Theoretical and empirical review of multinomial process tree modeling

WILLIAM H. BATCHELDER
University of California, Irvine, California

and

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

We review a current and popular class of cognitive models called *multinomial processing tree* (MPT) models. MPT models are simple, substantively motivated statistical models that can be applied to categorical data. They are useful as data-analysis tools for measuring underlying or latent cognitive capacities and as simple models for representing and testing competing psychological theories. We formally describe the cognitive structure and parametric properties of the class of MPT models and provide an inferential statistical analysis for the entire class. Following this, we provide a comprehensive review of over 80 applications of MPT models to a variety of substantive areas in cognitive psychology, including various types of human memory, visual and auditory perception, and logical reasoning. We then address a number of theoretical issues relevant to the creation and evaluation of MPT models, including model development, model validity, discrete-state assumptions, statistical issues, and the relation between MPT models and other mathematical models. In the conclusion, we consider the current role of MPT models in psychological research and possible future directions.

This article presents a detailed review of a current and popular class of cognitive models called *multinomial processing tree* (MPT) models. MPT models have been described formally in Riefer and Batchelder (1988) and in Hu and Batchelder (1994b), although models of this type have been around well before the class was first formalized in 1988 (e.g., Batchelder & Riefer, 1980; Chechile & Meyer, 1976; Greeno, James, DaPolito, & Polson, 1978; Humphreys & Bowyer, 1980; B. H. Ross & Bower, 1981). However, the last 10 years have witnessed a deeper understanding and an accelerated use of these models within psychology. This increased popularity of MPT models has resulted not only in the application of these models to new areas in psychology but has also led to a variety of new statistical techniques and a certain amount of theoretical debate. Because of these developments, a review article on this class of models seems timely both for researchers already working in this area and for others who might benefit from using this type of modeling.

MPT models are simple, substantively motivated statistical models that can be used to measure underlying or latent cognitive capacities. Psychological data often result from multiple, interacting processes, and operationally defined statistics are quite limited in determining which of these processes are involved in a particular experimental paradigm. One primary use of MPT models is as data-analysis tools, capable of disentangling and measuring the separate contribution of different cognitive processes underlying observed data. This approach can be helpful in settling theoretical issues, because psychological theories often focus on one process or another as the fundamental cause of a particular psychological phenomenon. The structural simplicity of the class of MPT models also makes it a useful framework for developing and testing quantitative theories. It is possible to make theoretical assumptions precise when incorporating them into MPT models, and testing these assumptions is relatively straightforward using standard statistical theory.

An important characteristic of MPT models is that they are developed exclusively for categorical data (i.e., situations in which each experimental observation falls into one and only one of a finite set of categories). Categorical data are convenient to model because it is not necessary to model measurement error, which is a source of concern with continuous data. With categorical data, there is virtually no uncertainty as to the appropriate category for each observation, and, thus, the probabilistic structure of the data can be thought to arise from underlying processes of interest uncontaminated by error variance.

In the case of categorical data, the most general and theoretically neutral statistical distribution is the multinomial distribution, a natural generalization of the binomial distribution to more than two categories. In the multinomial distribution, observations are independent and identically distributed over the categories, and each

category has a parameter representing the probability that a random observation falls into it. Viewed generally, statistical models for categorical data express the probability parameters for each category as functions of the model's parameters (i.e., they reparameterize the multinomial distribution). Models of this type have a long history in scientific research and include such standard model families as log-linear and logit models (e.g., Agresti, 1990; Bishop, Fienberg, & Holland, 1975). Log-linear and logit models are routinely used for categorical data in a number of fields in the social, behavioral, and biological sciences, and their development has paralleled developments in modeling continuous data (e.g., analysis of variance [ANOVA] and the general linear model). As with ANOVA models, all that is required is a suitable factorial experimental design, and then a model can be selected without much regard to the substantive nature of the paradigm being modeled.

In contrast, MPT models are tailored explicitly to particular psychological paradigms. Each MPT model is a reparameterization of the category probabilities of the multinomial distribution; however, unlike log-linear and logit models, the reparameterization is in terms of parameters that are designed to represent underlying psychological processes, rather than a canonical decomposition in terms of main effects and various interactions. MPT models reflect a particular type of cognitive architecture, one that is represented as a tree with a single root (a type of directed graph having no cycles; e.g., Gondran, Minoux, & Vajda, 1984). Each branch of the tree represents a different hypothesized sequence of processing stages, resulting in a specific response category. Thus, MPT models are designed to represent situations in which an observed response category can arise from one or more unobserved processing sequences, represented by branches in the tree structure. Both the sequential character of processing stages and the many-to-one mapping of processing sequences into response categories are psychological properties that are conveniently represented in the rooted tree architecture.

One important consequence of this architecture is that category probabilities are generally expressed as nonlinear functions of the underlying psychological parameters. This contrasts with the usual models for categorical data mentioned earlier that have linearity built in at some level. Thus, even though MPT models parameterize latent processing events in a straightforward way, the manifest category probabilities are usually nonlinear polynomial functions of the processing event parameters. In this sense, MPT models share a property with most successful scientific models in the behavioral sciences—namely, that simplicity at the theoretical level may have more complex consequences at the behavioral level.

Even though MPT models are more substantively based than are off-the-shelf statistical models, they are usually much less detailed than are more theoretical cognitive models, such as global memory models (Clark & Gronlund, 1996) or neural network (connectionist) models

(J. A. Anderson, 1995). MPT models capture some of the psychologically important variables in a paradigm, but they are necessarily approximate and incomplete and are usually confined to particular paradigms. In this way, MPT modeling may be viewed as a type of cognitive psychometrics in the spirit of such approaches as paired-comparison scaling (David, 1988), signal detection models (Macmillan & Creelman, 1991), or information-integration models (N. H. Anderson, 1982; Massaro, 1987). Despite their disadvantage as approximations to psychological theory, MPT models share the advantage with other psychometric models of being statistically tractable and thus provide clear choices for data analysis over standard, multipurpose statistical models.

The article is organized as follows. First, we present a description of two previous applications of MPT modeling from our own laboratory. The purpose is to familiarize the reader with the ideas behind MPT models and, later, to use these examples to help illustrate important theoretical and statistical points. Next, we formally describe the cognitive structure and parametric properties of the class of MPT models, providing an explicit statistical analysis of this class. Following this, we provide a comprehensive review of the applications of MPT modeling to a variety of substantive areas in cognitive psychology. The purpose of this review is twofold. First, it will give the reader an idea of the scope of MPT modeling as it is currently applied in psychology, with references for researchers who want to learn more about a particular area or application. Second, the applications that we describe raise a number of theoretical and statistical issues related to MPT modeling. These issues will be discussed in the section following the review, and they should be relevant to researchers who wish to develop and use this type of modeling. Finally, we consider the current role of MPT models in psychological research and possible future directions.

TWO EXAMPLES

Batchelder and Riefer's Pair-Clustering Model

The first example of MPT modeling is a model that we designed to disentangle cluster storage from cluster retrieval in a standard free-recall paradigm (Batchelder & Riefer, 1980, 1986). The paradigm involves a study list consisting of two types of items, clusterable pairs (e.g., lawyer, teacher) and singletons (items without a category partner), followed by free recall of the list. Recall of the pairs is scored into four mutually exclusive categories: (1) both items recalled adjacently, E_1, (2) both items recalled nonadjacently, E_2, (3) one and only one item recalled, E_3, and (4) neither item recalled, E_4. Singletons are scored into two categories: (1) recalled, F_1, and (2) not recalled, F_2. As we indicated earlier, MPT models can be represented by processing trees, and Figure 1 presents the tree diagram for the pair-clustering model. The model in Figure 1 postulates two processing trees—one for the pairs and one for the singletons—which technically makes it
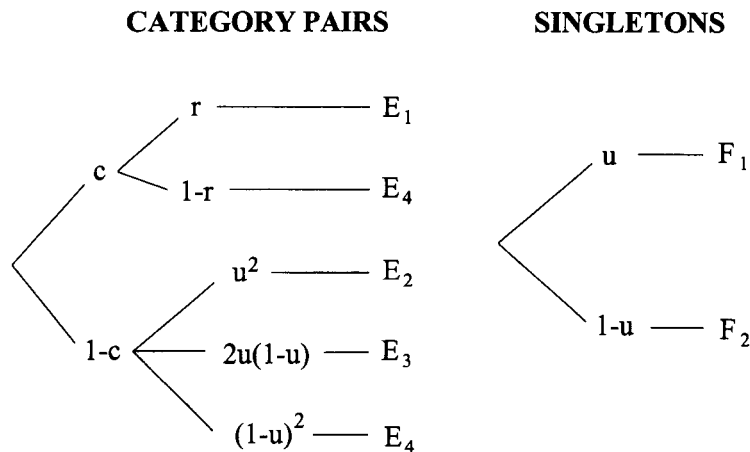
CATEGORY PAIRS   **SINGLETONS**

Figure 1. Batchelder and Riefer's (1980, 1986) pair-clustering model for measuring storage and retrieval processes.

a joint MPT model (Hu & Batchelder, 1994b, p. 30; Riefer & Batchelder, 1988, p. 324), a variety of MPT model described formally in the next section.

The model contains three parameters, c, r, and u, each designed to measure a different cognitive process. Parameter c measures the probability that the members of a pair are clustered and that the cluster is stored in memory during study. Parameter r is the conditional probability that a stored cluster is subsequently retrieved from memory during free recall. Because both parameters are probabilities, they must satisfy $0 \le c, r \le 1$. The top branch in Figure 1 has a probability cr, which is the probability of both storing and retrieving the cluster. The model assigns this branch to category E_1 on the assumption that retrieved clusters are recalled adjacently. The next branch has probability $c(1 - r)$, representing successful cluster storage but unsuccessful retrieval. This branch is assigned to category E_4, indicating recall of neither item in the pair. Notice that this is one of two branches that lead to category E_4, the other being at the lowest branch of the tree representing unsuccessful cluster storage. This illustrates one of the main characteristics of MPT models mentioned earlier—namely, that a particular response category can be obtained by more than one processing sequence or branch.

The third parameter, u, is the joint probability that a singleton is stored and retrieved. The tree for singletons is simple, reflecting a basic Bernoulli process in which singletons are either recalled or not with probability u or $1 - u$, respectively. Notice, however, that parameter u also appears in the tree for clusterable pairs. Here, it reflects a deeper assumption of the model—namely, that each item in a pair that is not clustered (with probability $1 - c$) is processed independently, with the same probability of recall as a singleton. In this case, u is the probability of storing and retrieving an item in a pair, conditional on it not being stored as a cluster. This dual interpretation of parameter u is a strong assumption, but

it can be evaluated by testing the model in Figure 1, which has three parameters and four degrees of freedom (three for the pairs and one for the singletons), against the more general model that allows the value of u for pairs to be different from that of the singletons.

Another strong assumption in the model concerns the recall of nonclustered pairs. Notice that with probability $(1 - c)u^2$, items in a pair are not clustered, and they are both recalled independently. This branch is assigned to category E_2, reflecting nonadjacent recall of the pair. This assumption is surely wrong in detail, because there is a random possibility for two nonclustered items to be recalled adjacently. The assumption is therefore clearly an approximation, but one that greatly simplifies the analysis of the model and still allows the model to reflect the main processing stages in the task. Even with this simplifying assumption, the model equations, representing category probabilities in terms of parameters, can be complicated. For example, for event E_4,

$$P(E_4) = c(1 - r) + (1 - c)(1 - u)^2, \tag{1}$$

which is a third-degree polynomial in the model's parameters. Among other things, this means that if $P(E_4)$ varies between groups in an experiment, it could be from changes in any subset of the three parameters.

Batchelder and Riefer's Source-Monitoring Model

The second example of MPT modeling is our model for source monitoring (Batchelder & Riefer, 1990). Source monitoring concerns the ability to remember the source of information acquired earlier. In a typical source-monitoring experiment, subjects study a list of items from two sources, A and B (e.g., List 1 or List 2, male or female voice). Subjects are then given a recognition-type test in which they must respond to old items and new distractors by classifying them as Source A, Source B, or New. The model for this paradigm is presented in Figure 2, and it specifies a processing tree for each of the three types of

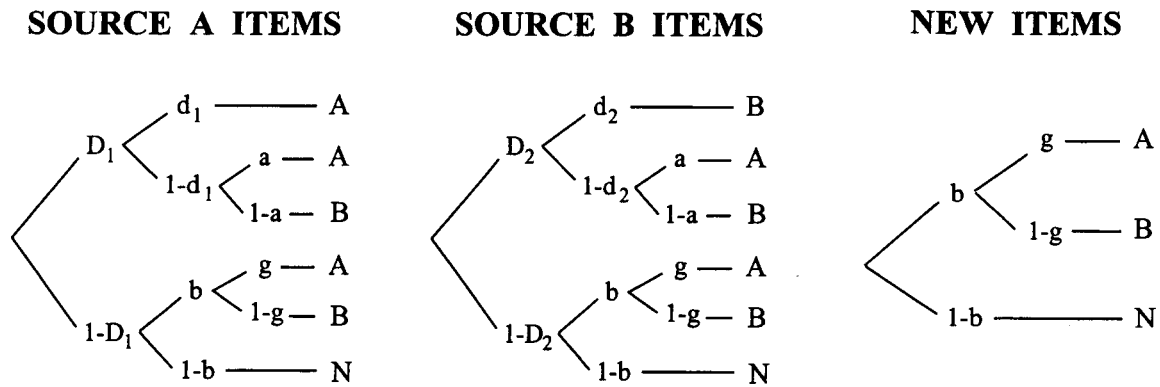SOURCE A ITEMS   SOURCE B ITEMS   NEW ITEMS

Figure 2. Batchelder and Riefer's (1990) source-monitoring model.

items. Considering all three trees, there is a total of 7 parameters in the model, 15 processing branches, and 9 response categories (3 for each tree, A, B, and N).

The model assumes that subjects first "detect" whether an item is old with probabilities D_1, D_2, or b for old A, old B, or new items, respectively. If an old item is detected as old, then d_1 and d_2 measure the capacity to discriminate the source of old A and B items, respectively. On the other hand, if the item is new but incorrectly detected as old, then only a guessing process, measured by parameter g, determines which source the item is assigned to. If discrimination of the source of an old item fails (with probability $1 - d_i$, $i = 1,2$), then another guessing process determines the source assignment, this time measured by parameter a. Finally, if an old item is not detected as old (with probability $1 - D_i$), it is treated as a new item, and the new-item tree is the continuation of the tree on the $1 - D_i$ branches.

There are a number of points that emerge from the model in Figure 2. First, it embeds some fairly strong psychological theory. For example, the model assumes that old–new detection precedes source discrimination and that correct source discrimination of an old item (apart from guessing) can occur only if the item is detected as old. A second point is that some parameters appear in more than one tree (i.e., a, g, and b). Most of these parameters can be interpreted as conditional probabilities, because they appear on links later in the tree, conditional on the success or failure of other processes. For example, in the two trees for old items, b is the conditional probability of deciding that an old item is old, given that it was not detected as old. But notice that b also appears in the tree for new items as an unconditional probability, that a new item is "detected" (biased) to be old. Thus, b plays a similar role in all trees (i.e., as a bias for responding "old"). However, it is evoked in the trees in different places and with different meanings when viewed as an event probability. A third point is that, as with most MPT models, the event probabilities are sums of products of the underlying parameters, a particular type of nonlinear function.

For example, the probability of correctly identifying an item from Source A is

$$P(\text{"A"} \mid A) = D_1 d_1 + D_1(1 - d_1)a + (1 - D_1)bg,$$

which is technically a third-degree polynomial in D_1, d_1, a, g, and b and represents an interaction of all the cognitive processes postulated in the model.

STRUCTURE OF MPT MODELS

Structural Properties

The structural characteristics of MPT models have been described in detail in two previous articles. In Riefer and Batchelder (1988), we discuss statistical inference for parametric multinomial models and give three examples. All three examples are represented in a tree diagram, and they satisfy strong structural properties. Hu and Batchelder (1994b) formally capture these structural properties for a class of models, which they term *general processing tree* (GPT) models. In this paper, we use the term *multinomial processing tree* (MPT) models to cover GPT models, as well as a slight extension described later in this section.

The formal requirements for constructing MPT models cover two aspects: (1) generating the tree structure, and (2) parameterizing the tree structure. We start by considering a single category system. For the tree structure, there are J observable categories, $C_1, C_2, \ldots, C_J$; for each category, there is one or more processing branches that lead to that category. Each branch is composed of a sequence of one or more directed links. Formally, the tree itself has three kinds of nodes: the root, the intermediate nodes, and the terminal nodes. The root is the starting point and is the unique node where there are no incoming links. The intermediate nodes correspond to various states of processing that may occur. The terminal nodes are the categories themselves. At any node, except for the category end nodes, there may be two or more directed links (out arcs), and these correspond to various directions that the processing sequence can take at that

point. Thus, the processing tree consists of a single root and a collection or processing branches, each terminating in a particular response category.

To parameterize the tree structure, the model must specify S functionally independent parameters, θ_s, $s = 1$, ..., S, and a parameter space $\Theta = (\theta_1, ..., \theta_s, ..., \theta_S) \in [0,1]^S$ (i.e., the S dimensional unit hypercube). The reason for the restriction that $0 \leq \theta_s \leq 1$ is that the parameters are interpreted as probabilities of various underlying cognitive processes. At each nonterminal category node, there is a probability distribution over the directed links that proceed from that node, and these distributions are required to satisfy a certain functional form in the underlying parameters, described in Hu and Batchelder (1994b). These requirements on the functional form of the link probabilities lead to a particular form for the branch probabilities, given by

$$P\left(B_{ij};\Theta\right) = c_{ij} \prod_{s=1}^{S} \theta_s^{a_{ijs}} \left(1-\theta_s\right)^{b_{ijs}}, \qquad (2)$$

where B_{ij} is the ith branch that leads to category C_j, c_{ij} is a positive number, and a_{ijs} and b_{ijs} are nonnegative integers (see Hu & Batchelder, 1994b, for details). The requirements in Equation 2 are often met by having many of the values for the a_{ijs} and b_{ijs} equal to zero.

