Minimum Description Length Model Selection of Multinomial Processing Tree Models

Hao Wu and Jay I. Myung
The Ohio State University

William H. Batchelder
University of California, Irvine

Word count of text and appendices: 7700

Running head: Multinomial Processing Tree Models

Corresponding author and address:

Hao Wu
Department of Psychology
Ohio State University
1835 Neil Avenue
Columbus, Ohio 43210-1351
E-mail: wu.498@osu.edu
Tel: 614-292-5510
Multinomial processing tree (MPT) modeling has been widely and successfully applied as a statistical methodology for measuring hypothesized latent cognitive processes in selected experimental paradigms. This paper concerns the problem of selecting the “best” MPT model from a set of scientifically plausible MPT models given observed data. The likelihood ratio test is often employed in model selection for MPT models, but it is a null hypothesis significance test that assesses the descriptive adequacy of a given null model, and as such, does not necessarily help identify the best approximating model to the truth, which is the hallmark of model selection. Model selection methods such as the Akaike Information Criterion and the Bayesian Information Criterion do not fully take into account all relevant dimensions of model complexity, such as the number of parameters, model structure, and parametric inequality constraints, the latter two of which are of particular importance for MPT models. In this paper, we introduce a minimum description length (MDL) based model selection approach that overcomes the limitations of the aforementioned methods and therefore is well suited for model selection of MPT models. To help ease the computational burden of implementing MDL, we provide a computer program in MatLab that performs MDL-based model selection for any MPT model, with or without inequality constraints. Finally, we discuss applications of the the MDL approach to well-studied MPT models with real data sets collected in two different experimental paradigms: source monitoring and pair-clustering. The aforementioned MatLab program may be downloaded from www.psychonomic.org/archive.
Introduction

Multinomial processing tree (MPT) modeling is a statistical methodology for measuring latent cognitive capacities in selected experimental paradigms (Batchelder & Riefer, 1986, 1990, 1999; Hu & Batchelder, 1994; Chechile, 2004; Riefer & Batchelder, 1988, 1991, 1995; Riefer, Hu & Batchelder, 1994). The data structure requires that participants performing a cognitive task make categorical responses to a series of test items. An MPT model parameterizes a subset of probability distributions over the response categories by specifying a processing tree designed to represent hypothesized cognitive steps, such as memory encoding, storage, discrimination, inference, guessing, and retrieval.

Since its introduction in the 1980s, MPT models have been successfully applied to modeling performance in a wide range of cognitive tasks including associative recall, source monitoring, eyewitness memory, hindsight bias, object perception, speech perception, propositional reasoning, social networks, and cultural consensus. Batchelder and Riefer (1999) lists over 80 applications of MPT models in various areas of cognitive and social psychology. MPT models have also been applied to estimate cognitive deficits in special populations (see Batchelder & Riefer, 2007; Chechile, 2007, for a review of such applications). The use of MPT models to assess special populations is often referred to as cognitive psychometrics representing the fact that theoretically motivated models are employed as measurement tools of cognitive functioning (Batchelder, in press; Batchelder & Riefer, 2007; Riefer et al., 2002). In all these applications, MPT models intended to offer researchers more instructive and informative interpretations of data than those based on the traditional data analytic approaches such as the analysis of variance (ANOVA).
MULTINOMIAL PROCESSING TREE MODELS

In the present study, we are concerned with the logic of selecting the “best” MPT model from a set of scientifically plausible MPT models that are available to account for a given data set. A researcher may entertain multiple scientific hypotheses about the underlying processes, each formulated as a distinct MPT model\(^1\), and may wish to determine which one of these models “best” describes the observed data in some defined sense; this is the problem of model selection (Myung & Pitt, 1997). By selecting among theoretically motivated models, the researcher is able to identify from alternative theories the one best supported by empirical observations. To illustrate, consider the question of how different languages of bilingual people are cognitively represented. Several theories addressing this issue differ as to whether or not information presented in a particular language retains a language specific tag. Source monitoring experiments were conducted to differentiate these theories (e.g. Saegert, Hamayan & Ahmar, 1975; Rose et al., 1975), and these theories, as represented by their corresponding source monitoring MPT models (to be elaborated in the next section) that assume different treatment effects on their parameters, can be compared by model selection (Batchelder & Riefer, 1990). Similarly, other theoretical issues in cognitive psychology such as sequential vs. non-sequential processes and automatic vs. control processes can also be addressed by comparing MPT models with different tree structures (e.g., Schweickert, 1993; Bishara & Payne, 2008).

In addition to evaluating multiple scientific theories behind different MPT models, model selection can also be employed as a tool for examining the validity of an MPT model.

\(^1\)In this paper, an MPT model may either refer to a model for a particular experimental paradigm (e.g. source monitoring, with three trees), or a set of such models each representing a different experimental condition in an experiment.
The validity of an MPT model concerns whether or not it is warranted to interpret a parameter in the model as representing the underlying cognitive process that it is explicitly postulated to represent (see, e.g., Batchelder & Riefer, 1999; Riefer et al., 2002; Schweickert & Chen, 2008). To establish validity, it is necessary to apply experimental treatments that have predictable selective influence on the parameters. For example, if a model has a parameter \( \theta \) that is postulated to measure the ability to retrieve items from memory, then experimental manipulations that should affect levels of retrievability should result in predictable changes in \( \theta \) but no change in parameters postulated to measure other things.

