli into storage and discriminability factors. The parameter estimates from the discrimination-learning experiment are easy to interpret, and the model does a good job of fitting the data. It is our opinion that this type of model has a large potential for supplementing research in the area of stimulus discrimination and recall. To cite a few examples, the model could be used to explore such research issues as the interference effects of similar stimuli (Kitahama, 1982; Nelson, Brooks, & Wheeler, 1975; Runquist & Runquist, 1978), developmental and personality differences (Carey, Diamond, & Woods, 1980; Craig et al., 1979), and comparisons of different types of stimuli (Skaalvik, 1977).

As we stated earlier, there are other methods, principally Markov models, for separately measuring cognitive factors in discrimination learning. In fact, the multinomial model presented here is related somewhat to a more complicated Markov model that was tested by Polson et al. (1965). It could be argued that from a theoretical standpoint, the stronger Markov is preferable to the multinomial model if both equally account for the data. But if one's goal is to statistically measure and experimentally study cognitive processes, then the multinomial model has the advantages of statistical simplicity and wider applicability. In our personal research, we have tried to fit Markov models like Polson et al.'s to the data analyzed here without much success, unless the models are made increasingly more complicated. For example, good fits can be obtained if the transition probabilities between states are allowed to change over trials (cf. Bush & Mosteller's, 1955, and Robbins's, 1970, "shrinkage" models). For basic theoretical purposes this may be the proper approach even if it does lead to increasingly complex models. But as methods for analyzing data, such models become unwieldy and difficult to use. If one instead wants to study the separate effects of an experimental variable on stimulus discrimination and associative learning, then the simpler multinomial model has many advantages over more complete and complex theories incorporating these processes.

Another advantage of multinomial models that we have attempted to emphasize is their improvement over traditional, general purpose statistical techniques such as ANOVAs. This certainly applies to the area of discrimination learning, because stimulus discrimination and recall may involve several separate cognitive processes, and empirical statistics only reflect the combined result of these processes. Because of this, using ANOVAs on data of this type may produce results that are hard to interpret or, as Batchelder and Riefer (1986) pointed out, totally misleading.

In fact, the limitations of using ANOVAs on empirical statistics can now be illustrated very convincingly with the following demonstration. In order to present this demonstration more clearly, we will focus on a current research area—memory and recognition of unfamiliar faces. This is a task that involves

Table 6
Goodness-of-Fit Statistics for the Discrimination-Learning Model

Trial	A–B items	A'–C items
1	3.10	6.44*
2	0.0002	0.62
3	0.02	0.08
4	0.11	0.06
5	0.05	0.05
6	2.98	0.84
7	0.01	0.34
8	0.20	0.20
9	1.02	1.45

Note. $df = 1$ for each test.
*$p < .01$.

learning and discrimination processes, because people must first store a representation of a face in memory and then discriminate it from a set of distractors in order to recognize it correctly. A well-established finding in this area is that recognition of faces is much more difficult when they are viewed upside down than when they are viewed upright (Goldstein, 1965; Rock, 1974). Moreover, a change in orientation seems to hurt facial recognition more than it hurts the recognition of other types of stimuli (Diamond & Carey, 1986; Yin, 1969). This Orientation × Stimulus Type interaction has led many researchers to conclude that the encoding of faces involves "special" processes, unique to facial stimuli, that are disrupted when faces are inverted. However, exactly what these processes are and what generally causes the inversion effect remain open questions.

For the purposes of this demonstration, we will now make some hypothetical assumptions about the cognitive processes involved in memory for faces and other stimuli. First, suppose that stimulus inversion affects only the discriminability of stimuli, and not their storage in memory. Assuming that the multinomial model is true, this would mean that upright versus inverted stimuli have different discriminability parameters (c_1 and c_2), but the same storage parameter s. For the sake of argument, also assume that facial stimuli are not special (i.e., that the memory advantage that faces have over other stimuli is merely due to general encoding processes and nothing more). This would imply that the memory differences between faces and nonfaces lie in their memorability (parameters s_1 vs. s_2) and not in their discriminability.