To illustrate, consider the pair-clustering model in Figure 1. The tree reveals that categories E_1, E_2, and E_3 have only one branch, whereas E_4 has two branches. The parameter vector is $\Theta = (c, r, u)$, and, to take one example, the branch probabilities for category E_4 in terms of Equation 2 are

$$P(B_{1,4};\Theta) = 1 \times c^1 r^0 u^0 (1 - c)^0 (1 - r)^1 (1 - u)^0$$

$$= c(1 - r)$$

$$P(B_{2,4};\Theta) = 1 \times c^0 r^0 u^0 (1 - c)^1 (1 - r)^0 (1 - u)^2$$

$$= (1 - c)(1 - u)^2.$$

In practice, the tree structure of many MPT models has a binary, single-parameter form where there are only two directed links from any nonterminal node, with one link having probability θ_s and the other $1 - \theta_s$. In this case of binary, single-parameter links, it is easy to see that Equation 2 holds, with $c_{ij} = 1$, because a_{ijs} and b_{ijs} are just the number of links in the branch that have θ_s and $1 - \theta_s$, respectively. This is the case with all the MPT models depicted in the figures, except Figure 1, and the tree for clusterable pairs in Figure 1 is easily redesigned to meet the binary structure if the extension of the $1 - c$ link is rewritten as two successive binary, single-parameter links with parameter u.

Once the tree structure of the model has been established, the category probabilities are given, from Equation 2, by

$$p_j\left(\Theta\right) = \sum_{i=1}^{I_j} P\left(B_{ij};\Theta\right), \qquad (3)$$

where I_j is the number of branches terminating in Category j, and $j = 1, ..., J$. Finally, the probabilistic structure of the model requires that $\Sigma\, p_j(\theta) = 1$, for all $\Theta = (\theta_1, ..., \theta_s, ..., \theta_S) \in [0,1]^S$, which allows each parameter to vary independently in $[0,1]$. Thus, the parameters are functionally independent, and each has its full range in $[0,1]$. Given the restrictions on the branch probabilities, it is easy to see that the category probabilities are polynomial functions of the parameters. To illustrate from Figure 1,

$$P(E_4) = P(B_{1,4}; \Theta) + P(B_{2,4}; \Theta)$$

$$= c(1 - r) + (1 - c)(1 - u)^2,$$

which is a derivation of Equation 1.

The key to the statistical analysis of MPT models is the fact that, if the branch frequencies are known in addition to the category frequencies, then the maximum likelihood estimates (MLEs) of all the parameters can be written in a simple closed-form expression. Hu and Batchelder (1994b) show this in general, but to illustrate suppose that we have an MPT model written as a binary, single-parameter tree. Further suppose that each branch leads to a unique category (i.e., $I_j = 1$ for all $j = 1, 2, ..., J$), and, thus, the branch frequencies, N_j, are all known. Then, it is easy to see that the MLEs of the θ_s are given by

$$\hat{\theta}_s = \frac{\sum\limits_{j=1}^{J} a_{1js} N_j}{\sum\limits_{j=1}^{J} \left(a_{1js} + b_{1js}\right) N_j}, \qquad (4)$$

where $s = 1, 2, ..., S$. Equation 4 is easily interpreted: It is the number of times a link with probability θ_s is taken, divided by the number of times a binary link with parameter θ_s is encountered. In other words, Equation 4 is just a version of the well-known fact that the proportion of "successes" is the MLE for the probability of a success in a Bernoulli process.

The structural and parametric requirements of MPT models, especially those of Equations 2 and 3, are sufficiently restrictive to yield many consequences that are described in detail by Hu and Batchelder (1994b). Most importantly, they show that statistical inference, including goodness of fit, point and interval parameter estimation, and hypothesis testing, is straightforward for members of the model family by employing the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977; McLachlan & Krishnan, 1997). The EM algorithm is a well-known iterative method for obtaining MLEs for certain statistical models, in which some of the data can be regarded as missing. In this case, the missing data are assumed to be the branch frequencies, subject to constraint by the category frequencies (which are known). The EM algorithm starts by selecting initial estimates for the branch frequencies, then MLEs of the parameters are computed by a formula, such as Equation 4, then the branch frequencies are reestimated, and so on. Under

fairly general conditions, this algorithm necessarily leads to at least a local maximum in the likelihood function for MPT models, as shown in Hu and Batchelder (1994b, Observation 4).

The EM algorithm has a number of statistical advantages in working with MPT models over ad hoc iterative search routines, such as STEPIT (Chandler, 1969). First, it is not necessary to specify step size. Second, it provides asymptotic approximations to the variance–covariance matrix of the parameter estimators at the end of the search. Third, for the class of MPT models, it is guaranteed to produce a local minimum of the loglikelihood measure G^2, the quantity whose global minimization yields the MLEs. Fourth, it enables statistical inference to be accomplished within a single framework.

There are two potential disadvantages of the EM algorithm that Hu and Batchelder (1994b) have shown do not apply in the case of MPT models. First, for some classes of models, the EM algorithm is slow relative to other iterative search routines, such as ones based on gradient search methods. However, for the MPT class, both the expectation step and the maximization step are in simple, closed-form expressions; so no iterative search is required within each cycle of the algorithm. Second, the EM algorithm is designed to obtain only MLEs; however, in Hu and Batchelder (1994b), it is extended to the entire Read and Cressie family of goodness-of-fit statistics for categorical models (Batchelder, 1991; Read & Cressie, 1988). This family includes not only G^2 but other traditional fit methods, such as minimum chi-square and modified minimum chi-square.

There exists computer software based on the EM algorithm that has been developed by Xiangen Hu to handle all forms of MPT models. Researchers can use the graphic capabilities of this program to construct and display model trees, modify the trees, and combine many trees to form joint MPT models. Multiple data sets for analysis by the model can be input and saved, and the program performs all of the basic statistical calculations needed for MPT modeling, including parameter estimation (with confidence intervals), hypothesis testing, and model simulation. The program can be obtained by accessing the GPT website established in 1996 at http://irvin.psyc.memphis.edu/gpt/, where researchers can view documentation and download the program. Other computer programs specifically designed for other models can also be downloaded from this site, such as a program based on Batchelder and Riefer's (1990) source-monitoring model.

Global Identifiability

The statistical inference results for MPT models described above require that models satisfy a property called *global identifiability*. To illustrate, the source-monitoring model in Figure 2 has seven parameters, but there are only 6 degrees of freedom (df) in the data (i.e., each item type is scored into three response categories, so each item type yields 2 df). One consequence of this surplus of parameters is that it is not possible to uniquely estimate the mod-

el's parameters from data. Global identifiability of a model holds if there is at most one parameter vector underlying a given probability distribution over the categories.

More formally, suppose one has an MPT model with categories $C_1, \ldots, C_J$ and parameters $\Theta = (\theta_1, \ldots, \theta_S)$. Let $p(\Theta) = [p_1(\Theta), \ldots, p_J(\Theta)]$ express the category probabilities as a function of the parameter vector Θ. Then, the model is *globally identifiable* if $p(\Theta) = p(\Theta*)$ implies $\Theta = \Theta*$ for all $\Theta, \Theta* \in (0,1)^S$. A consequence of global identifiability is that if a probability distribution $p = (p_1, \ldots, p_J)$ is satisfied by the MPT model, then it is satisfied uniquely for some Θ. Thus, if the model holds, then, in principle, one can identify the latent parameter vector Θ given knowledge of the category probabilities. Global identifiability is a desirable condition for using the model as a measurement tool, because the process of deriving parameter values from category data leads to unique parameter estimates. Of course, a model may be testable (i.e., falsifiable) even if it fails to satisfy global identifiability (see Bamber & van Santen, 1985).

Batchelder and Riefer (1990) were able to achieve global identifiability for the source-monitoring model in Figure 2 by considering several psychologically motivated ways to reduce the number of parameters. In particular, they considered three restrictions of the model: (1) equal source detection probabilities $(D_1 = D_2)$, (2) equal source discrimination probabilities $(d_1 = d_2)$, and (3) equal response-bias probabilities $(a = g)$. By imposing one or more of these restrictions, a family of seven new models can be generated (six of which are globally identifiable); Batchelder and Riefer (1990) have proposed that this family, rather than a single model, be used to analyze source-monitoring data.

Joint MPT Models

Psychological paradigms yielding categorical data often involve more than one system of categories—for example, where each category system reflects a particular type of item. If responses to item types are independent, then the most general statistical model is a product of multinomial distributions, one for each category system. Joint MPT models postulate a separate processing tree for each of these category systems; however, if there is some overlap in processing events among the types of items, then a given parameter may occur in more than one of these trees. Another natural way that joint MPT models arise is the case in which one has data from two or more groups of subjects in the same paradigm. To test hypotheses about possible differences in parameters between groups, one constructs an identical processing tree for each group and then compares the fit of the versions where a given parameter is the same or different between groups.

Joint MPT models join the separate processing trees by representing the item types or subject groups as initial branches in one tree leading to each of the separate processing trees. Thus, a joint MPT model may be viewed as a hierarchical, two-stage model, in which the first stage reflects an experimenter assignment of item types

or subject groups, and the second reflects the subjects' responses to each item. Both the pair-clustering model and the source-modeling model described above are examples of joint MPT models. Hu and Batchelder (1994b) formalize joint MPT models and show that the statistical inference theory for MPT models extends naturally to joint MPT models.

Reparameterization

Several MPT models that will be discussed later do not satisfy the stringent parametric form required in Equations 2 and 3. For example, Batchelder, Hu, and Riefer (1994) extended the source-monitoring model in Figure 2 to cover the case of three or more sources. The extension required guessing parameters for each source (e.g., g_1, g_2, and $g_3 = 1 - g_1 - g_2$ for three sources). Clearly, this extension requires that $g_1 + g_2 \leq 1$, which restricts the parameter space from the requirement in Equation 3 that each parameter independently can take any value in [0,1]. However, in the case of three sources one can reparameterize the model by introducing two new parameters, g_1* and g_2*, $0 \leq g_1*, g_2* \leq 1$, where $g_1 = g_1*$, $g_2 = (1 - g_1*)g_2*$, and $g_3 = (1 - g_1*)(1 - g_2*)$. The reparameterized model is equivalent to the original model, in the sense that it can generate exactly the same set of probability distributions. Furthermore, it has the same number of parameters and satisfies the structural requirements of Equations 2 and 3.

More generally, in this article, we include as MPT models any parametric processing-tree form that can be reparameterized as an equivalent model, with the same number of parameters, satisfying Equations 2 and 3. So far, all of the models that we classify as MPT models in this article are easily reparameterized to satisfy these constraints. In fact, using the above definition, the general multinomial distribution for, say, J categories, is itself an MPT model. This is because it is always possible to reparameterize this distribution into MPT form by using an extension to J categories of the strategy above for reparameterizing the guessing parameters in the source-monitoring model for three sources.

It is not the case, however, that all probabilistic models for categorical data belong to the MPT class. Any model for J categories corresponds to a particular subset of all possible probability distributions on J categories, and there are many subsets (most, in fact) that cannot be modeled exactly by the MPT form in Equations 2 and 3. Thus, the MPT class does not contain all parametric models for categorical data. In this sense, it would be incorrect to view MPT modeling as a "universal framework," incapable of being falsified and able to fit categorical data over any set of experiments.

APPLICATION AREAS

In this section, we review the scope of MPT modeling by describing its application to a number of different areas in cognitive psychology. There is a rapidly growing literature on developing and testing MPT models for various paradigms, and the review includes over 80 articles that involve MPT models. There are several cognitive areas that utilize these modeling efforts, and we use these to organize the review. The examples below also provide illustrations for various theoretical and statistical issues that we discuss in subsequent sections.

Models for Traditional Memory Paradigms

Perhaps the most common application of MPT modeling has been in the area of human memory. Much of our own published work has been in this area, including the two MPT models in the previous section. In this section, we describe a series of models that have been developed for three long-standing and traditional issues in memory research: interference, association, and short-term memory (STM).

Interference effects. One of the first examples of an MPT type of model is a model for proactive inhibition developed by DaPolito (1967; reported in Greeno et al., 1978). The experimental paradigm behind the model is known as *modified-modified free recall* (MMFR), involving a paired-associate task in which subjects learn a list of A–B pairs, followed by a second list containing the same stimuli paired with different responses (A–C). Subjects are given the stimuli and asked to generate both responses. Proactive inhibition occurs when the recall of the C terms is poorer than a control group that does not study the A–B list. The basic idea behind the DaPolito model is that items sharing a common stimulus are stored within the same retrieval network, which is accessed with probability p. If the network is successfully accessed, then retrieval of each item occurs independently, with probability q for Item B and r for Item C. Although Greeno et al. (1978) never presented or analyzed the model in MPT form, Riefer and Batchelder (1988) demonstrated how an MPT structure for the model could be generated and derived closed-form MLEs and asymptotic confidence intervals for the model's parameters.

A version of the DaPolito (1967) model is shown in Figure 3. The model in Figure 3 is a general version that is appropriate not only for situations involving two distinct types of items (such as the DaPolito application) but also for some tasks involving two successive tests of memory. Like the DaPolito model, the model assumes that items are stored in memory with probability s, and then retrieval is conditionally independent on successful storage, with probability r_1 for the first test and r_2 for the second. It should be easy to see that parameters p, q, and r in the DaPolito model correspond respectively to parameters s, r_1, and r_2 here. There are four recall events for the model: successful recall on both tests (SS), successful recall on Test 1 but not Test 2 (SF), successful recall on Test 2 but not Test 1 (FS), and recall failure on both tests (FF). Both Riefer and Batchelder (1995) and Bender, Wallsten, and Ornstein (1996) have used this model for situations involving repeated testing of memory, and details of these applications will be reviewed later.
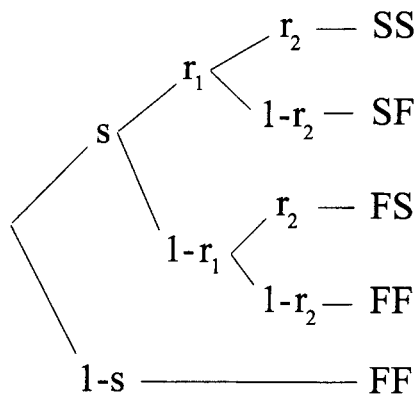
Figure 3. An MPT model for measuring storage and retrieval processes in paradigms involving two type of items or two successive memory tests.

Greeno et al. (1978) were interested in investigating the independent retrieval phenomenon, which is the observation that the recall of B and C is stochastically independent even when strong interference effects occur. They applied the model to an experiment in which they manipulated the number of A–B presentations before the A–C list. Even though strong interference effects were obtained in the experiment, the number of A–B presentations did not significantly influence the probability of recalling C. In terms of the model's parameters, Riefer and Batchelder (1988) were able to show that the number of A–B presentations significantly increased parameter q but had no reliable effect on parameter r. This pattern of results provided strong support for the retrieval independence theory of proactive inhibition.

Riefer and Batchelder (1988) have also explored a paradigm similar to the one for proactive inhibition described above. In their experimental task, subjects were presented with two lists in which similar stimuli were paired with different responses (e.g., taxi–"2" and cab–"7"). This paradigm therefore involves an A–B, A'–C task, in contrast to the A–B, A–C task used by Greeno et al. (1978). In addition, both presentation and testing were alternated between the A–B and A'–C items, unlike the Greeno et al. task in which all study trials on the A–B list occurred before the study trials of the A–C list. Because the stimulus terms in the Riefer–Batchelder (1988) task were related and potentially confusable, subjects had to learn to discriminate between A and A' before they could successfully associate the proper responses.

A key component in Riefer and Batchelder's (1988) analysis was their expansion of the traditional error–success data to include confusion errors. These are errors in which the response for one stimulus is actually given for the related stimulus (e.g., responding "7" to taxi). They also included data from unique paired associates (i.e., ones with stimulus terms that were distinct from one another). This yielded enough response events for the creation of an MPT model capable of measuring the pro-

cess of stimulus–response learning separately from stimulus discrimination. Riefer and Batchelder (1988) were able to derive closed-form solutions for the parameter estimates, including asymptotic confidence regions. The model was applied to a repeated-trials experiment, and it showed that, over early trials, the A'–C pairs exhibited less stimulus–response learning and more stimulus confusion than did the A–B pairs, which had the advantage of being presented first on each trial. This difference disappeared over later trials, however, as subjects learned to discriminate between the similar stimuli.

Associative recall. The study of memory for related or associated information has a long history in psychological research. B. H. Ross and Bower (1981) explored different theoretical accounts of associative memory in order to determine which theories do the best job of describing empirical data. Their basic paradigm required subjects to memorize clusters of four or five words related by a common theme (e.g., apron, chair, brush, clip: haircut), followed by a memory test in which one, two, or three words from the cluster were provided as retrieval cues. As B. H. Ross and Bower noted, this task goes beyond the traditional paired-associate methodology in which items consist of word pairs, with one of the words cuing the other. The purpose of using larger clusters and potentially more cues was to provide a more detailed set of data for discriminating between different theories.

B. H. Ross and Bower (1981) tested three formal models of associative recall: the horizontal model, the schema model, and Jones's fragment model. For the horizontal model, items are stored separately but have direct associations to each other. The two parameters of the model represent the probability that each item is stored within the associative structure and the probability that a stored item leads to retrieval of an associated item. For the schema model, items are also stored separately but are connected to a common schema instead of being directly associated to each other. This model also has two parameters: the probability that an item has access to the schema, and the conditional probability that an accessed schema leads to retrieval of an associated item. The final model tested was Jones's fragment model. This model assumes that items are not stored separately but instead are associated within fragments, or "chunks," of different sizes, with no associated links between the fragments. Parameters of the model represent the probabilities that each type of fragment is stored within memory. The structure of each fragment determines whether one item will successfully cue another, because an item can cue another item only if they share the same fragment.