To determine whether the desired selected influence is present for a particular MPT model, it is necessary to select among different versions of the model assuming different patterns of treatment effects.

Because of its importance in evaluating scientific theories and establishing validity of MPT models, model selection is of particular interest in MPT modeling. To perform model selection, one must account for the effect of model complexity. This is because model complexity can affect the predictive capacity or accuracy of a model, which is the hallmark of model selection (Myung, 2000; Myung & Pitt, 1997). In the case of MPT models, it has been shown that they can vary greatly in complexity due to not only the number of parameters but also importantly, functional form of the models such as tree structure and parameter constraints (Wu, Myung & Batchelder, submitted). However, as will be discussed later in this paper, the likelihood ratio test (LRT: Read & Cressie, 1988), currently in wide use for MPT modeling, does not select models based on their predictive accuracy. Other popular selection methods such as Akaike information criterion (AIC: Akaike, 1973) and Bayesian information criterion (BIC: Schwartz, 1978) do not fully account for all dimensions
of model complexity. Given these limitations, a model selection method that fully accounts for model complexity is called for.

In the present paper we introduce such a method for MPT modeling. This is minimum description length (MDL) model selection. MDL has been successfully applied to addressing various model selection problems in cognitive modeling (e.g., Lee, 2001; Pitt, Myung & Zhang, 2002; Navarro & Lee, 2004; Lee & Pope, 2001; Myung, Pitt & Navarro, 2007) but is entirely absent in MPT modeling, with the exception of our own work (Wu, Myung & Batchelder, submitted). To help researchers not familiar with numerical computing, in this paper we make available a general purpose computer program that implements MDL based model selection for virtually all types of MPT models.

The rest of the paper is organized as follows. We first begin with a formal definition of MPT models. We then briefly review the extant methods of model selection such as LRT, AIC, and BIC, before introducing MDL, the focus of the present work. The discussion then turns to the computer program, and we provide an instruction in detail of how to use it in a given situation of MPT modeling. Finally, two application examples of MDL based model selection with real data sets are presented before concluding the paper.

Multinomial Processing Tree Models

Multinomial processing tree (MPT) models assume that the observed categorical responses in an experiment follow from a series of latent cognitive events. These events are represented by a tree structure, in which non-terminating nodes represent the events, branches that follow from a node represent all possible outcomes of the event, with the probabilities of these outcomes being either parameters in the model or known constants,
and leaves (terminating nodes) of the tree structure represent the observed responses from subjects. Because different sequences of events may lead to the same response, a response category may include more than one leaf in the tree.

To illustrate how an MPT model works, consider the one-high-threshold model (1HTM) for source monitoring experiments as depicted in Figure 1. In a source monitoring experiment, participants first study a list of items from two sources, A and B, and then are asked to judge the source of test items as either from A, from B, or new (N; i.e. a new item from neither source). The 1HTM for such experiments consists of three distinct trees (Batchelder & Riefer, 1990), each modeling hypothetical processes assumed to be involved in responding to a given type of items. A distinguishing feature of this model is that it assumes that old items can be correctly detected with probabilities $D_1$ and $D_2$ for items from sources A and B, respectively. If an old item is correctly detected as old, a discriminating decision on its source is made, with success probabilities $d_1$ and $d_2$ for the two sources, respectively. If any of the two processes fails, guessing processes follow. For new items, however, the model assumes no detection process and instead response selection is governed by guessing processes only. The model postulates three types of guessing processes represented by parameters $b$, $g$, and $a$ (see Figure 1 for details). By putting various constraints on the model parameters, a hierarchy of sub-models can be derived from the model, which is shown in Figure 2. For instance, the equality constraints of $D_1 = D_2$ and $d_1 = d_2$, which amount to saying that the detection and discrimination probabilities both stay the same across items from different sources, results in 1HTM-5a. On the other hand, if we assume that only the source discrimination probabilities are the same for both sources ($d_1 = d_2$) but not the detection probabilities ($D_1 \neq D_2$), then 1HTM-6b, which
nests 1HTM-5a, is obtained instead.