Thus, we have created a hypothetical scenario in which one variable (orientation) affects only discrimination, and another variable (stimulus type) affects only storage. One way of examining this would be to run a paired-associate experiment that manipulates these variables in a factorial design and then to measure the effect of these variables on some commonly used statistic such as the proportion of correct responses, $P(C)$. If such an experiment were run, one might expect to see significant main effects for each variable but no interaction, because it has been assumed that these variables affect storage and discriminability independently. However, we now show that strong empirical interactions are possible under just these circumstances. To illustrate this for the current example, Figure 7 presents the values of $P(C)$ (from Equation 41c) computed for a hypothetical paired-associate experiment where c_1 (upright) = .3, c_2 (inverted) = .9, s_1 (faces) = .8, s_2 (nonfaces) = .2, and g = .2. As can be seen, using $P(C)$ as the dependent measure results in a strong interaction between these two variables, even though in the example the underlying cognitive processes of storage and discrimination are assumed to be completely independent. Moreover, because it is a crossover interaction, no monotonic transformation of the dependent measure will eliminate it.

Of course, the previous example has been specifically constructed to make a particular statistical point, and we are not advocating the aforementioned ideas as a new theory for the memory of faces. However, we feel that the above demonstration effectively illustrates our main point, that the common practice of applying ANOVAs to single empirical statistics may not provide an accurate picture of the underlying cognitive processes involved. Interactions or nonmonotonic trends are often found between variables in discrimination tasks (e.g., MacLeod & Nelson, 1976). As long as the factorial analysis of empirical measures constitutes the principal methodology for studying stimulus discrimination, the true properties of the cognitive processes involved may remain elusive. The multinomial model developed in this article can potentially provide a clearer picture of these processes, not just for the research questions outlined but for many others as well.

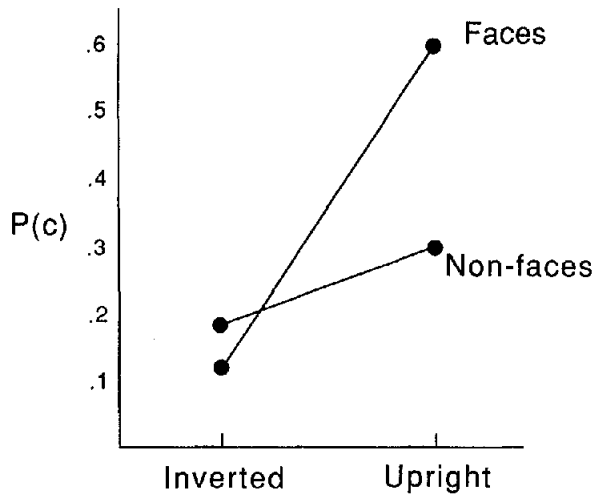

Figure 7. Results of a hypothetical experiment on the recall of inverted facial and nonfacial stimuli, in which orientation and type of stimulus are assumed to affect discrimination and storage independently. (P(c) = probability correct.)

Summary

In this article we have attempted to evaluate a statistical methodology, called multinomial modeling, that we feel can be used very effectively for certain types of research in cognitive psychology. By making a few situation-specific assumptions, multinomial models are capable of measuring the types of psychological processes that cognitive psychologists frequently study. In addition, hypothesis-testing procedures can be used to determine which experimental variables significantly influence these processes. In this way, multinomial models can be a useful supplement to more traditional methods of data analysis. In the second section, we pointed out that many explanations for cognitive phenomena postulate that discrete behavior derives from finitely many underlying cognitive states. We have seen that multinomial modeling can provide a useful statistical methodology for assessing such explanations without the necessity of constructing complete processing theories.

The three empirical examples presented in this article illustrate a number of important aspects of multinomial modeling. For example, the model for proactive inhibition demonstrates how easily theoretical ideas can be expressed as a multinomial model and how data analysis based on such a model can reveal additional insights into cognitive processing. The theoretical ideas behind the proactive-inhibition model were originally formulated by Greeno et al. (1978); however, they did not statisti-

cally analyze these ideas to their fullest extent. Figure 1 illustrates how Greeno et al.'s theoretical assumptions can be represented in a tree diagram that leads directly to a simple multinomial model. Using the techniques outlined in the article, it is easy to obtain point estimates and confidence intervals for the model's parameters. Tests of hypotheses relating experimental variations to changes in the underlying parameters can also be easily conducted within the multinomial framework. These procedures provide a substantial amount of information that would not be available without the analysis of the multinomial model.