All three models can be represented in the MPT class, and B. H. Ross and Bower (1981) were able to obtain minimum chi-square parameter estimates and assess the goodness of fit of the models. They accomplished this by using STEPIT (Chandler, 1969) to minimize the chi-square difference between the models' predictions and the actual data frequencies. In general, the fragment model performed poorly across a series of three experiments.

The horizontal model fared somewhat better, providing a good quantitative fit to data in one experiment but poor fits on two others. The best overall results occurred for the schema model, which provided the best fits over the three studies. B. H. Ross and Bower discuss the difficulty of choosing a specific theoretical framework when exploring the underlying structure of semantic memory, and they conclude that the schema model "may be recommended for use by investigators working on associative learning but wishing to avoid the strong commitment to specific structural representations of memories made by extant theories of the day" (p. 15).

Short-term memory. Schweickert (1993) has developed a simple MPT model designed to capture the basic processes in immediate recall of short lists of items from STM. The model assumes that if a memory trace is intact (with probability I), then an item is directly recalled. If the trace is too degraded for a direct readout (with probability $1 - I$), it may be reconstructed (with probability R) using other cognitive processes. This model predicts that successful recall from STM has the function $P(\text{recall}) = I + (1 - I)R$, an equation in the same form as the one originally proposed by Waugh and Norman (1965). The model itself does not have enough degrees of freedom to obtain separate estimates for both I and R from a single set of empirical data. However, Hulme et al. (1997) were able to apply the model to a series of experiments examining the effects of word frequency and word length on serial position curves. By making two assumptions— that the value of I is the same for low- and high-frequency words, and that R is a constant across serial position— Hulme et al. were able to apply the model to data, estimate parameter values, and measure goodness of fit. The model did a generally good job of fitting the data and correctly predicted that recall differences between high- and low-frequency words would increase across serial position.

Storage–Retrieval Models

Figure 1 presents an MPT model for separately measuring cluster storage from cluster retrieval. In general, the measurement of storage and retrieval processes in memory represents one of the most common applications of MPT models. This is because a storage–retrieval account of memory allows for memory failure to be explained by more than one process (e.g., storage failure, or retrieval failure despite successful storage). A recurring issue in memory research is whether memory phenomena are caused primarily by differences in storage or retrieval capacities, and a wide variety of storage–retrieval hypotheses has been proposed in many areas of memory research (e.g., Payne & Roediger, 1987; Schonfield & Robertson, 1966; Tulving & Psotka, 1971). The MPT models reviewed below represent different attempts to measure these two processes separately. As will be seen, each model is applied to a different paradigm, and, thus, each makes different theoretical assumptions about how empirical statistics can be translated into measures of storage and retrieval. Most of the following models are motivated by

informal assumptions that have been previously used by researchers to make qualitative evaluations of storage and retrieval. The advantage of the models is that they are able to take these informal assumptions and make them more explicit, thus providing quantitative measures of these processes that may vary as a function of experimental manipulations.

Chechile's storage–retrieval model. One of the first explicit applications of process-tree modeling in psychology was a storage–retrieval model developed by Chechile (1987; Chechile & Meyer, 1976). The experimental paradigm for the model involves a standard memory task in which recall trials are randomly supplemented with recognition trials. On the recognition trials, subjects are required to give a "yes"–"no" response along with a confidence rating on a 3-point scale. The data events for the model arise from the combination of recall data, recognition data, and confidence ratings, and these data provide enough information to estimate storage (θ_S), retrieval (θ_R), and various guessing and response-bias parameters.

The basic idea behind the Chechile model rests on a common assumption in memory research—that recall requires both storage and retrieval (represented as θ_S and θ_R in the model), whereas recognition depends only on sufficient storage. By incorporating these assumptions into a formal model, measurement of storage and retrieval capacities becomes possible. More recently, Chechile (1993) has extended this modeling approach to measure fractional storage. The assumption behind this new work is that storage failure ($1 - \theta_S$) can be subdivided into fractional (or partial) storage and no storage. Thus, the new model does not make storage an all-or-none process, and, therefore, it gives a more detailed and possibly more correct measure of storage processes.

One advantage of the Chechile (1987) model is that it can be applied to a number of different memory paradigms. For example, the model has been used to measure storage and retrieval processes in serial list learning (Skoff & Chechile, 1977), paired-associate learning (Chechile & Ehrensbeck, 1983), and the Brown–Peterson paradigm (Chechile & Butler, 1975; Gerrein & Chechile, 1977). Furthermore, Chechile and his associates have applied the model to a wide range of basic memory issues, including interference effects (Butler & Chechile, 1976), the serial position curve (Skoff & Chechile, 1977), acoustic similarity (Chechile, 1977), semantic memory (Chechile & Richman, 1982), and others. This corpus of experimental results also provides a number of validation tests of the model as a tool for separately measuring storage and retrieval factors. Experimental manipulations have been shown to have a selective influence on specific parameters; for example, increasing search time improves θ_R without affecting θ_S (Chechile & Meyer, 1976), acoustic similarity influences θ_S but not θ_R, and manipulating the similarity of foils in the recognition task changes the guessing parameter without influencing θ_S or θ_R.

Riefer and Rouder (1992). The Chechile (1987) model uses the contrast between recall and recognition to

measure storage and retrieval. Riefer and Rouder (1992) have proposed a model for measuring storage and retrieval that examines the contrast between free and cued recall. The model is based on the assumption that free recall reflects both storage and retrieval processes, whereas cued recall is a more direct indicator of storage alone. The experimental paradigm behind the model involves a paired-associate task in which memory for noun pairs is tested successively with free recall followed by cued recall. For free recall, there are three possible responses—both nouns, one noun, or neither noun in a pair can be recalled—whereas cued recall results in successful or unsuccessful recall of the second noun given the first as the cue. This results in six (3 × 2) response categories when performance is combined across both free recall and cued recall. Rouder and Batchelder (1998) have developed a family of closely related alternative versions of this model and have explored a number of their statistical properties.

Riefer and Rouder (1992) used the model to investigate the bizarreness effect, the observation that bizarre or unusual stimuli are recalled better than common stimuli. The model revealed that bizarre stimuli are retrieved better from memory, but not stored better, than common stimuli. Conversely, Riefer and LaMay (1998) demonstrated in a follow-up study that common stimuli are stored in memory better than bizarre stimuli. These findings help explain why the bizarreness effect has been weak or nonexistent in many previous studies that have relied solely on traditional statistical analyses, such as ANOVA. Depending on the relative contribution of storage and retrieval processes, the combined empirical effect of these two factors, as measured by ad hoc statistics, such as percent correct, can produce conflicting memory results that may be hard to interpret. The above examples are good illustrations of how MPT models can disentangle the relative contributions of separate cognitive processes, especially when these processes have opposite effects on the empirical data.

Batchelder and Riefer (1980, 1986). Earlier, we described in detail our pair-clustering model for measuring storage and retrieval (see Figure 1). The basic assumption behind that model is that adjacent recall is taken as an indication of successful storage and retrieval of a pair cluster, whereas nonadjacent recall is an indication of failure to store a pair as a cluster. Variations of this model include an extension by Riefer (1982) to clusters of more than two items and Markov versions of the model by Batchelder and Riefer (1980) and Bäuml (1996a). Like the Chechile (1987) model, the Batchelder–Riefer model has been applied to a number of different theoretical issues and experimental variables. For example, Riefer and Batchelder (1988) showed that memory loss caused by retroactive inhibition (RI) is due to poorer retrievability and not storage, a finding that was replicated by Bäuml (1996b). However, another experiment by Bäuml (1991) revealed that storage factors can play a role in RI when spaced presentation of the word lists is used, in contrast to the blocked presentation used by Riefer and Batchelder

(1988). The model has also been used to conduct storage–retrieval analyses on a number of basic variables that are known to affect memory. For example, we (Batchelder & Riefer, 1980, 1986; Riefer & Batchelder, 1987) have shown that storage capacity is improved by longer presentation rates, higher category association, and smaller interitem lags. In contrast, retrieval capacity benefits from cuing during recall and large interitem lags. These findings are consistent with reasonable psychological assumptions about the separate effects of these variables on storage and retrieval.

Children's memory. Issues of storage and retrieval play an important role in theories of children's memory. The question is whether the main developmental changes in memory are due to changes in storage or retrieval capacities, or possibly both. For example, the Chechile (1987) model has been applied to this issue, showing that both storage and retrieval abilities improve across age groups, although not necessarily at the same rate (Chechile & Richman, 1982; Chechile, Richman, Topinka, & Ehrensbeck, 1981). A number of other MPT models have been proposed to measure these processes in children. The models use different memory paradigms; however, they all involve repeated recall, in which the child's memory for an event or word list is tested over a series of successive recall tests.

Howe's trace-integrity model. The most extensive application of MPT modeling to children's memory has been conducted by Howe and his associates (e.g., Howe, 1991, 1995; Marche & Howe, 1995). In their experimental paradigm, children learn a list of words to criterion, followed at various delays by four successive recall attempts of the list. The data events consist of the 16 (2^4) four-tuples that result when all combinations of successful and unsuccessful recall are tabulated for each item across the four trials. The model itself is based on trace-integrity theory (Howe & Brainerd, 1989), which assumes that stored memory traces can be forgotten but that it is also possible to restore decayed memory traces through reminiscence. The model thus contains storage and retrieval parameters, as well as parameters for various forms of reminiscence.

The trace-integrity model has been used by Howe and his associates to examine developmental trends in a large corpus of experiments (see Howe & O'Sullivan, 1997, for a review). The overall results of these studies are argued to support the hypothesis that age-related memory differences tend to reflect storage capacity more strongly than retrieval capacity. This is based on experiments in which developmental differences are found for storage but not retrieval (e.g., Brainerd, Reyna, Howe, & Kingma, 1990, Experiments 1 and 3; Howe, 1991), as well as studies in which storage and retrieval both increase with age but with stronger changes for storage (e.g., Brainerd et al., 1990, Experiment 2; Howe, 1995). In addition to age differences, the trace-integrity model has also been applied to a number of other factors that have been previously known to influence memory, with theoretically plausible results. For example, the model has shown that pictures

are both stored and retrieved better than words (Brainerd et al., 1990; Howe, Kelland, Bryant-Brown, & Clark, 1992), that semantically related items are stored better in memory than unrelated items (Brainerd et al., 1990), and that extra presentation trials benefit storage but not retrieval (Howe, 1995). Other studies have examined the influence of postevent information on children's memories. Howe, Courage, and Bryant-Brown (1993) found that reinstating memories after a 3-week interval improves storage and retrieval rates, but with a stronger effect on storage, whereas Howe (1995) observed that RI adversely affects storage capacity, with weak or no effects on retrieval capacity.

Memory in language-impaired children. A model similar to Howe's was developed earlier by Kail, Hale, Leonard, and Nippold (1984). Their model is also based on the theoretical framework that stored traces can be strengthened by recall procedures, although the specifics of their model differ somewhat from Howe's. In addition, the experimental task behind the Kail et al. model involves only three successive recall attempts instead of four, resulting in eight (2^3) data events. Despite fewer data categories, there are still enough degrees of freedom in the data to estimate the model's parameters.

Kail et al. (1984) used this model to explore memory deficits in language-impaired children. Their empirical analysis focused on the differences between free and cued recall performance and indicated the presence of strong storage deficits. However, these empirical results were not clear-cut concerning whether retrieval deficits were also a contributing factor. The model clearly revealed that memory problems in language-impaired children were the result of both storage and retrieval deficits. This example and the previous studies by Howe and his associates nicely illustrate the advantages of formal modeling over ad hoc statistics in the measurement of cognitive processes. When multiple cognitive processes influence behavior, it is sometimes difficult using traditional statistical techniques to determine whether a single factor or a combination of factors is the primary cause of the behavior. MPT models can provide a more formal and quantitative measure of each cognitive process, which can be helpful in establishing their relative contributions in different experimental situations.

Memory for personal experiences. Bender et al. (1996) have also applied a storage–retrieval model to children's memory; however, unlike the above studies, they examined memory for the real-life events experienced during a visit to the doctor. Children (3–7 years old) made two recall attempts for these events, one immediately after the visit and one after a delay. This created four (2^2) possibilities for successful (S) or unsuccessful (F) recall: SS, SF, FS, and FF. As Bender et al. pointed out, the model they developed was equivalent in form to the one developed by DaPolito (1967) for proactive inhibition, represented in Figure 3. As stated earlier, the model has three parameters, a storage parameter plus a conditional retrieval parameter for each of the two recalls. Thus, there

are enough degrees of freedom (3) to uniquely estimate parameter values, but not enough to test the fit of the model to data from a single group of subjects.

However, Bender et al. (1996) were able to generate a nested hierarchy of models by constraining certain parameters to be equal across age and recall attempts. By doing so, they were able to test the fit of different nested models, especially those making assumptions about storage or retrieval changes across age groups. The results showed that while storage capacity improved significantly with age, the strongest developmental improvements were due to retrieval. This finding contrasts to the analyses based on Howe's model discussed earlier, which has recorded developmental gains for both storage and retrieval abilities in a different paradigm. However, Howe's research suggests that storage improvements across age tend to be much larger than improvements in retrieval. Of course, the memory paradigms differ in the two studies, and further research and modeling analysis will be needed to determine the precise circumstances leading to storage or retrieval differences across age groups.

Recognition failure. A curious observation in memory research is that it is possible to correctly recall words that earlier could not be recognized, a phenomenon known as *recognition failure of recallable words*. This occurs within the recognition-failure paradigm, in which paired associates (A–B) are presented to subjects followed by two memory tests: a recognition test for the B terms, followed by a cued recall test of the B terms given the A terms as cues. The data consist of a 2 × 2 table of correct and incorrect responses for both recognition and recall. Riefer and Batchelder (1995) have used a simple MPT model to measure storage and retrieval processes in this paradigm and, in particular, to explore the Tulving–Wiseman (Tulving & Wiseman, 1975) function, which predicts a systematic relationship between recognition versus recognition given recall. One version of the model is structurally equivalent to the one in Figure 3, where r_1 and r_2 are the retrieval probabilities of recognition and cued recall, respectively. Riefer and Batchelder (1995) show that this model is actually a special case of the more elaborate retrieval-independence theory of recognition failure (Flexser & Tulving, 1978), which assumes that item retrieval during recognition and during cued recall are conditionally independent events, given that the item is successfully stored.

In particular, we have applied the model in Figure 3 to a large corpus of data from recognition-failure studies collected by Nilsson and Gardiner (1993). The simplicity of the model allowed us to explore some of the basic properties of this paradigm, as well as the Tulving–Wiseman function itself. For example, we were able to generate specific predictions, in the form of model equations, as a function of certain theoretical assumptions. Among other things, the model revealed that exceptions to the Tulving–Wiseman function occur only when weak storage is coupled with strong retrieval. By assuming that storage and retrieval capacities are positively correlated within a par-

ticular condition, we were able to show that the MPT model sheds light on why the Tulving–Wiseman function appears to fit data so well. However, we also showed that the Tulving–Wiseman function has surprising consequences for data that are not satisfied, and, thus, the function itself (but not the MPT model) has a questionable status.

Another application of MPT modeling to the recognition-failure paradigm has been conducted by Humphreys and Bowyer (1980). Their contention is that priming during the recognition test plays a significant role in this paradigm, due to the fact that the presentation of the B term during recognition could boost the memory for B on the subsequent recall task. To explore this, Humphreys and Bowyer developed an MPT model that incorporates a parameter for priming, as well as parameters for storing items and correctly recognizing them given sufficient or insufficient storage. Given the precise structure of the model, they were able to incorporate different theoretical assumptions and then determine the precise predictions of the model based on those assumptions. For example, retrieval independence was incorporated in the model by constraining two of the model's parameters to be equal. By comparing the model's predictions with the actual data, Humphreys and Bowyer were able to argue that priming is in fact an important factor in the recognition-failure paradigm. A related MPT model for this paradigm can be found in Humphreys and Bain (1983).

Models of Source Memory

A common experimental task in memory research requires subjects to memorize information that comes from multiple sources. An example of this is the source-monitoring paradigm described earlier. Sometimes, subjects are aware of these sources and are specifically instructed to keep track of which items come from which sources. At other times, subjects are not informed that different information may be coming from other sources, such as occurs when inaccurate text information is presented in the eyewitness suggestibility effect (Lindsay, 1993). In either case, memory performance can be evaluated not only by accuracy on source memory but also by the types of errors made when information is attributed to the wrong source. The combination of successes and errors to different sources yields a potentially rich set of categorical data for the creation of MPT models. In turn, these models are capable of measuring basic cognitive processes in these tasks, such as item detection, source discrimination, and the cognitive effect that information from one source can have on memories based on a different source.

Source monitoring. Our source-monitoring model described in Figure 2 has been used in several studies to investigate a variety of theoretical issues. For example, Johnson, Kounios, and Reeder (1994) used a speeded-response paradigm (e.g., Corbett & Wickelgren, 1978) to compare the time course of source-monitoring judgments.

Subjects were required to make source judgments at lags ranging from 300 to 1,500 msec, and analysis by the model showed that recognition accuracy (as measured by D) develops more quickly than source accuracy (as measured by d). This is consistent with the model's assumption that item detection necessarily precedes source discrimination. In another study, Mulligan (1996) examined the influence of perceptual interference (e.g., brief exposure to stimuli followed by a pattern mask) on various aspects of memory, including source monitoring. He found, somewhat surprisingly, that certain levels of perceptual interference during encoding can actually enhance later recognition memory, although it has no effect on source memory. Other studies have used the model, as a supplement to more traditional analyses, to investigate age differences in source monitoring (Ferguson, Hashtroudi, & Johnson, 1992; Light, LaVoie, Valencia-Laver, Albertson Owens, & Mead, 1992; Lindsay, Johnson, & Kwon, 1991) and discrimination between true and false statements (Begg, Anas, & Farinacci, 1992).

The source-monitoring model in Figure 2 has also been extended in a number of important ways. Batchelder, Hu, and Riefer (1994) have expanded the model to handle situations involving any number of sources, and, in particular, Riefer, Hu, and Batchelder (1994) used a three-source version of this model to explore the role that response bias plays in source judgment. Batchelder, Riefer, and Hu (1994) have proposed a low-threshold version of the model, as well as one based on signal-detection theory. In a more extensive study, Bayen, Murnane, and Erdfelder (1996) compared the original source-monitoring model with various high- and low-threshold variations and tested these models on data from a factorial experiment manipulating item similarity and source similarity. They argue that the only model to accurately account for their data in a psychologically plausible way was a two-high-threshold (2HT) version, which differs from the original model by postulating an additional parameter, D_3, representing the probability that new distractors can be detected as new. Bayen and Murnane (1996) successfully used this model to investigate how older adults use perceptual and temporal information to make source-monitoring judgments. Erdfelder and Bredenkamp (1998) used a variation of the 2HT model to study source memory for script-typical versus script-atypical information, presented in either whole or fragmentary form. Finally, Klauer and Wegener (1998) provide an important new model for the "Who said what?" paradigm in social categorization.

Process dissociation. An area closely related to source monitoring is the process-dissociation paradigm. Jacoby and coworkers (e.g., Jacoby, 1991, 1998; Jacoby, Toth, & Yonelinas, 1993) have developed a two-process theory of recognition memory based on Mandler's (1980) idea that recognition of an item can occur either through a specific, conscious "recollection" of the item or from a sufficient feeling of "familiarity." Arguments for this distinction come from experiments designed to show that these two processes can be dissociated (i.e., that they can be

affected differentially by experimental factors or individual differences). In an effort to separately measure recollection and familiarity, Jacoby (1991) invented the process-dissociation procedure. Two groups of subjects each study two successive lists, followed by a "yes"–"no" recognition test with old list items and new distractors. Group 1 (the inclusion group) is instructed to say "yes" for old items in either list, whereas Group 2 (the exclusion group) is instructed to say "yes" only to old List 2 items and to say "no" to old List 1 items and to distractors.

Jacoby (1991) developed a model for this paradigm, which assumes there is a probability R that a list item is recollected and probability F that it has sufficient familiarity to be called "old." The model further assumes that the two processes are independent, yielding the equations

$$p_1 = P(\text{"yes"} \mid \text{List 1 item, Group 1})$$
$$= R + (1-R)F \tag{5}$$

and

$$p_2 = P(\text{"yes"} \mid \text{List 1 item, Group 2})$$
$$= (1-R)F. \tag{6}$$

Estimates of R and F are generated by solving for Equations 5 and 6, yielding $R = p_1 - p_2$ and $F = p_2 / (1 - R)$.

As Buchner, Erdfelder, and Vaterrodt-Plunnecke (1995) note, it is easy to view the model in Equations 5 and 6 as a simple MPT model, with separate trees for Group 1 and Group 2 responses. But Buchner et al. (1995) also argue that there need to be trees for List 2 items as well as distractors. In addition, unlike most MPT models for recognition memory, there are no guessing probabilities in Jacoby's (1991) original formulation. On the other hand, because of the experimental design, it might be expected that the exclusion group would have a lower bias for "yes" responses than the inclusion group. As a consequence, Buchner et al. (1995) expanded the model to include identifiable guessing probabilities and argued that the model fit the data in a series of experiments designed to influence guessing but not memory processes. They also challenged the necessity of postulating that recollection and familiarity are independent processes. In fact, Buchner and Erdfelder (1996) discuss models that make recollection and familiarity mutually exclusive processes, as well as ones in which one process is conditional on the other.

Yonelinas and Jacoby (1996) have proposed an alternative processing-tree model in which familiarity is handled by a Gaussian signal detection model, rather than a discrete-state process characteristic of Buchner et al.'s MPT model. Yonelinas and Jacoby's model is motivated by the desire to account for receiver operator characteristics (ROCs) based on confidence ratings, where subjects indicate the confidence of their response assignment on a discrete scale. With the addition of confidence ratings, the data structure is still categorical, and, in fact, there are

even more categories to support a richer model. Yonelinas and Jacoby (1996; also see Yonelinas, 1994; Yonelinas, Regehr, & Jacoby, 1995) argue that the Buchner et al. model is based on "high-threshold" assumptions (discussed later), which predict linear ROCs based on confidence ratings. In a well-argued reply, Erdfelder and Buchner (1998b) show that this is not the case, and they extend their MPT model to predict confidence rating data, as well as the more traditional data structure in the process-dissociation paradigm.

Currently, the process-dissociation paradigm is very popular, and a number of researchers are developing models, including MPT models, to handle data in this and a variety of related paradigms (e.g., Buchner & Erdfelder, 1996; Buchner, Steffens, Erdfelder, & Rothkegel, 1997; Buchner & Wippich, 1996; Dehn & Engelkamp, 1997; Mulligan & Hirshman, 1997; Wainwright & Reingold, 1996). An interesting application comes from Buchner, Erdfelder, Steffens, and Martensen (1997), who directly compared the process dissociation and source-monitoring paradigms. They point out the close correspondence between the response categories in each paradigm, and, by constructing related MPT models, they argue from an analysis of experimental data that both paradigms share the same psychological processes. In general, the process-dissociation paradigm appears to be an area where MPT models, with their structural simplicity and statistical advantages, are likely to have a substantial impact.

Eyewitness memory. The eyewitness suggestibility effect is the observation that misleading information presented after an event can distort the memory for that event. In a typical experiment, subjects are presented with information about an event (e.g., a car approaching a *traffic light*) and later read a text containing information that is consistent (traffic light), inconsistent (stop sign), or neutral (intersection). An important theoretical question concerns what happens to the memory trace of the original stimulus. Loftus has theorized that original information can be destroyed and updated with postevent information (e.g., Loftus, Donders, Hoffman, & Schooler, 1989), whereas McCloskey and Zaragoza (1985) claim that the original information is still intact.

Wagenaar and Boer (1987) explored this issue by representing these informal theories as formal models, which they referred to as *event-tree models*. Each of the models satisfied the MPT properties described earlier. Three theories were examined: (1) destructive updating, in which the original stimulus can be erased and replaced, (2) coexistence theory, in which the original stimulus is always intact but sometimes inaccessible, and (3) no-conflict theory, in which inaccurate memories only occur on those portion of trials when the original information is insufficiently encoded. The experimental paradigm was as described above, except that Wagenaar and Boer added a final "second-guess" phase in which subjects were informed that a traffic light was in fact the correct

response and were then asked to state the color of the light. The data events consisted of the combination of correct and incorrect responses to questions about the traffic light and its color.

The advantage of representing each theory in the form of an MPT model was to make the assumptions behind each theory explicit, thereby providing a more precise test of each theory's ability to account for experimental results. In particular, Wagenaar and Boer (1987) were able to generate a series of equations representing the specific predictions made by each theory. The results showed that the no-conflict theory produced the best fit to the data with the fewest parameters.

Hindsight bias. Situations sometimes arise in which people make predictions or judgments about an event, followed by feedback on the event's actual outcome. Under these conditions, it is necessary to discriminate between two sources of information: one's original judgment and the true outcome. However, research shows that people's memory for their original judgment often tends to be skewed in the direction of the true outcome, a phenomenon known as *hindsight bias* (Hawkins & Hastie, 1990).

Erdfelder and Buchner (1998a) examined hindsight bias using an experimental task in which subjects gave their best answers to a series of difficult questions requiring unique numerical responses (e.g., "What is the melting temperature of lead?"). After a delay, subjects were provided with feedback of the correct answers for half of the questions, with the other half serving as controls. Subjects' recall of their original responses was later tested in a final memory test. The data events consisted of the rank ordering of three numerical judgments: the original judgment (OJ), the correct judgment (CJ), and recollection of the original judgment (ROJ). Hindsight bias occurs, for example, when the recollected judgment is different from the original judgment, and in the direction of the correct judgment (e.g., OJ < ROJ < CJ). Allowing for ties between judgments, and assuming that subjects' original responses are incorrect (i.e., OJ ≠ CJ), there are 10 possible rank orderings. This creates a total of 20 response categories when both the feedback and the control items are analyzed.

Erdfelder and Buchner (1998a) identified a number of cognitive factors that can be instrumental in causing hindsight bias and developed an MPT model that incorporated these factors as parameters of the model. The full version of the model contained 13 parameters, representing processes of recollection, biased and unbiased reconstructions of the original judgment, and guessing. To test the model, Erdfelder and Buchner (1998a) conducted a series of validation experiments designed to show that certain experimental manipulations selectively influence some parameters but not others. Among other things, these tests showed that (1) providing the correct answers at the time of the final recall test increased reconstruction bias without affecting other parameters, (2) casting doubt on the accuracy of the correct answers reduced hindsight bias by improving the recollection parameters,

and (3) manipulating the number of response alternatives on the final memory test affected only the model's guessing parameters. On the basis of these and other experimental validation studies with the model, Erdfelder and Buchner (1998a) concluded that reconstruction bias seems to play a larger role in creating hindsight bias than recollection bias. All of these results are psychologically plausible and help validate the model as a viable tool for measuring cognitive processes in hindsight bias. In a follow-up paper, Dehn and Erdfelder (1998) have proposed a modification of the Erdfelder–Buchner model to handle situations lacking unique correct judgments.

Models of Perception

It should be evident from the examples so far that many of the applications of MPT models have been in the area of human memory. But MPT modeling is a framework that, in principle, can be applied to measure any type of cognitive processing, provided those processes result in categorical data. This is illustrated with three MPT models that have been developed to study attention and perceptual processes.

Object perception. Ashby, Prinzmetal, Ivry, and Maddox (1996) have proposed an MPT model for feature binding in object perception. When subjects are briefly exposed to two stimuli, they may occasionally report percepts in which the visual features of the stimuli are perceived correctly but combined incorrectly, a phenomenon known as *illusory conjunction* (Treisman & Schmidt, 1982). For example, if two letters of different colors (e.g., a blue C and a red X) are presented quickly, subjects may report "seeing" a blue X. In tasks such as this, subjects can respond with the correct target, a distractor error (based on the other letter in the display), or a nondistractor error (a letter or color not on the current display). These types of responses can be combined for both letter and color features, resulting in a large number of response categories.

One issue in this research area concerns whether illusory conjunctions occur though some specific cognitive mechanism or whether they result purely from guessing errors. Ashby et al. (1996) assert that traditional empirical measures of illusory conjunctions have been fundamentally flawed, precisely because they fail to adequately account for guessing. For this reason, Ashby et al. developed a series of models that incorporated different assumptions about guessing and feature-binding processes. Models that incorporated a specific parameter for feature binding provided the best account of the data, whereas pure guessing models performed poorly. This provided strong, theoretically based evidence for a cognitive basis to illusory conjunctions that previously could not be established using purely empirical measures. Moreover, an MPT model based on Ashby et al.'s location-uncertainty theory provided a better fit to the data than one based on Treisman and Schmidt's (1982) feature-integration theory.

One advantage of the Ashby et al. (1996) model is that it allows researchers to measure feature perception separately from feature binding. This can prove to be useful
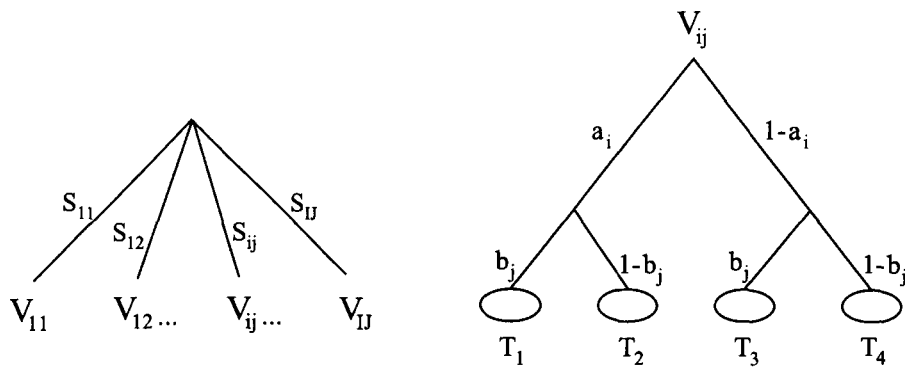
Figure 4. A version of Batchelder and Crowther's (1997) joint MPT representation of the fuzzy logic model of perception.

in research on illusory conjunctions because it allows investigators to measure the effect of experimental variables more precisely for each process. For example, Ashby et al. found that increasing the interstimulus distance reduced correct feature detection and created more feature-binding errors. Prinzmetal, Henderson, and Ivry (1995) observed that lengthening exposure durations and reducing attentional demands had no significant effect on feature detection and still resulted in illusory conjunction errors (as measured by the feature-binding parameter). This is an important result because it demonstrates that illusory conjunctions can be obtained without extremely brief exposures and attention-demanding tasks.

Speech perception. Batchelder and Crowther (1997) have recently applied MPT modeling to the area of speech perception. In one of their experiments, synthetic speech stimuli were generated by combining two acoustic factors: 9 levels of vowel duration with 3 levels of the $F1$ offset frequency of the vowel, creating 27 stimuli, $<i,j>$, $1 \leq i \leq 9$, $1 \leq j \leq 3$. Subjects listened to each stimulus 30 times and were required to classify each stimulus as one of four syllables: [b Λ k] ("buck"), [bak] ("bock"), [b Λ g] ("bug"), and [bag] ("bog"). This specific experimental task is an example of a more general paradigm called *factorial categorization*. In factorial-categorization experiments subjects classify stimuli into a small set of response categories. Each stimulus factor has a finite number of levels, and the stimuli are obtained by combining all possible ways of conjoining one level from each factor. Many areas of cognitive psychology employ such paradigms, including concept identification, pattern recognition, and speech perception (N. H. Anderson, 1990; Ashby, 1992; Trabasso & Bower, 1964).

Batchelder and Crowther (1997) constructed a nested family of MPT models for the special case of two factors and four responses, in which the most specific model was equivalent to a version of Massaro and Oden's fuzzy logic model of perception (FLMP; e.g., Massaro, 1987). In particular, Crowther, Batchelder, and Hu (1995) showed

that the FLMP for any four-response, two-factor paradigm is equivalent to an MPT model depicted in Figure 4 (i.e., it generates identical parametric probability functions). The model for the experiment above has a total of 12 parameters, a_i, b_j for $1 \leq i \leq 9$, $1 \leq j \leq 3$; because there are four response categories (and thus 3 df) per stimulus, it is a joint MPT model with 69 df [(3 × 27) − 12]. The left panel of Figure 4 depicts its 27 separate trees, V_{ij}, corresponding to each stimulus; the right panel depicts the processing tree for stimulus $<i,j>$, where the T_k ($k = 1,2,3,4$) correspond to the four syllable responses discussed earlier. The idea behind the MPT representation is that each factor-level defines its own scale value, and the decision process on conjointly defined stimuli depends on the joint presence or absence of two latent cognitive events—namely, support for vowel [Λ] or [a] ("u" or "o") and support for final consonant [k] or [g]. In Batchelder and Crowther (1997), the model in Figure 4 was a special case of a more general model, called the *conditional independence model*, where the a_i and b_j in the right panel of Figure 4 were replaced with a_{ij} and b_{ij}, allowing both factor levels to influence the probability of each latent event. In this case, the joint MPT model has 54 parameters (the a_{ij} and b_{ij}) and 27 df (81 − 54). Interestingly, the data strongly supported the conditional independence model but not the FLMP in Figure 4.

Object identification. A recent study by Brown (1998) has addressed the issue of whether object identification in a multiobject array is a parallel or a sequential process. The purpose was to test a strong sequential assumption made by LaBerge and Brown (1989) that only one object at a time in a visual scene is identified by the visual attention system. In Brown's paradigm, observers were given a fixed set of target letters and then tried to identify the number of target letters in a series of horizontal two-letter displays. The duration and luminance of each display were varied, as was the distance between the two letters. There were four possible stimulus arrangements for each two-letter display (two targets, one target on left, one

target on right, no targets) and, in turn, four corresponding responses that observers could make. This created a rich data structure represented by a 4 × 4 item-by-response table of category frequencies.

Brown (1997) developed and tested four MPT models, two with a parallel-processing interpretation and two with a sequential-processing interpretation. The parallel models (in Brown's terminology) satisfied the condition that the order of processes could be interchanged in the tree without affecting the category probabilities. In contrast, the sequential models had a fixed order to the processes in the tree. Brown concluded that the best-fitting model was a parallel version, in which observers may "chunk" the letters and process the stimuli as a whole or process the letters independently in parallel if the chunking process fails. It is encouraging that different plausible accounts of object recognition can be sharply differentiated when formulated as MPT models applied to the entire data structure.

Models of Reasoning

A number of standard tasks exist for the study of logical reasoning abilities, working with forms such as syllogisms, the propositional calculus, analogies, and so on. Performance on these tasks is often measured simply in terms of whether or not the logically correct response is given. However, restricting the data analysis to correct versus incorrect responding ignores a wide range of different response errors that can be made to these problems. A more careful analysis of response errors can potentially reveal important insights into the underlying reasoning processes that produce them. As can be seen in the examples to follow, MPT modeling provides a useful tool for accomplishing this.

Wason card-selection task. The card-selection task is a method developed by Wason (1966) for studying propositional reasoning. In this paradigm, subjects are presented with a rule in the form *if p then q*, followed by four cards. Subjects are told that each card depicts the status of *p* on one side and *q* on the other, but subjects can see only one side of each card. The four possible card sides represent *p*, *not p*, *q*, and *not q*, and the subject's task is to indicate which cards need to be turned over to verify if the rule is true or false. Because subjects can choose any combination of cards to turn over, there are 2^4, or 16, possible responses to this task (the correct response being the choice of cards *p* and *not q*). However, as Evans (1977) pointed out, most studies using this task fail to report their results in such detail, and some report only whether performance is correct or incorrect.