Another important aspect of multinomial models is their ability to disentangle opposing cognitive processes. The storage–retrieval model, for example, is capable of measuring the separate contributions of storage and retrieval to recall. Batchelder and Riefer (1980, 1986) have shown how these processes can combine, sometimes in opposing ways, leading to empirical recall data that may not reflect the individual effect of either process. For example, the analysis of retroactive inhibition in Tulving and Psotka's (1971) design demonstrated how an experimental variable can independently affect one process but not the other. The data directly revealed that overall recall decreased as a function of retroactive inhibition, but it was the analysis of the multinomial model that provided a rigorous argument that the decrease was due to retrieval and not to storage factors.

In conclusion, it should be emphasized that even though multinomial models have a certain theoretical motivation, our intention in this article is not to develop yet another class of strong theoretical models. Instead, our goal has been to evaluate a class of models that can express substantive theoretical ideas as statistical models and that can be used to supplement more general-purpose models such as ANOVA. By their very design, these models are more situation specific and less of an approximation to the true operation of cognitive systems than strong theoretical models. Nevertheless, multinomial models have enhanced scientific understanding in other fields, such as statistical genetics. It is our hope that multinomial modeling will be used more extensively in cognitive psychology to fill the gap between more empirical data-processing approaches and strong, theoretically motivated models.

References

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

Atkinson, R. C., Bower, G. H., & Crothers, E. J. (1965). *An introduction to mathematical learning theory.* New York: Wiley.

Atkinson, R. C., & Estes, W. K. (1963). Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 121–268). New York: Wiley.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Bamber, D., & van Santen, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology, 29,* 443–473.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97–105.

Batchelder, W. H. (1971). A theoretical and empirical comparison of the all-or-none multilevel theory and the mixed model. *Journal of Mathematical Psychology, 8,* 82–107.

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. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology, 39,* 129–149.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80*(3, Pt. 2).

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge: MIT Press.

Brainerd, C. J., Howe, M. L., & Desrochers, A. (1980). Interpreting associative-learning stages. *Journal of Experimental Psychology: Human Learning & Memory, 6,* 754–765.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models of learning.* New York: Wiley.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition—A maturational component? *Developmental Psychology, 16,* 257–269.

Chechile, R., & Meyer, D. L. (1976). A Bayesian procedure for separately estimating storage and retrieval components of forgetting. *Journal of Mathematical Psychology, 13,* 269–295.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics.* London: Chapman and Hall.

Craig, M. J., Humphreys, M. S., Rocklin, T., & Revelle, W. (1979). Impulsivity, neuroticism, and caffeine: Do they have additive effects on arousal? *Journal of Research in Personality, 13,* 404–419.

DaPolito, F. J. (1966). *Proactive effects with independent retrieval of competing responses.* Unpublished doctoral dissertation, Indiana University.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Earhard, B. (1976). Retroactive inhibition: Associative responses or response suppression? *American Journal of Psychology, 89,* 385–401.

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

Elandt-Johnson, R. C. (1971). *Probability models and statistical methods in genetics.* New York: Wiley.

Feigenbaum, E. A. (1970). Information processing and memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 451–468). New York: Academic Press.

Gibson, E. J. (1941). Retroactive inhibition as a function of degree of generalization between tasks. *Journal of Experimental Psychology, 82,* 93–115.

Goldstein, A. G. (1965). Learning of inverted and normally oriented faces in children and adults. *Psychonomic Science, 3,* 447–448.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greeno, J. G., & Bjork, R. A. (1973). Mathematical learning theory and the new "mental forestry." *Annual Review of Psychology, 24,* 81–116.

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. (1972). On testing the independence of associations. *Psychological Review, 79,* 261–264.

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

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (3rd ed.). New York: Macmillan.

Houston, J. P. (1981). *Fundamentals of learning and memory* (2nd ed.). New York: Academic Press.

Humphreys, M. S., & Bain, J. D. (1983). Recognition memory: A cue and information analysis. *Memory & Cognition, 11,* 583–600.

International Mathematical and Statistical Libraries (IMSL) (1982). *IMSL library. Reference manual* (Vol. 3, 9th ed.). Houston, TX: Author.

Johnson, N. L., & Kotz, S. (1972). *Distributions in statistics: Continuous multivariate distributions.* New York: Wiley.