Evans (1977) also noted that formal mathematical modeling is rare in the area of thinking and reasoning, and he set as his goal to develop a simple stochastic model for the card-selection task based on the assumption that subjects' choices for each of the four cards are made independently, although with different marginal probabilities. The parameters of the model represent the probabilities

of making responses based on logical reasoning versus nonlogical response tendencies. In generating data for this model, Evans not only considered the selection or nonselection of each card but also included rules other than *if p then q*, created by adding negative statements (e.g., *if not p then not q*). This generates a large set of response categories for developing and interpreting the model. Although not presented in tree form, Evans' model is in fact an MPT model; specifically, it is a joint MPT model, because 16 separate trees can be generated for the selection of the four cards across four different types of rules.

Evans (1977) used the model to reanalyze data from an experiment by Evans and Lynch (1973) and found that the independence model did a good job of accounting for the data with logically interpretable parameter values. In a follow-up study, Krauth (1982) reformulated the model, adding parameters for specific cognitive processes. These included parameters for the logical operations of falsification and verification and for nonlogical processes, such as matching. Krauth's model revealed that logical and nonlogical processes both occur in the card-selection task, a result that is not directly evident from prior analyses of this paradigm.

Klauer and Oberauer (1995). A more comprehensive MPT modeling approach for studying propositional reasoning has been developed by Klauer and Oberauer. In their task, subjects read a major premise involving propositions *p* and *q* and then a minor premise giving the state of one of these two propositions (*true* or *false*). Subjects had to categorize the state of the other proposition ("true" or "false") if it could be inferred and otherwise select a "nothing follows" response. There were four major premise types, corresponding not only to the conditional form used in the Wason (1966) card-selection task but also exclusive disjunction, inclusive disjunction, and biconditional. When these were coupled with the four forms of the minor premise and the three response categories, the data structure was a 16 × 3 product multinomial, with 32 *df* for modeling. Klauer and Oberauer constructed a joint MPT model based on a general theory of propositional reasoning proposed by Johnson-Laird, Byrne, and Schaeken (1992). The model included cognitive processes reflecting "reasoning difficulty," "inconsistent interpretation" (e.g., reading a conditional as a biconditional), and response bias. The model fit the data well and allowed for the separate measurement of parameters for these three cognitive processes. Furthermore, the rank ordering of the reasoning-difficulty parameters was consistent with that predicted by the Johnson-Laird et al. theory.

Rips's (1983) ANDS model. One of the most elaborate information-processing theories for propositional reasoning has been developed by Rips, called ANDS (a natural deductive system). Rips's model can handle a wider set of inferences than those of Wason (1966) and Klauer and Oberauer (1995), and it has been implemented in a computer program in the LISP programming language. The model makes assumptions about memory and con-

trol processes and couples these with mental inference rules. It also constructs proofs for a variety of inferences, and Rips found the model's proofs to be similar to subjects' proofs in many settings.

In one application, Rips (1983) decided to make ANDS probabilistic and fit the proportion of valid responses to a set of 32 inference problems. These problems were created by using different combinations of qualifiers such as *and, not, or, if . . . then*, and so on (see Rips, 1983, Table 4). The data structure was a product binomial with 32 problems and two response categories for each problem ("valid" or "invalid"), thus creating 32 *df*. The model was made probabilistic by postulating that each of 12 inference rules may or may not be available for a given problem. To simplify the calculations, Rips assumed that each rule had its own fixed probability of availability and that rule availability is independent. With these assumptions, explicit expressions of the response probabilities for each problem were derived in terms of the 12 availability parameters, plus one guessing parameter, and the data were fit using the STEPIT program (Chandler, 1969) with a least squares criterion. The fit was judged to be satisfactory.

With the above assumptions, ANDS can be viewed as a joint MPT model with 32 trees, where each branch reflects a sequence of available and unavailable reference rules. In fact, a reanalysis of the data within the MPT framework would provide statistically interpretable fit measures and confidence intervals for the estimates. What we think is interesting in this case is that a very detailed information-processing model can be reduced to MPT form with a few simplifying assumptions, such as rule independence and availability probabilities that are identical over problems. These assumptions are, at best, approximations; however, the gain in statistical analysis may compensate for the degree of simplification. In any event, analyses like the one Rips (1983) provides supplements the analyses of the information-processing model in a productive way.

Class inclusion. Piaget (1952) has developed a number of simple tasks to study reasoning abilities in children. One set of tasks examines children's understanding of the concept of class inclusion. For example, a child may be informed that there are five black dogs and three orange cats and then asked the question, "Are there more dogs or more animals?" Rabinowitz, Howe, and Lawrence (1989) have developed a model that satisfies the MPT constraints and is capable of separately measuring memory and reasoning processes in this task. They extend Piaget's standard paradigm by asking additional questions concerning the subclasses (e.g., "Are there more dogs than cats?") plus a series of more complicated questions about the color of the class members (e.g., "Is the blackest dog the same color as the blackest animal?"). Altogether, there are 11 different questions that can be asked in these formats. Moreover, responses can be categorized into successes ("more animals") and two types of errors

("more dogs" or "same number"). This generates a fairly rich set of response categories for MPT modeling.

Rabinowitz et al. (1989) did not explicitly represent their model in processing-tree form, but, in essence, their model can be constructed as a series of joint MPT models when applied to each different question. Their two main goals were to assess the relative contributions of memory and reasoning processes in the class-inclusion task and to examine how these processes change across the life span. The results of their analysis showed that children were poorer than adults on both memory and reasoning abilities. In fact, 7- and 10-year-old children exhibited no understanding at all of class-inclusion logic, on the basis of the finding that the model parameter measuring that process was estimated to be zero. This conclusion, based on the model's analysis, is the type of result that is often difficult to extract from ad hoc statistics because they are typically confounded by guessing and response-bias factors.

In a more extensive study, Howe and Rabinowitz (1996) found that the main developmental trend is the emergence of more sophisticated reasoning processes with the decline of less sophisticated strategies. Howe and Rabinowitz also manipulated memory load (having the original statement available or not at the time of questioning) and information load (one or two dimensions per statement), and these manipulations affected both memory and reasoning parameters. From this, they concluded that memory and reasoning processes are not independent, contrary to some earlier theories.

Psychometric Models

An area of psychology that has seen great use of probabilistic models is psychometrics. Psychometric models, unlike most models in cognitive psychology, generally involve individual-difference parameters and item parameters. One large subarea of psychometric modeling is test theory, in which the data structure consists of the performance of a set of subjects who each answer the same set of questions. Although most of the models in this area are not MPT models, two applications of MPT modeling are described below.

Cultural consensus analysis. In a series of papers, Batchelder and Romney (1986, 1988, 1989; Romney, Batchelder, & Weller, 1987; Romney, Weller, & Batchelder, 1986) developed a set of formal models for information pooling called *cultural consensus analysis*. The paradigm involves a set of respondents, each answering the same set of objective questions. Neither the answers to the questions nor the relative knowledge of the respondents is known a priori. This is the case for many situations in the social sciences, especially cultural anthropology, in which researchers attempt to learn about a cultural group by asking questions to members of the group.

In one of the models designed for multiple-choice questions, each question is assumed to have a single correct answer, and each respondent is assumed to have a com-

petency parameter that measures the probability that the respondent "knows" the correct answer to any particular question. The model allows the researcher to estimate both the competency parameters for each respondent and the correct answer to each item, by using information from the question–response matrix. Klauer and Batchelder (1996) showed that the multiple-choice model described above can be formulated as an MPT model. More specifically, they explored the two-respondent case and applied it to the problem of interjudge reliability in subjective categorization. The model extends previous work on this issue by explicitly postulating and measuring category response biases and interjudge reliability.

In another application, Batchelder, Kumbasar, and Boyd (1997) have constructed a consensus-analysis MPT model for sociometric measurement. The paradigm requires each member of a social network to provide information about the friendship ties between every pair of members in the network. If the network has N members, then the data consist of N friendship digraphs on the same set of N nodes, one from each member. The model permits one to estimate a "consensus" digraph of the network, as well as competency and response-bias parameters for each member of the network. Interestingly, these member-competency parameters can vary from region to region in the digraph, reflecting the fact that one's social knowledge and biases can depend on how close one is tied to other members of the social network.

Item-response theory. A popular psychometric approach to test theory is item-response theory, where models postulate subject-ability parameters and item-characteristic parameters (see Baker, 1992; Fischer & Molenaar, 1995). The models are applied to a subject × item matrix, in which each entry is a measure of the performance of a particular subject to a particular item. In most cases, models of this type are designed for categorical data but, in general, are not MPT models (e.g., the Rasch, 1960, model discussed earlier). One exception are the finite-state models of Hutchinson (1982) and their elaboration by Garcia-Perez (1987, 1989, 1990; Garcia-Perez & Frary, 1991a, 1991b). In one version of these models, the test taker is assumed to know the status of each alternative in a multiple-choice test independently with probability λ, $0 \leq \lambda \leq 1$. For an item with, say, three response alternatives (only one of which is correct), the test taker can know the status of k of the items with probability $p_k = \binom{3}{k} \lambda^k (1 - \lambda)^{3-k}$, for $k = 0,1,2,3$. The model takes into account whether the correct answer is among the known ones, and the model includes guessing parameters if it is not.

In another application, Garcia-Perez (1993) constructed an MPT model for the case in which the last option is "none of the above." He used the model to compare the confidence-interval estimates of the test taker's abilities for the case of "none-of-the-above" items compared with more conventional items involving an equal number of options. Interestingly, Garcia-Perez's (1993) analysis using the MPT model showed that a "none-of-the-above"

option can reduce the size of the confidence interval of λ, which is a different conclusion from some earlier work claiming that there is no statistical advantage to including such items.

THEORETICAL ISSUES

In this section, we address a number of theoretical issues relevant to the application of MPT modeling. We start by outlining some basic considerations that are important in the development and testing of new models. Following this, we address a series of topics that have generated some recent debate, including high-threshold and discrete-state assumptions behind MPT models and their role as approximations to more complete psychological theory. Finally, we compare MPT modeling with other general classes of mathematical models.

Model Development

It is generally the case that given a category system C_1, C_2, ..., C_J, many potential MPT models can be developed. There is no algorithm for creating models; however, there are a few useful heuristics. If one wishes to develop an MPT model for a particular research paradigm, three basic steps must be accomplished. First, a category system of possible response events must be established. Second, it is necessary to identify which cognitive processes are involved in the paradigm. Third, one must specify how these processes lead to each of the response categories (i.e., the construction of the tree model itself).

Response categories. A crucial step in the success of any MPT model is to define response categories in a way that creates a rich and informative data structure. A detailed, fine-grained analysis of data into psychologically meaningful categories provides more information with which to extract the contribution of different cognitive processes. Also, the more response categories that can be meaningfully identified, the more degrees of freedom that can be created for estimating parameters and testing the fit of the model. If there are J categories, then an identifiable model can have at most $J - 1$ parameters. This poses quite a constraint on the modeler. If one is interested in measuring a particular cognitive capacity, other things being equal, it makes sense to seek relevant paradigms in which there are many more degrees of freedom to permit more model parameters.

One solution to this problem is to select paradigms involving more than one category system. Sometimes, each separate category system can be represented by a different tree, creating a joint MPT model. The advantage of this strategy is that if certain cognitive processes operate in more than one tree, it may be possible to capture these multiple processes using a single parameter. This, in turn, keeps the number of parameters to a minimum, creating more degrees of freedom.

Fortunately, there are several standard ways to develop research paradigms with multiple category systems. One common method is to investigate research situations that

involve more than one type of stimulus item. The source-monitoring model in Figure 2 is a good example, because separate trees are generated for Source A, Source B, and New items. The fact that the response-bias parameters (a, g, and b) occur in more than one tree helps to reduce the total number of parameters needed for the model. Other examples of joint MPT models based on multiple stimuli include our pair-clustering model in Figure 1, which examines the recall of clusterable pairs and singleton items and creates separate trees for each. The various models of reasoning reviewed earlier also fall into this category. For example, Rabinowitz et al. (1989) expanded the standard Piagetian paradigm by asking their subjects multiple questions about class inclusion, creating a different tree structure for each equation. Evans (1977) and Krauth (1982) created multiple items by including both positive and negative propositional statements, whereas Rips (1983) created 32 inference problems by combining different logical arguments.

A special case involving multiple items is the factorial categorization paradigm described earlier in the section on speech perception. For example, in a two-factor experiment with I levels of one factor and J levels of the other, there are $I \times J$ factorially generated stimuli, with only $I + J$ factor levels. If a subject repeatedly classifies each stimulus into one of K categories, there are $I \times J \times (K - 1)$ degrees of freedom in the data structure. If a modeler can assign individual parameters to each of the $I + J$ levels and postulate a simple combining rule (as Batchelder & Crowther, 1997, did with the FLMP), then there may be a large surplus of degrees of freedom.

Using paradigms with multiple category systems is one method for creating a large set of response categories for MPT modeling. However, there are other techniques that can be used for this purpose that do not necessarily lead to the creation of a joint MPT model. For example, many of the memory models reviewed earlier create a sufficient number of response categories by giving subjects multiple tests of their memory. Howe's (1991, 1995) model and Kail et al.'s (1984) model achieved this by using repeated recall tests of the same material. Other models have examined memory performance using a combination of different testing procedures. For example, Riefer and Rouder's (1992) model examines both free recall and cued recall, whereas Chechile's (1987) model looks at both recall and recognition along with confidence judgments.

Another useful heuristic is to examine stimuli that are multidimensional (i.e., items that have more than one attribute or characteristic). If subjects' responses are classified according to these multiple characteristics, this can lead to a potentially large set of response categories for model development. The pair-clustering model in Figure 1 is one example of this, in that recall of pair clusters is examined both in terms of the number of items recalled and in terms of whether the items are recalled adjacently. Along similar lines, the B. H. Ross and Bower's (1981) model of associate recall examines memory sets

containing multiple items and probes these sets with one or more cues. Brown's (1998) paradigm varies the number and positioning of targets in a two-letter display, thus creating different display arrangements with multiple response categories. Perhaps the most prototypical example of this approach is the model of object perception by Ashby et al. (1996). In their paradigm, items are classified by such attributes as letter, color, and location, and Ashby et al. were able to generate a large set of response categories from this classification system.

Cognitive processes. Given a sufficiently rich data structure, the next step is to identify and parameterize the cognitive processes contributing to those data. One can usually imagine a large number of psychological processes that may be operating in concert to produce categorical data in a given research paradigm. Some of these processes will be more important than others and, thus, will explain more of the variance in the data. However, it is often the case that even if less central processes are eliminated from the model, there may still remain more processes than categories. The consequence of this is that if the model includes all of these processes, then it will not be globally identifiable as defined earlier.

In dealing with this problem, one useful heuristic is to build a more complete but nonidentifiable model and then create a series of nested submodels based on certain restrictions on the parameters. This will result in a family of identifiable models. For example, we (Batchelder & Riefer, 1990) were able to create identifiable versions of the source-monitoring model in Figure 2 by equating detection ($D_1 = D_2$), discrimination ($d_1 = d_2$) or response bias parameters ($a = g$). Other examples of this strategy can be found in Rouder and Batchelder (1998), Hulme et al. (1997), and Bender et al. (1996). In particular, Bender et al. created nested submodels by collecting data on three different age groups of children. They were able to test the goodness of fit for different submodels by assuming that various parameters were constant across the three groups. Of course, when a model's parameters are restricted in this way, any comparisons between experimental conditions must be made under the same set of restrictions. This has the potential to limit theoretical conclusions, depending on which parameters are constrained. However, one can usually place restrictions on ancillary parameters, leaving crucial parameters free to vary for comparisons across conditions.

In general, the process of model building certainly involves a compromise between the goals of having an incomplete but identifiable model, which permits unique measurement of the model's parameters, and a more complete and psychologically valid model, which represents all of the relevant cognitive processes. Each MPT model is at best an approximation to a complete process description of categorical data, and the task of the modeler is to select the most important processes and capture them in a valid way.

Tree structure. There are, of course, many MPT models that could be constructed for a given categorical data

structure. Unlike general linear models that routinely decompose data into main effects and interactions, there is no algorithm for generating useful processing trees or even for discovering if they exist for a given set of data. MPT model building for a particular paradigm is a creative process that requires a researcher to commit to certain processes and how they are arranged structurally in a conditional processing tree. Furthermore, because there are many possible tree structures, there will often be psychologically uninterpretable MPT models that nevertheless fit a given set of data well. Thus, the process of developing a valid model requires that one fit a number of data sets in the same paradigm and that the resulting parameter estimates be interpretable in terms of the underlying processing assumptions.

As a consequence, a key question in the development of an MPT model is whether the model's parameters are, in fact, valid measures of their respective cognitive capacities. Ideally, a parameter designed to measure, say, storage capacity should reflect only storage factors and should not be influenced by the operation of other cognitive processes. Therefore, in the course of developing and testing a new MPT model, it is necessary to conduct validity testing of the model's parameters. This typically involves a series of experiments in which basic independent variables are manipulated and shown to have a selective influence on certain processes that is interpretable on theoretical or logical grounds. In memory research, for example, it is relatively straightforward to find variables that should have their primary influence on storage and not retrieval, or vice versa. In general, validity testing of a model's parameters is essential if one wishes to have confidence in an MPT model as a valid measurement tool. If validity tests reveal that parameter values are affected by variables in a psychologically implausible manner, it is at least an indication that the current tree structure of the model is wrong or, more seriously, that the basic assumptions behind the model are incorrect.

Numerous examples of validity testing can be found in the applications of MPT modeling reviewed earlier. Both the storage–retrieval models of Chechile and Howe have been applied to a number of data sets examining basic experimental manipulations. Riefer and Batchelder (1995, pp. 617–620) applied the model in Figure 3 to a large corpus of experiments on recognition failure and generally found the model's parameters to behave logically as a function of many experimental variables. Bayen et al. (1996) manipulated item similarity and source similarity in a factorial design to test different models of source memory and concluded that only a 2HT model adequately accounted for their data. Buchner et al. (1995) validated their MPT model for the process-dissociation procedure by showing that experimental manipulations designed to effect response bias did, in fact, have a selective influence on the response-bias parameters in their model. Prinzmetal et al. (1995) applied their feature-binding model to experiments that manipulated stimulus duration, interstimulus distance, and levels of attention,

and they observed that the parameter values were affected in a theoretically interpretable manner as a function of these manipulations.

Another important facet of validity is to connect the assumptions of an MPT model for a particular paradigm to those of a more complete and accepted formal theory, if one exists. In the best of worlds, the MPT model will be a special case, or a provable mathematical approximation, of the more complete theory. One example of this type of connection is our model of the recognition-failure paradigm (Riefer & Batchelder, 1995) in Figure 3, which is a special case of the retrieval-independence theory of Flexser and Tulving (1978). A second example is the pair-clustering model in Figure 1, which is mathematically related to Markov models of pair clustering in multitrial free recall (Batchelder & Riefer, 1980; Bäuml, 1996a). Of course, it is not reasonable to expect that a tight, mathematical connection can always be drawn between an MPT model and a more complete theory for any given phenomenon. In such cases, it is still important that there be a connection at some level between the processes in the MPT model and those of the more complex theory. For example, we have seen this in Klauer and Oberauer's (1995) model for propositional reasoning, which was directly motivated by the more general theory of propositional reasoning by Johnson-Laird et al. (1992). The probabilistic version of Rips's (1983) information-processing ANDS model is another example of a close connection between an MPT model and a more elaborate processing theory.

Statistical Issues

Earlier, we discussed the statistical inference for MPT models, as described in Hu and Batchelder (1994b). This approach is based on the classical theory of inference that utilizes asymptotic approximations requiring large data samples (e.g., Read & Cressie, 1988). The idea is that as sample size increases, and if a model is true, then, under rather mild mathematical conditions, goodness-of-fit measures such as G^2 or minimum chi-square discussed earlier have approximate chi-square distributions when minimized over the parameter space. This is the idea behind the efficiency of MLEs, confidence intervals of the estimators based on the observed Fisher information matrix, and likelihood ratio tests of nested models. Thus, if the sample is suitably large, then goodness of fit and hypothesis tests can be conducted with reference to standard chi-square tables.

However, there is good reason to doubt that any statistical model is "true" in the sense that the observations are independent and identically distributed and with a probability distribution exactly consistent with the model. Furthermore, one often has to deal with situations in which the sample size is smaller than desired for asymptotic approximations. It is natural to consider some of these issues and how to deal with them. The next subsections take up the use of computer simulations to supplement asymptotic inference theory, how to deal with approximate

models, and whether the MPT model framework is falsifiable in principle.

Computer simulation. A useful technique in the statistical analysis of an MPT model is to conduct computer simulations. Basically, this technique involves setting the values of the model's parameters and then conducting a series of simulated runs using various sample sizes. Parameter estimates, confidence intervals, and hypothesis tests from these simulations can be compared with the results of asymptotic estimates based on standard statistical theory. A number of important issues can be addressed in this way, such as the effect of small sample sizes or the amount of error (bias) in parameter estimates. Another use of computer simulation is to introduce small extensions of the model, such as allowing individual differences in the parameters (discussed in detail in the next subsection) or adding an additional cognitive process. The ability of the original model to account for data from these extensions is a measure of the robustness of the model. Riefer and Batchelder (1991b) provide a detailed illustration of how these techniques can be applied to the pair-clustering model in Figure 1. Other examples of using computer simulations to explore MPT models include Bender et al. (1996, p. 194), Erdfelder and Buchner (1998a), Riefer and Batchelder (1988, p. 328), and Riefer and Rouder (1992, p. 604).

What is especially beneficial about computer simulations is that they can be used to explore the sample size necessary to achieve sufficient statistical power when using a particular MPT model. Of course, the issue of power and sample size is an important one in research generally, but it is particularly relevant to MPT modeling because, as we discuss later, MPT models are simple and approximate models of cognitive processing. Large samples sizes may be necessary to reveal significant differences between groups in terms of the model's parameters; however, sample sizes that are too large also inflate goodness-of-fit statistics. Computer simulations can be used to explore the potentially complicated relationship between sample size, power, and choice of Type I error probability for any particular MPT model. A very useful approach to assessing power in MPT models has been developed by Erdfelder, Faul, and Buchner (1995) based in part on the extensive work by Cohen (1988, 1992). Their computer program, called GPOWER, performs power analysis for a range of common statistical tests and can be used to establish significance levels when applied to MPT modeling (e.g., Buchner, Erdfelder, et al., 1997, p. 512). In several of these applications of GPOWER, it is shown that sufficient power to test hypotheses about the parameters of a model can be achieved using much lower levels of Type I error for goodness-of-fit tests than the conventional $\alpha = .05$. This approach is especially reasonable when the model has already been shown to fit data in prior studies.

Approximation. Because of the requirement on a useful MPT model to be globally identifiable, the number of categories places an upper bound on the number of parameters an MPT model can have. As we have stated earlier, unless the paradigm provides many categories, only an incomplete processing account is possible. In this case, it is important to select processes that control most of the variance in category responses and then to regard the model as approximate. However, even if a model is only approximate, there is reason to believe that estimates of the parameters may provide measurements that are valid and useful. More technically, most MPT models, especially ones that have a surplus of degrees of freedom, are "misspecified." A misspecified statistical model is one that cannot exactly model the "true" probability distribution over the categories (i.e., the probability distribution is not in the set that is generated by varying the model's parameters). Most stochastic modelers would readily agree that all statistical models with a surplus of degrees of freedom (technically called *unsaturated*) are misspecified, and this remains true for more complex and detailed accounts, such as neural network or connectionist models (J. A. Anderson, 1995) or global memory models (Clark & Gronlund, 1996).

There is a technical theory of statistical inference with misspecified models that is well developed in statistics and econometrics (e.g., White, 1994); however, this work has yet to make an impact in psychological modeling (see Golden, 1995). White (1994) presents the theory of misspecification from the standpoint of classical statistics, as opposed to a Bayesian approach, and his early work (e.g., White, 1982) shows that MLEs computed from misspecified models, called *pseudo-MLEs*, nevertheless have useful asymptotic properties under mild regularity conditions.

One way that an MPT model may be misspecified is that there may be parameter heterogeneity underlying the observations. In this case, even if the processing account is accurate, the multinomial assumption that observations are identically distributed over the categories is violated. In the past, there have been many concerns raised in the cognitive modeling area about parameter heterogeneity over subjects, items, and even subject–item interactions (e.g., Curran & Hintzman, 1995; Hintzman, 1980, 1992, 1993). However, there have been only a few efforts to productively incorporate parameter heterogeneity into cognitive models (e.g., Batchelder, 1975; Offir, 1972; Riefer & Batchelder, 1991b).

Fortunately, Riefer and Batchelder (1991b) describe a straightforward way to augment an MPT model to handle parameter heterogeneity. The idea is to assume that each parameter is a random variable that has a marginal beta distribution across the experimental observations. More specifically, suppose θ_s is a parameter of an MPT model. The beta distribution on θ_s takes the form

$$f(\theta_s) = \frac{\Gamma(\alpha_s + \beta_s)}{\Gamma(\alpha_s)\Gamma(\beta_s)^2} \theta_s^{\alpha_s-1}(1-\theta_s)^{\beta_s-1}, \quad 0 < \theta_s < 1$$
$$0 \qquad\qquad\qquad\qquad , \text{ otherwise,}$$

where α_s, $\beta_s > 0$. The beta distribution has mean, $E(\theta_s) = \alpha_s/[\alpha_s + \beta_s]$, and variance

$$\text{Var}\,(\theta_s) = \frac{\alpha_s \beta_s}{(\alpha_s + \beta_s + 1)(\alpha_s + \beta_s)^2}. \qquad (7)$$

The idea in Riefer and Batchelder (1991b) is to assume independence of the parameters, but with marginal beta distributions. Observations are then produced from the MPT model conditional on the value of the parameter vector, creating a conditional MPT model. This extension of the MPT model family is explicitly analyzed in Hu and Batchelder (1994a). The idea is well known in stochastic modeling (e.g., the conditional Poisson Process; S. M. Ross, 1983), and it was used for adding individual differences to the all-or-none and linear operator learning models in Batchelder (1975).

The approach of creating a conditional MPT model has the effect of doubling the number of parameters, in that each θ_s yields α_s and β_s for the beta distribution. Sometimes, when there are many degrees of freedom remaining after modeling, specifically $S \le (J - 1)/2$, the resulting conditional MPT model may be an identifiable model, and it can be analyzed in the usual way. But, in practice, it is more usual that $S > (J - 1)/2$; thus, this approach yields a nonidentifiable model. Nevertheless, there is a useful strategy in data analysis that can still be accomplished. The recommended approach, illustrated in Riefer and Batchelder (1991b), is to first analyze the empirical data with the identifiable MPT model obtaining, among other things, the MLE $\hat{\theta}_s$ for each parameter. Next, select α_s and β_s, so that the mean $\alpha_s/[\alpha_s + \beta_s]$ is approximately equal to $\hat{\theta}_s$, and the variance in Equation 7 is set to some prespecified level. Finally, simulate many data sets, each the size of the original, from the resulting conditional MPT model. By examining goodness of fit, MLEs, and confidence intervals, one can explore the robustness of the original MPT model against controlled amounts of parameter heterogeneity. This strategy is straightforward and computationally fast on the current generation of personal computers, and Xiangen Hu's MPT software program described earlier incorporates this option.

Analysis of several MPT models by our group (e.g., Riefer & Batchelder, 1991b; Riefer et al., 1994; Rouder & Batchelder, 1998) has led to some optimism that MPT models may be rather robust under realistic levels of parameter heterogeneity. In particular, the MLEs are often quite robust under individual-difference assumptions; however, the size of the asymptotic confidence intervals is more sensitive to individual differences. Also, it is clear that robustness may not hold for MPT models that contain many levels of the tree structure. To see this, suppose there are large individual differences in overall processing ability in a subject pool. To reach any particular node near the terminal node of the tree requires a sequence of conditional processing steps. Because some nodes are reached through accurate processing and others through imperfect processing or even guessing, the structure will tend to separate or select good processors (i.e., ones with high values of the process parameters) from poor processors as conditional processing nodes are successively encountered in the tree. A consequence of this is that the estimates of the parameters that occur only near the terminal nodes of the tree may be severely biased one way or the other by subject selection.

Falsifiability. The fact that many MPT models can be developed for a given set of categories raises the question of whether the class of MPT models can be falsified with data. Falsifiability is an important issue that has been raised, for example, with connectionist models. It has been proved that the connectionist framework can model any mapping of stimuli into response categories, each represented by one–zero vectors, if no bounds are placed on the model (Rumelhart, McClelland, & the PDP Research Group, 1986, p. 65). However, MPT models are not like connectionist models in this sense because, as we noted earlier in the section on reparameterization, there are many models for categorical data that are not in the MPT class. If one of these models accounts for data in a particular paradigm, then, technically, one can infer that the MPT class is falsified for that paradigm. Of course, it may be possible to design an MPT model that closely mimics or approximates the successful fits of the non-MPT model; thus, it may be difficult to argue that the MPT framework is falsifiable in practice.

On the other hand, an acceptable MPT model must not only be able to fit data but its parameters must be globally identifiable, must be psychologically interpretable, and must pass appropriate validation experiments. When these criteria are imposed, we see no reason to expect that an MPT model satisfying these constraints can always be found for a given paradigm. In fact, we ourselves have tried and failed to adequately model several cognitive paradigms with MPT models, despite some promising initial ideas. Thus, we argue that MPT models are a useful, but by no means a universal, family of cognitively interpretable statistical models.

High-Threshold, All-or-None, and Discrete-State Assumptions

MPT models on occasion have been described as postulating high-threshold or all-or-none assumptions (cf. Kinchla, 1994; Wagenaar & Boer, 1987; Yonelinas & Jacoby, 1996). It is true that a particular model may evidence one of these characteristics, but it is important to realize that MPT models as a general class are neither high-threshold nor all-or-none. The high-threshold (HT) model of yes–no signal detection (e.g., described in Macmillan & Creelman, 1991, chap. 4) can be viewed as a joint MPT, with a separate tree for old and new items. It is "high threshold" in the sense that new distractors are never detected or recognized as "old." Kinchla (1994) describes the source-monitoring model in Figure 2 as an HT model because if old Source A and old Source B responses are collapsed into a single "old" category, then the model reduces to the HT model. However, because subjects are re-

quired to distinguish Source A and Source B responses, Batchelder, Riefer, and Hu (1994) argue that this collapsing may not result in a psychologically meaningful category.

If the term *high threshold* is to be applicable to MPT models in general, a wider definition is needed. For example, it is possible to formally define this term for MPT models as applied to certain types of recognition paradigms. Suppose subjects study a list of various types of items and are then given a recognition test that includes new distractors. Suppose also that the processing tree for new distractors is a subtree of the processing trees for each type of old item, so that when old items are undetected and require guessing, the responses are governed by the distractor processing tree. (This can be seen in Figure 2, in which the tree for new items constitutes a part of the tree for Source A and Source B items.)

The preceding definition clearly classifies the HT model, as well as the source-monitoring model in Figure 2, as having a high-threshold assumption. In the latter case, it provides a logical sense to Kinchla's (1994) claim that the source-monitoring model in Figure 2 is an HT model. However, it seems to us that it would be difficult to capture the essence of this informal definition in a useful and precise way, especially for a wider class of models. In any event, many MPT models for recognition memory do not satisfy this definition. For example, Luce's (1963) low-threshold model and Bayen and Murnane's (1996) source-monitoring model based on the "double-high-threshold" model, are both joint MPT models that fail to satisfy the collapsing condition suggested by this informal definition.

The attribute of all-or-none as applied to MPT models is also difficult to formalize. The term *all-or-none* was popular in the literature on Markov learning models, and it refers to a particular model by Bower (1961) and others, in which subjects either completely master an item on a study trial or learn nothing. In other Markov models, the term all-or-none is more restrictive and refers to the assumption that transitions to the next stages in the learning process occur either completely or not at all. Most Markov learning models for a finite number of learning trials, including the all-or-none model of Bower, belong to the MPT model family. However, outside of repeated-trial learning and memory experiments, the notion of all-or-none as used with Markov models does not apply to MPT models.

One way that the term all-or-none might be applied to an MPT model of memory is the assumption that memory storage is either completely successful or completely unsuccessful. This assumption is embodied in several MPT models (e.g., the model in Figure 3). However, several other models allow for memory storage to be in one or more intermediate states. Examples include the Schweickert (1993) model of STM, in which traces can be intact or degraded, Chechile's (1993) model of fractional storage, Howe's (1991, 1995) model, which allows redintegration of nonstored information, and the pair-

clustering model in Figure 1, which assumes that clusterable pairs can be stored as clusters or as two separate individual traces.

In summary, neither the attribute high-threshold nor the attribute all-or-none can be applied to MPT models in general, and, outside of the learning-model context, they can be reasonably applied only to a few special cases. What, then, is the best way to describe the basic characteristics of MPT models? In our opinion, MPT models are best characterized as a particular class of discrete-state models (cf. Riefer & Batchelder, 1988, pp. 319–320). They are discrete models in the sense that they postulate only a finite number of processing states, characterized by the nodes of the processing trees. The assumption of finitely many discrete processing stages may seem restrictive and incorrect to some theorists, but it seems to us quite plausible, at least as an approximation of more complete processing accounts.

The tension between finite-state and "continuous" processing models is reflected in the literature on signal detection models (e.g., Macmillan & Creelman, 1991). Some theorists (e.g., Yonelinas & Jacoby, 1996) have argued that recognition memory cannot be modeled with a finite-state processing model and instead claim that evidence supports a Gaussian signal detection model that separates hits and false alarms into d' and a response criterion. However, the Gaussian assumption itself is, at best, a convenient approximation, especially with the assumption of equal variances for the signal and noise distributions. Furthermore, many other models, such as ones obtained by substituting a logistic distribution for the Gaussian or even some finite-state models, are practically indistinguishable from each other on recognition-memory data (e.g., Green & Birdsall, 1978).

Despite considerable discussion on this issue (see, e.g., Mulder, Sanders, & van Galen, 1995), we know of no conclusive empirical argument that finite-state models can be ruled out or that cognitive psychology has discovered that processing is continuous and not discrete. Even if the discreteness assumption is found unacceptable in detail, that does not argue against its usefulness as an approximation in modeling. Of course, there may exist arguments against specific models, both finite-state and continuous, but these arguments do not support a generalization to all such models. In much of the theoretical work since the 1980s, continuous processing models have been the dominant variety. The issue, we think, rests with one's goals and has a more pragmatic than scientific rationale. In particular, if the goal is to model a variety of empirical phenomena with the same theoretical system, continuous processing models of global matching, such as CHARM, TODAM, SAM, and MINERVA (see Clark & Gronlund, 1996), or various connectionist models (J. A. Anderson, 1995) have proven the more successful. However, if the goal is to measure processing capacities in precisely defined categorical paradigms, the MPT class of models has demonstrated success in
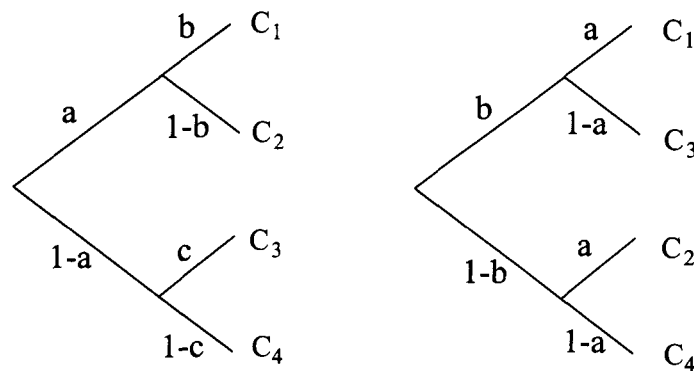
Figure 5. Serial and parallel versions of a simple MPT model. The left panel portrays a serial model in which the *a* process occurs before (and determines) the *b* and *c* processes. The right panel reconstructs the tree under the assumption that $b = c$, with no necessary serial ordering to the *a* process and the *b* process.

part because it has a well-worked-out and computationally simple statistical inference theory. From the measurement point of view, it is essential to have an apparatus for workable statistical inference, and it is in this area that the more complex models have yet to be developed.