Kitahama, T. (1982). Effects of differentiating cues on interference by stimulus semantic similarity in paired associate learning. *Japanese Journal of Psychology, 53,* 232–235.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* New York: Elsevier.

Lehmann, E. L. (1983). *Theory of point estimation.* New York: Wiley.

Levine, G., & Burke, C. J. (1972). *Mathematical model techniques for learning theories.* New York: Academic Press.

MacLeod, C. M., & Nelson, T. O. (1976). A nonmonotonic lag function for false alarms to associates. *American Journal of Psychology, 89,* 127–135.

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

Martin, E. (1981). Simpson's paradox resolved: A reply to Hintzman. *Psychological Review, 88,* 372–374.

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology, 53,* 173–203.

Miller, R. R., Kasprow, W. J., & Schactman, T. R. (1986). Retrieval variability: Sources and consequences. *American Journal of Psychology, 99,* 145–218.

Minsky, M. L. (1967). *Computation: Finite and infinite machines.* Englewood Cliffs, NJ: Prentice-Hall.

Murphy, M. D., & Puff, C. R. (1982). Free recall: Basic methodology and analysis. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 99–128). New York: Academic Press.

Nelson, D. L., & Brooks, D. H. (1974). Retroactive inhibition of rhyme categories in free recall: Inaccessibility and unavailability of information. *Journal of Experimental Psychology, 102,* 277–283.

Nelson, D. L., Brooks, D. H., & Wheeler, J. W. (1975). Sensory and meaning features in stimulus recognition and associative retrieval. *Journal of Experimental Psychology: Human Learning & Memory, 1,* 711–719.

Newton, J. M., & Wickens, D. D. (1956). Retroactive inhibition as a function of the temporal position of the interpolated learning. *Journal of Experimental Psychology, 51,* 149–154.

Polson, M. C., Restle, F., & Polson, P. G. (1965). Association and discrimination in paired-associates learning. *Journal of Experimental Psychology, 69,* 47–55.

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology, 79,* 168–177.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

Pribram, K. H., Newer, M., & Baron, R. J. (1974). The holographic hypothesis of memory structure in brain function and perception. In D. H. Krantz, R. D. Luce, R. C. Atkinson, & P. Suppes (Eds.), *Contemporary development in mathematical psychology* (Vol. 2, pp. 416–457), San Francisco: Freeman.

Pylyshyn, Z. W. (1986). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge: MIT Press.

Restle, F. (1964). Sources of difficulty in learning paired-associates. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 116–172). Stanford, CA: Stanford University Press.

Reynolds, J. H. (1977). Unavailable and inaccessible information in retroactive inhibition of paired associates. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 68–77.

Riefer, D. M. (1982). The advantage of mathematical modeling over traditional methods in the analysis of category clustering. *Journal of Mathematical Psychology, 26,* 97–123.

Robbins, D. (1970). Stimulus selection in human discrimination learning and transfer. *Journal of Experimental Psychology, 84,* 282–290.

Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230,* 78–85.

Ross, B. H., & Bower, G. H. (1981). Comparisons of models of associative recall. *Memory & Cognition, 9,* 1–16.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing* (Vol. 1). Cambridge: MIT Press.

Runquist, W. N., & Runquist, P. A. (1978). Interference reduction with conceptually similar paired associates. *Journal of Experimental Psychology: Human Learning & Memory, 4,* 370–381.

Skaalvik, E. M. (1977). Pictures and words as stimuli in paired associate learning. *Scandinavian Journal of Education Research, 21,* 1–15.

Sowder, C. D. (1976). Category length and retroactive inhibition in cued recall. *Psychological Reports, 39,* 711–714.

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology, 87,* 1–8.

Underwood, B. J. (1983). *Attributes of memory.* Glenview, IL: Scott, Foresman.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50,* 1–26.

Wichawut, C., & Martin, E. (1971). Independence of A-B and A-C associations in retroaction. *Journal of Verbal Learning and Verbal Behavior, 10,* 316–321.

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

Yang, J. (1985). Retroactive inhibition without intentional new learning: Support for a stimulus-encoding theory. *Acta Psychologica, 58,* 63–74.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145.

Received November 4, 1986
Revision received August 26, 1987
Accepted September 23, 1987 ■