Parallel and Serial Processes

An important issue in MPT modeling is the nature of the processing architecture assumed. For example, the branches of an MPT model are designed to represent possible processing pathways that lead to a particular response category. Each link in a branch is interpreted substantially as one stage in a hypothetical sequence of stages that results in a particular response. This interpretation raises the issue of whether these stages are necessarily accomplished serially, as suggested by the directed graph representation of the tree or if one or more processes can occur in parallel.

Actually, it is quite easy to represent parallel processes as well as serial ones in tree form. Consider the simple tree in the left panel of Figure 5, with parameter vector $\Theta = (a,b,c)$ and response categories C_1, C_2, C_3, and C_4. In this representation, the first process governed by parameter a occurs before the ones governed by b and c, because the result of the first process determines which of the other processes operates. But if we assume that $b = c$, then the tree can be restructured to create the tree in the right panel of Figure 5. It should be easy to see, under this assumption, that the two MPT models in Figure 5 are equivalent models. For this reason, the model with $b = c$ can be thought of as representing a case in which there is no necessary serial ordering of the *a* process and the *b* process, perhaps a case of parallel processing. In general, if there is a node in a tree where each link from the node leads to an identically parameterized node, then one can restructure the model by reversing the two processes as in Figure 5. In this case, an MPT model is necessarily neutral concerning whether two underlying processes occur in series or in parallel. This observation led Ashby et al. (1996, p. 170) to conclude from their models

of object perception that it was not possible to order the processes of perceiving a target's form from perceiving its color. This is also the logic that Brown (1998) used to differentiate parallel versus serial models of object identification, discussed earlier.

A related approach to representing sequential versus nonsequential processes in MPT models can be found in the model of STM by Schweickert (1993), discussed earlier. In one version of this model, reconstruction of the short-term trace can occur through either phonological processes (with probability S) or lexical processes (probability L). The question is whether these phonological and lexical processes occur sequentially. Figure 6 presents two versions of this model, with a nonsequential version on the left and a sequential (or serial) version on the right. The sequential version assumes that lexical processes occur only after successful completion of phonological processes. The nonsequential version assumes that correct responding is determined by lexical processes with probability A or by semantic processes with probability $1 - A$. Thus, only one of these two processes is instrumental in determining any given response. Schweickert has shown that the two models in Figure 6 make qualitatively different predictions in specially designed factorial experiments.

Comparisons to Other Mathematical Models

As we have indicated above, MPT models share some of the goals of psychometric measurement models. For example, consider the areas of paired-comparison scaling (David, 1988) or item-response models of testing (Baker, 1992). In both of these paradigms, the goal is to estimate parameters that reflect latent factors underlying the manifest data, and these models typically have many parameters. In the case of paired-comparison scaling, the model of Bradley, Terry, and Luce (see Luce, 1959) defines a nonnegative parameter v_i for each choice object i, and the model postulates that the paired comparison probabilities are given by $p_{ij} = v_i/(v_i + v_j)$, where p_{ij} is the probability that object i is chosen over object j. If there are N
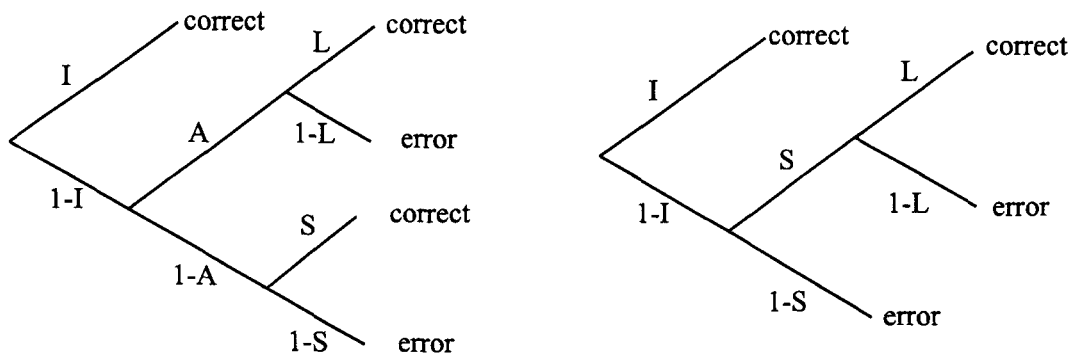
Figure 6. Two versions of Schweickert's (1993) MPT model for short-term memory. The left panel depicts a version in which the lexical process (L) and phonological process (S) occur nonsequentially. The right panel depicts a sequential version in which the phonological process S occurs before the lexical process L.

choice objects, then there are $N(N - 1)/2$ choice probabilities p_{ij} and N parameters (actually $N - 1$ identifiable parameters). Despite the large number of parameters, nothing about the detailed cognitive processes of choice is modeled, only the basic fact that some objects are more attractive than others.

A similar situation can be found in models of item-response theory. In this paradigm, N examinees each takes a test consisting of M items, and the data consist of an $N \times M$ matrix, where the ij term is given by

$$X_{ij} = \begin{cases} 1 & \text{if subject } i \text{ is correct on item } j \\ 0 & \text{otherwise.} \end{cases}$$

One very highly studied model for this situation is the Rasch (1960) model. In one version of this model, there are N subject-ability parameters (s_i) and M item-difficulty parameters (d_j), and the model postulates that

$$P(X_{ij} = 1) = \frac{s_i(1-d_j)}{s_i(1-d_j) + (1-s_i)d_j},$$

where $0 \le s_i, d_j \le 1$. Again, the purpose of the Rasch model is not to account for the complex cognitive processes that subjects use to answer questions. Instead, it is a measurement model with only the most rudimentary psychological assumptions behind it—namely, that there are individual differences at both the subject-ability and item-difficulty levels and that ability and difficulty are unidimensional and trade off in a simple way.

There is a vast literature concerning the statistical theory behind both the Bradley–Terry–Luce model and the Rasch model, including statistical inference issues such as parameter estimation, goodness of fit, and robustness. The reason why we have described these two psychometric models in some detail is to contrast them to most of the modeling work in cognitive psychology, in which complex models are created that are rich in psychological assumptions. Yet, for these complex models, it is often the case that very little is known about the statistical inference

issues that are relevant to using the parameter estimates as measurements of underlying or latent processes.

Our pair-clustering model in Figure 1 provides a good illustration of the use of MPT models as psychometric tools and how they contrast with more complex cognitive models. Note that the tree for singletons merely models free recall as a Bernoulli process, in order to complete the model that focuses on item pairs. It provides no insight into the actual processing stages for single item recall. Our model is therefore completely unsatisfactory if viewed as a deep theory of memory. What our model does instead is provide a means of separately measuring two latent processes in free recall that are confounded in observable data—namely, cluster storage and cluster retrieval. The goal of models like the one in Figure 1 is to separately measure latent processes such as these without commitment to a strong processing account of recall. Like psychometric models, this allows for a certain simplicity in the service of statistical analysis. In fact, we have been able to provide a complete asymptotic analysis of the model in the maximum likelihood framework, and have studied the model from a preasymptotic viewpoint and under conditions of individual differences in the parameters (Batchelder & Riefer, 1986; Riefer & Batchelder, 1991b).

Because they are designed to be used as measurement tools, MPT models may be viewed as less parsimonious in their description of data than complex mathematical theories. As we have discussed previously, MPT models are designed for very specific experimental paradigms, which may yield only a limited set of response categories. Moreover, parameters for a model need to capture the main cognitive factors, including noncentral processes such as guessing, response bias, singleton recall, and so on. As a consequence, there is usually a large number of parameters used to account for a small number of categories, leaving few, if any, degrees of freedom for testing the model's fit. It is not unusual for researchers who have developed MPT models to express concern that their models do not provide a particularly parsimonious account of the data. For example, Wagenaar and Boer (1987) felt that applying their four-parameter model to an experimental paradigm

with six data points was "verging on triviality." Kail et al. (1984) considered it a "drawback" that their model accounted for 14 *df* across two experimental conditions with 12 parameters. However, in our opinion, it is the measurement of the cognitive processes in the form of parameter estimates, and not the data-fitting capacity, that characterizes the usefulness of MPT models.

The issue of parsimony depends primarily on the purpose of a mathematical model. Complex theoretical models attempt a detailed account of cognitive processing, often by specifically modeling the dynamics of how this processing changes across the levels of continuous variables, such as serial position, lag, trials, or time. They provide a theoretically based account for these changes, usually with only a few central assumptions and parameters. In contrast, MPT models in their capacity as measurement tools normally do not specify the precise theoretical mechanisms that account for why parameters increase or decrease as a function of different variables. Once an MPT model has been established as a valid measurement tool, the purpose of the model is to measure changes in parameters over experimental conditions, and this typically requires a separate measure of each parameter for each condition. The result is a proliferation of estimated parameters and a model that appears not to provide a very parsimonious description of the data.

We do not view this as an inherent problem for MPT models, at least no more than it is for other mathematical models for measurement. Models of signal detection, for example, derive separate measures for item detection and response bias from only two data observations (hits and false alarms). These models yield many estimates of the parameters across conditions, and, in general, there are no degrees of freedom to evaluate the model's fit to data. The same can be said about some of the psychometric models mentioned earlier; however, little concern is expressed about the parsimony of these models. The main reason for this, of course, is that all of the above models have been extensively and rigorously tested, and they have proven to offer useful and valid measures of their underlying cognitive constructs. This brings us back to the point of validity testing for MPT models. As important as it is to determine whether or not an MPT model fits the data, we believe that an even more crucial test of a model's validity is to show that the model performs well under basic experimental manipulations. If the model's parameters behave in a psychologically interpretable fashion, then the model gains credence as a valid measurement tool. Researchers should then be able to use the model without concern, even if it does contain many parameters that have to be estimated when applied to data.

CURRENT AND FUTURE DIRECTIONS

In this article, we have described a general class of statistical models, discussed some relevant theoretical issues related to this type of modeling, and given a broad overview of its many applications in psychology. MPT models have been applied to a wide range of different con-

tent areas and to a number of diverse theoretical issues within memory and cognition. In general, however, most of these applications can be classified into one or both of two broad categories: (1) models that formalize psychological theory, and (2) models designed to measure cognitive capacities. In this section, we discuss these two general uses for MPT models plus one other that we feel holds great potential for future applications.

MPT modeling is particularly suitable as a methodology for formally representing psychological theory. As Ashby et al. (1996, pp. 188–189) have pointed out, MPT models are flexible and easy to modify for the purpose of incorporating and testing different theoretical assumptions. Their simple mathematical properties allow one to evaluate in detail the consequences that theories have on empirical data. This approach has been used successfully, for example, by Riefer and Batchelder (1995) to explore the recognition-failure paradigm, by Schweickert (1993) to investigate STM, and by Evans (1977) and Krauth (1982) to study propositional reasoning. Along these lines, MPT models are also particularly useful for directly comparing different theories. The studies by Wagenaar and Boer (1987), B. H. Ross and Bower (1981), and Ashby et al. are all excellent examples of how precise comparisons between competing theories can be made when put into MPT form.

As stated earlier, an equally important use of MPT models is as a tool for the measurement of cognitive capacities. MPT models can be used to disentangle the empirical effects of opposing latent processes and, thus, can extract information from data in ways that ad hoc statistical techniques, such as ANOVA, cannot. Numerous examples of this approach have appeared in the literature, including models for measuring storage versus retrieval in memory (Batchelder & Riefer, 1986; Chechile, 1987; Howe, 1995), detection versus discrimination in source monitoring (Batchelder & Riefer, 1990; Bayen et al., 1996), recollection versus familiarity in process dissociation (Buchner et al., 1995), and perception versus binding in object perception (Prinzmetal et al., 1995). Because their mathematical properties have been worked out in detail, a large number of convenient statistical techniques are available to researchers when analyzing data using MPT models, including confidence intervals, hypothesis testing, goodness of fit, and computer simulations. MPT models are capable of providing specific measures for different cognitive processes and, as such, have been used with great success to explore a variety of important theoretical issues.

However, in our opinion, there is one other application of MPT modeling that has great potential. Because they are designed to measure cognitive functioning, MPT models can also be used as diagnostic tools to evaluate cognitive deficits in clinical populations (e.g., Batchelder, 1998). There have, on a few occasions, been studies in which MPT models have been applied to special populations. For example, both Riefer and Batchelder (1991a) and Howe and Hunter (1986) have applied storage–retrieval models to measure cognitive deficits in elderly

populations. In terms of clinical applications, Chechile's model has been applied to measure storage and retrieval deficits due to alcohol-induced amnesia (Gerrein & Chechile, 1977), developmental dyslexia (Chechile & Roder, 1998), and in mildly retarded adults (Gutowski & Chechile, 1987). Kail et al. (1984) conducted a storage–retrieval analysis of cognitive deficits in language-impaired children, and finite-trial Markov models have been used to measure storage and retrieval deficits in dementia due to alcoholism and Alzheimer's disease (e.g., Howe, 1990; Kraemer, Peabody, Tinklenberg, & Yesavage, 1983).

Perhaps a prototypical example of a clinical application of MPT models can be found in a recent study by Batchelder, Chosak-Reiter, Shankle, and Dick (1997). Batchelder, Chosak-Reiter, et al. examined multitrial, free-recall data taken from the CERAD (consortium to establish a registry for Alzheimer's disease), and analyzed these data using both traditional statistics and an MPT model. The specific experimental task involved a memory test consisting of three study-test trials, followed by two delayed test trials. However, unlike the usual way that data are obtained from controlled experimental conditions, these data were extracted from a large data bank of individuals undergoing a test battery to assess possible memory deficits. Nine groups of subjects were compared on this task, representing different severity levels of Alzheimer's disease and vascular dementia, plus a control group of healthy elderly adults. The traditional analysis involved an ANOVA conducted on the aggregate group means, which revealed significant improvement in memory performance across trials and significant memory differences between some of the clinical groups.

However, Batchelder, Chosak-Reiter, et al. (1997) demonstrated that, by analyzing the error success patterns for individual items, more information could be obtained from this task than by analyzing the aggregate data with ANOVA. Specifically, examining the successful or unsuccessful recall of items over the three study-test trials creates eight (2^3) response protocols. This data structure is similar to the one examined by Kail et al. (1984), except their experiment involved three successive recalls following a single study presentation. Batchelder, Chosak-Reiter, et al. developed an MPT model to analyze these data on the basis of assumptions taken from traditional Markov learning models (Wickens, 1982). The model postulates that items can reside in an unlearned, a temporary, or a permanent state of storage (cf. Howe, 1990; Kraemer et al., 1983) and contains parameters for the transitions between these states, as well as memory retrieval parameters for both intermediate and long-term learning states. When applied to the response protocols, including the two delayed test trials, the model revealed significant differences between the groups that were not evident using the traditional ANOVA.

As indicated in the introduction, in the 10 years since Riefer and Batchelder's (1988) study, there has been a great increase in the development and application of MPT models. In fact, approximately two thirds of all the empirical articles dealing with models in this class and almost all of the theoretical articles dealing specifically with statistical inference for MPT models have been produced during this period. So far, the vast majority of MPT models have been developed within the area of human memory and cognition. However, categorical data are not the exclusive domain of cognitive psychology, and, as MPT modeling continues to grow in popularity, it is reasonable to expect an even wider range of new applications. It is our hope that in the future, even more substantive areas of psychology will benefit from the application of MPT models.

REFERENCES

AGRESTI, A. (1990). *Categorical data analysis*. New York: Wiley.

ANDERSON, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

ANDERSON, N. H. (1982). *Foundations of information integration theory*. New York: Academic Press.

ANDERSON, N. H. (1990). Integration psychophysics. In H.-G. Geissler (Ed.), *Psychophysical explorations of mental structures* (pp. 71-93). Lewiston, NY: Hogrefe & Huber.

ASHBY, F. G. (ED.) (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

ASHBY, F. G., PRINZMETAL, W., IVRY, R., & MADDOX, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, **103**, 165-192.

BAKER, F. B. (1992). *Item response theory*. New York: Marcel Dekker.

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.

BATCHELDER, W. H. (1975). Individual differences and the all-or-none vs. incremental learning controversy. *Journal of Mathematical Psychology*, **12**, 53-74.

BATCHELDER, W. H. (1991). Getting wise about minimum distance measures. *Journal of Mathematical Psychology*, **35**, 267-273.

BATCHELDER, W. H (1998). Multinomial processing, tree models and psychological assessment. *Psychological Assessment*, **10**, 331-344.

BATCHELDER, W. H., CHOSAK-REITER, J., SHANKLE, W. R., & DICK, M. B. (1997). A multinomial modeling analysis of memory deficits in Alzheimer's and vascular dementia. *Journal of Gerontology: Psychological Sciences*, **52B**, 206-215.

BATCHELDER, W. H., & CROWTHER, C. S. (1997). Multinomial processing tree models of factorial categorization. *Journal of Mathematical Psychology*, **41**, 45-55.

BATCHELDER, W. H., HU, X., & RIEFER, D. M. (1994). Analysis of a model for source monitoring. In G. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 51-65). New York: Springer-Verlag.

BATCHELDER, W. H., KUMBASAR, E., & BOYD, J. P. (1997). Consensus analysis of three-way social network data. *Journal of Mathematical Psychology*, **22**, 29-58.

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 & Statistical Psychology*, **39**, 120-149.

BATCHELDER, W. H., & RIEFER, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, **97**, 548-564.

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

BATCHELDER, W. H., & ROMNEY, A. K. (1986). The statistical analysis of a general Condorcet model for dichotomous choice situations. In B. Grofman & G. Owen (Eds.), *Information pooling and group decision making* (pp. 103-112). Greenwich, CT: JAI Press.

BATCHELDER, W. H., & ROMNEY, A. K. (1988). Test theory without an answer key. *Psychometrika*, **53**, 71-92.

BATCHELDER, W. H., & ROMNEY, A. K. (1989). New results in test theory without an answer key. In E. Roskam (Ed.), *Advances in mathematical psychology: Vol. II* (pp. 229-248). New York: Springer-Verlag.

BÄUML, K.-H. (1991). Experimental analysis of storage and retrieval processes involved in retroactive inhibition: The effect of presentation mode. *Acta Psychologica*, **77**, 103-119.

BÄUML, K.-H. (1996a). A Markov model for measuring storage loss and retrieval failure in retroactive inhibition. *Acta Psychologica*, **92**, 231-250.

BÄUML, K.-H. (1996b). Revisiting an old issue: Retroactive interference as a function of the degree of original and interpolated learning. *Psychonomic Bulletin & Review*, **3**, 380-384.

BAYEN, U. J., & MURNANE, K. (1996). Aging and the use of perceptual and temporal information in source memory tasks. *Psychology & Aging*, **11**, 293-303.

BAYEN, U. J., MURNANE, K., & ERDFELDER, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 197-215.

BEGG, I. A., ANAS, A., & FARINACCI, S. (1992). Dissociation of processing in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, **121**, 446-458.

BENDER, R. H., WALLSTEN, T. S., & ORNSTEIN, P. A. (1996). Age differences in encoding and retrieving details of a pediatric examination. *Psychonomic Bulletin & Review*, **3**, 188-198.

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

BOWER, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika*, **26**, 255-280.

BRAINERD, C. J., REYNA, V. F., HOWE, M. L., & KINGMA, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development*, **55**(3-4, Serial No. 222).

BROWN, V. R. (1998). Comparing parallel and sequential multinomial models of letter identification. In C. Dowling, F. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology* (pp. 253-284). Hillsdale, NJ: Erlbaum.

BUCHNER, A., & ERDFELDER, E. (1996). On assumptions of, relations between, and evaluations of some dissociation measurement models. *Consciousness & Cognition*, **5**, 581-594.

BUCHNER, A., ERDFELDER, E., STEFFENS, M. C., & MARTENSEN, H. (1997). The nature of memory processes underlying recognition judgments in the process-dissociation procedure. *Memory & Cognition*, **25**, 508-517.

BUCHNER, A., ERDFELDER, E., & VATERRODT-PLUNNECKE, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, **124**, 137-160.

BUCHNER, A., STEFFENS, M. C., ERDFELDER, E., & ROTHKEGEL, R. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *Quarterly Journal of Experimental Psychology*, **50A**, 631-663.

BUCHNER, A., & WIPPICH, W. (1996). Unconscious gender bias in fame judgments. *Consciousness & Cognition*, **5**, 197-220.

BUTLER, K., & CHECHILE, R. [A.] (1976). "Acid bath" effects on storage and retrieval PI. *Bulletin of the Psychonomic Society*, **8**, 349-352.

CHANDLER, J. P. (1969). STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, **14**, 81-82.

CHECHILE, R. [A.] (1977). Storage–retrieval analysis of acoustic similarity. *Memory & Cognition*, **5**, 535-540.

CHECHILE, R. A. (1987). Trace susceptibility theory. *Journal of Experimental Psychology: General*, **116**, 203-222.

CHECHILE, R. A. (1993, August). *Two models for measuring fractional storage for list learning tasks*. Paper presented at the 26th Annual Meeting of the Society for Mathematical Psychology, Norman, OK.

CHECHILE, R. A., & BUTLER, K. (1975). Storage and retrieval changes that occur in the development and release of PI. *Journal of Verbal Learning & Verbal Behavior*, **14**, 430-437.

CHECHILE, R. A., & EHRENSBECK, K. (1983). Long-term storage losses: A dilemma for multistore models. *Journal of General Psychology*, **109**, 15-30.

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

CHECHILE, R. A., & RICHMAN, C. L. (1982). The interaction of semantic memory with storage and retrieval processes. *Developmental Review*, **2**, 237-250.

CHECHILE, R. A., RICHMAN, C. L., TOPINKA, C., & EHRENSBECK, K. (1981). A developmental study of the storage and retrieval of information. *Child Development*, **52**, 251-259.

CHECHILE, R. A., & RODER, B. (1998). Model-based measurement of group differences: An application directed toward understanding the information-processing mechanisms of developmental dyslexia. In S. A. Soraci & W. J. McIlvane (Eds.), *Perspectives on fundamental processes in intellectual functioning: Vol. 1. A survey of research approaches* (pp. 91-112). Norwich, CT: JAI Press.

CLARK, S. E., & GRONLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3**, 37-60.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

COHEN, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155-159.

CORBETT, A. T., & WICKELGREN, W. A. (1978). Semantic memory retrieval: Analysis by speed accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology*, **30**, 1-15.

CROWTHER, C. S., BATCHELDER, W. H., & HU, X. (1995). A measurement-theoretic analysis of the fuzzy logic model of perception. *Psychological Review*, **102**, 134-148.

CURRAN, T., & HINTZMAN, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 531-547.

DAPOLITO, F. J. (1967). Proactive effects with independent retrieval of competing responses. *Dissertation Abstracts*, **27**, 2522-2523.

DAVID, H. A. (1988). *The method of paired comparisons* (2nd ed.). London: Griffin.

DEHN, D. M., & ENGELKAMP, J. (1997). Process dissociation procedure: Double dissociations following divided attention and speeded responding. *Quarterly Journal of Experimental Psychology*, **50A**, 318-336.

DEHN, D. M., & ERDFELDER, E. (1998). What kind of bias is hindsight bias? *Psychological Research*, **61**, 135-146.

DEMPSTER, A. P., LAIRD, N. M., & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, **39**, 1-38.

ERDFELDER, E., & BREDENKAMP, J. (1998). Recognition of script-typical versus script-atypical information: Effects of cognitive elaboration. *Memory & Cognition*, **26**, 922-938.

ERDFELDER, E., & BUCHNER, A. (1998a). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 387-414.

ERDFELDER, E., & BUCHNER, A. (1998b). Process dissociation measurement models: Threshold theory or detection theory? *Journal of Experimental Psychology: General*, **127**, 83-96.

ERDFELDER, E., FAUL, F., & BUCHNER, A. (1995). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, **28**, 1-11.

EVANS, J. ST. B. (1977). Toward a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology*, **29**, 621-635.

EVANS, J. ST. B., & LYNCH, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391-397.

FERGUSON, S. A., HASHTROUDI, S., & JOHNSON, M. K. (1992). Age differences using source-relevant cues. *Psychology & Aging*, **7**, 443-452.

FISCHER, G. H., & MOLENAAR, I. W. (EDS.) (1995). *Rasch models: Foundations, recent developments, and applications*. New York: Springer-Verlag.

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

GARCIA-PEREZ, M. A. (1987). A finite theory of performance in multiple-choice tests. In E. E. Roskam & R. Suck (Eds.), *Progress in mathematical psychology 1* (pp. 455-464). Amsterdam: Elsevier.

GARCIA-PEREZ, M. A. (1989). Item sampling, guessing, partial information and decision-making in achievement testing. In E. E. Roskam (Ed.), *Mathematical psychology in progress* (pp. 249-265). Berlin: Springer-Verlag.

GARCIA-PEREZ, M. A. (1990). A comparison of two models of performance in objective tests: Finite state versus continuous distributions. *British Journal of Mathematical & Statistical Psychology*, 43, 73-91.

GARCIA-PEREZ, M. A. (1993). In defence of "none of the above." *British Journal of Mathematical & Statistical Psychology*, 46, 213-229.

GARCIA-PEREZ, M. A., & FRARY, R. B. (1991a). Finite state polynomic item characteristic curves. *British Journal of Mathematical & Statistical Psychology*, 44, 45-73.

GARCIA-PEREZ, M. A., & FRARY, R. B. (1991b). Testing finite-state models of performance in objective tests using items with "none of the above" as an option. In J.-P. Doignon & J.-C. G. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 273-291). New York: Springer-Verlag.

GERREIN, J. R., & CHECHILE, R. A. (1977). Storage and retrieval processes of alcohol-induced amnesia. *Journal of Abnormal Psychology*, 86, 285-294.

GOLDEN, R. M. (1995). Making correct statistical inferences using a wrong probability model. *Journal of Mathematical Psychology*, 39, 3-20.

GONDRAN, M., MINOUX, M., & VAJDA, S. (1984). *Graphs and algorithms*. New York: Wiley.

GREEN, D. M., & BIRDSALL, T. G. (1978). Detection and recognition. *Psychological Review*, 85, 192-206.

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

GUTOWSKI, W. E., & CHECHILE, R. A. (1987). Encoding, storage, and retrieval components of associative memory deficits of mildly retarded adults. *American Journal of Mental Deficiency*, 92, 85-93.

HAWKINS, S. A., & HASTIE, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107, 311-327.

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

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

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

HOWE, M. L. (1990). Development of a mathematical model of memory for clinical research applications in aging. In M. L. Howe, M. J. Stones, & C. J. Brainerd (Eds.), *Cognitive and behavioral performance factors in atypical aging* (pp. 3-36). New York: Springer-Verlag.

HOWE, M. L. (1991). Misleading children's story recall: Forgetting and reminiscence of the facts. *Developmental Psychology*, 27, 746-762.

HOWE, M. L. (1995). Interference effects in young children's long-term retention. *Developmental Psychology*, 31, 579-596.

HOWE, M. L., & BRAINERD, C. J. (1989). Development of children's long-term retention. *Developmental Review*, 9, 301-340.

HOWE, M. L., COURAGE, M. L., & BRYANT-BROWN, L. (1993). Reinstating preschooler's memories. *Developmental Psychology*, 29, 854-869.

HOWE, M. L., & HUNTER, M. A. (1986). Long-term memory in adulthood: An examination of the development of storage and retrieval processes at acquisition and retention. *Developmental Review*, 6, 334-364.

HOWE, M. L., KELLAND, A., BRYANT-BROWN, L., & CLARK, S. L. (1992). Measuring the development of children's amnesia and hypermnesia. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 56-102). New York: Springer-Verlag.

HOWE, M. L., & O'SULLIVAN, J. T. (1997). What children's memories tell us about recalling our childhood: A review of storage and retrieval processes in the development of long-term retention. *Developmental Review*, 17, 148-204.

HOWE, M. L., & RABINOWITZ, F. M. (1996). Reasoning from memory: A lifespan inquiry into the necessity of remembering when reasoning about class inclusion. *Journal of Experimental Child Psychology*, 61, 1-42.

HU, X., & BATCHELDER, W. H. (1994a, August). *Empirical Bayes approach for GPT models*. Paper presented at the 27th Annual Meeting of the Society for Mathematical Psychology, Seattle.

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

HULME, C., ROODENRYS, S., SCHWEICKERT, R., BROWN, G. D., MARTIN, S., & STUART, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 5, 1217-1232.

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

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

HUTCHINSON, T. P. (1982). Some theories of performance in multiple choice tests and their implications for variants of the task. *British Journal of Mathematical & Statistical Psychology*, 35, 71-89.

JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, 30, 513-541.

JACOBY, L. L. (1998). Invariance in automatic influences of memory: Towards a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24, 3-26.

JACOBY, L. L., TOTH, J. P., & YONELINAS, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122, 139-154.

JOHNSON, M. K., KOUNIOS, J., & REEDER, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20, 1409-1419.

JOHNSON-LAIRD, P. N., BYRNE, R. M., & SCHAEKEN, W. (1992). Propositional reasoning by model. *Psychological Review*, 99, 418-439.

JONES, G. V. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. *Journal of Experimental Psychology: General*, 105, 277-293.

KAIL, R., HALE, C. A., LEONARD, L. B., & NIPPOLD, M. (1984). Lexical storage and retrieval in language-impaired children. *Applied Psycholinguistics*, 5, 37-49.

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

KLAUER, K. C., & BATCHELDER, W. H. (1996). Structural analysis of subjective categorical data. *Psychometrika*, 61, 199-240.

KLAUER, K. C., & OBERAUER, K. (1995). Testing the mental model theory of propositional reasoning. *Quarterly Journal of Experimental Psychology*, 48A, 671-687.

KLAUER, K. C., & WEGENER, I. (1998). Unraveling social categorization in the "Who said what?" paradigm. *Journal of Personality & Social Psychology*, 75, 1155-1178.

KRAEMER, H. C., PEABODY, C. A., TINKLENBERG, J. R., & YESAVAGE, J. A. (1983). Mathematical and empirical development of a test of memory for clinical and research use. *Psychological Bulletin*, 94, 367-380.

KRAUTH, J. (1982). Formulation and experimental verification of models in propositional reasoning. *Quarterly Journal of Experimental Psychology*, 34, 285-298.

LABERGE, D., & BROWN, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, 96, 101-124.

LIGHT, L. L., LAVOIE, D., VALENCIA-LAVER, D., ALBERTSON OWENS, S. A., & MEAD, G. (1992). Direct and indirect measures of memory for modality in young and older adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 1284-1297.

LINDSAY, D. S. (1993). Eyewitness suggestibility. *Current Directions in Psychological Science*, 2, 86-89.

LINDSAY, D. S., JOHNSON, M. K., & KWON, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, 52, 297-318.

LOFTUS, E. F., DONDERS, K., HOFFMAN, H. G., & SCHOOLER, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition*, 17, 607-616.

LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review, 70*, 61-79.

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

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252-271.

Marche, T. A., & Howe, M. L. (1995). Preschoolers report misinformation despite accurate memory. *Developmental Psychology, 31*, 554-567.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

McCloskey, M., & Zaragoza, M. (1985). Postevent information and memory: Reply to Loftus, Schooler, and Wagenaar. *Journal of Experimental Psychology: General, 114*, 381-387.

McLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensions.* New York: Wiley.

Mulder, G., Sanders, A. F., & van Galen, G. P. (Eds.) (1995). Discrete and continuous information processing [Special issue]. *Acta Psychologica, 90* (1-3).

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 1067-1087.

Mulligan, N. W., & Hirshman, E. (1997). Measuring the bases or recognition memory: An investigation of the process-dissociation framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23*, 280-304.

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

Offir, J. (1972). Stochastic learning models with distributions of parameters. *Journal of Mathematical Psychology, 9*, 404-417.

Payne, D. G., & Roediger, H. L., III (1987). Hypermnesia occurs in recall but not recognition. *American Journal of Psychology, 100*, 145-165.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International University Press.

Prinzmetal, W., Henderson, D., & Ivry, R. (1995). Loosening the constraints on illusory conjunctions: Assessing the roles of exposure duration and attention. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 1362-1375.

Rabinowitz, F. M., Howe, M. L., & Lawrence, J. A. (1989). Class inclusion and working memory. *Journal of Experimental Child Psychology, 48*, 379-409.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Paedagogiske Institute.

Read, T. R. C., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer-Verlag.

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

Riefer, D. M., & Batchelder, W. H. (1987, November). *Further tests of a model for measuring storage and retrieval.* Paper presented at the annual meeting of the Psychonomic Society, Seattle.

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

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

Riefer, D. M., & Batchelder, W. H. (1991b). Statistical inference for multinomial processing tree models. In. J.-P. Doignon & J.-C. G. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 313-336). New York: Springer-Verlag.

Riefer, D. M., & Batchelder, W. H. (1995). A multinomial modeling analysis of the recognition-failure paradigm. *Memory & Cognition, 23*, 611-630.

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 680-693.

Riefer, D. M., & LaMay, M. L. (1998). Memory for common and

bizarre stimuli: A storage-retrieval analysis. *Psychonomic Bulletin & Review, 5*, 312-317.

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition, 20*, 601-611.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*, 38-71.

Romney, A. K., Batchelder, W. H., & Weller, S. C. (1987). Recent applications of consensus theory. *American Behavior Scientist, 31*, 163-177.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and accuracy. *American Anthropologist, 88*, 313-338.

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

Ross, S. M. (1983). *Stochastic processes.* New York: Wiley.

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired-associate learning. In C. Dowling, F. Roberts, & P. Theuns (Eds.), *Progress in mathematical psychology* (pp. 195-225). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations.* Cambridge, MA: MIT Press.

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

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*, 168-175.

Skoff, B., & Chechile, R. A. (1977). Storage and retrieval processes in the serial position effect. *Bulletin of the Psychonomic Society, 9*, 265-268.

Trabasso, T., & Bower, G. H. (1964). *Attention in learning: Theory and research.* New York: Wiley.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107-141.

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.

Tulving, E., & Wiseman, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society, 6*, 79-82.

Wagenaar, W. A., & Boer, J. P. A. (1987). Misleading postevent information: Testing parameterized models of integration in memory. *Acta Psychologica, 66*, 291-306.

Wainwright, M. J., & Reingold, E. M. (1996). Response bias correction in the process dissociation procedure: Approaches, assumptions, and evaluation. *Consciousness & Cognition, 5*, 232-254.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology I* (pp. 135-151). Harmondsworth, U.K.: Penguin.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72*, 89-104.

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

White, H. (1994). *Estimation, inference, and specification analysis.* New York: Cambridge University Press.

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

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 1341-1354.

Yonelinas, A. P., & Jacoby, L. L. (1996). Response bias and the process dissociation procedure. *Journal of Experimental Psychology: General, 125*, 422-434.

Yonelinas, A. P., Regehr, G., Jacoby, L. L. (1995). Incorporating response bias in a dual-process theory of memory. *Journal of Memory & Language, 34*, 821-835.