Speaking in formal terms, an MPT model parameterizes a subset of multinomial probability distributions over response categories. Because every MPT model can be reparameterized into a binary MPT (BMPT) model in which every decision node has only two processing possibilities (Hu & Batchelder, 1994), we will only discuss the mathematical formulation of BMPT models. Suppose a BMPT model has $S$ parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_S)'$ and $J$ categories ($C_1, C_2, \ldots, C_J$), and category $C_j$ includes leaves $B_{ij}$ ($i = 1, 2, \ldots, I_j; j = 1, 2, \ldots, J$). Because of its binary nature, non-constant probabilities on the branches must be of the form $\theta_s$ or $(1 - \theta_s)$. The probability of taking the decision path to a leaf $B_{ij}$ is given by the product of all probabilities along this path

$$ p_{ij}(\theta) = c_{ij} \prod_{s=1}^{S} \theta_s^{a_{ijs}}(1 - \theta_s)^{b_{ijs}} $$

(1)

where $a_{ijs}$ and $b_{ijs}$ are, respectively, the number of times $\theta_s$ and $1 - \theta_s$ appear on the path to $B_{ij}$, and $c_{ij}$ is the product of all constant probabilities along the same path or set to unity if there is no constant probability along this path. The probability of category $C_j$ is the sum of the probabilities of all leaves it includes, i.e.,

$$ p_j(\theta) = \sum_{i=1}^{I_j} p_{ij}(\theta) $$

(2)

For example, each tree in 1HTM discussed above is a BMPT model. The probability for a subject to respond “source A” given a stimulus from source A is given by $D_1 d_1 + D_1(1 - d_1)a + (1 - D_1)bg$.

Now let us assume that several participants make categorical responses to the same set of items and that their responses are independently and identically distributed into the $J$ categories of a model. Let $n_j$ be the number of these responses that fall into category $C_j$,.
MULTINOMIAL PROCESSING TREE MODELS

\[ n = (n_1, n_2, \ldots, n_J)' \] and \[ N = \sum_j n_j. \]

Then \( n \) is distributed as a multinomial probability distribution given by

\[
f(n|\theta) = \binom{N}{n_1, \ldots, n_J} \prod_{j=1}^{J} p_j^{n_j}(\theta)
\]

where the multinomial probabilities \( p_j \) follows the computational rules in equations (2) and (1).

The above mathematical description of BMPT models with constants \( a_{ijs}, b_{ijs} \) and \( c_{ij} \), though uniquely and sufficiently specifying the distribution of the data, can be cumbersome as an input to computer programs. For this purpose, Purdy and Batchelder (in press) has devised a much more concise and elegant representation of BMPT models. Their string representation scheme exploits the recursive properties of the tree structure and includes only branching probabilities and categories in the model. To illustrate, the string representation of a coin flipping Bernouli model is given by \( pHT \), where H and T are outcomes of the process and \( p \) is the probability of obtaining the outcome of H. To obtain the string representation for a more complex BMPT model, one begins with representation of the decision process at the root node, and then replaces the two outcomes with the representations of the decision processes that follow those outcomes. To illustrate, take the tree of source A items in the 1HTM in Figure 1. We first represent the item detection process with \( D_1 \) (“detected”) (“undetected”). We then replace the outcome “detected” with the representation of the discrimination process \( d_1A \) (“source unidentified”) and the outcome “undetected” with that of the guessing process \( b \) (“guess as old”) \( N \). Now we get \( D_1d_1A \) (“source unidentified”) \( b \) (“guess as old”) \( N \). We continue the replacement until the string contains only branching probabilities and response categories. The string representation of the tree is \( D_1d_1AaABbgABN \). This representation makes the input to computer
programs much easier and will be exploited in our *MatLab* program described later in the present article.

**Methods of Model Selection**

As mentioned in the Introduction, model selection is a necessary and crucial step in the application of MPT models. Various model selection methods have been proposed in the past for this purpose. In the following we review some of these methods including the minimum description length method.

**Likelihood Ratio Test**

The $G^2$-based likelihood ratio test (LRT) is the most commonly used method of inference in MPT modeling (e.g., Riefer & Batchelder, 1988; Hu & Batchelder, 1994; Hu & Phillips, 1999). At the center of this approach is the likelihood ratio based test statistic by which the adequacy of a model is evaluated in the null hypothesis significance testing framework. Specifically, LRT requires the setting of two models, full and reduced, such that the reduced model is nested in the full model with a reduction in the number of parameters. The $G^2$ test statistic is then defined as $G^2 = -2 \ln LR$, where $LR$ is the ratio of the maximum likelihood of the reduced model to that of the full model and $\ln$ is the natural logarithm. Under the null hypothesis that the reduced model is correct, when the sample size $N$ is large enough, the sampling distribution of $G^2$ is shown to follow a $\chi^2$-distribution with the degrees of freedom equal to the difference in numbers of parameters between the models, provided that certain regularity conditions are satisfied (e.g. Read & Cressie, 1988). If the value of $G^2$ is large enough to fall in the rejection region of the sampling distribution, then the null hypothesis is rejected and the full model is chosen. Otherwise, the reduced model
is chosen over the full model.

The $G^2$-based LRT is generally a useful method of model evaluation, but has several limitations in its use as a model selection method for MPT models. First, the method can only be used for comparing pairs of nested models, one pair at a time. This effectively excludes its application to the situation in which multiple models with or without nesting relationships are being compared. Second, the regularity conditions of the test require that the maximum likelihood estimate (MLE) under either model should not be on the boundary of the parameter space (see Shapiro, 1988, for an alternative procedure). This implies that LRT is not able to take into account inequality constraints in the models. To see this, because the inclusion of inequality constraints does not change the degrees of freedom of the LRT, it changes the result of the test only when parameters of either model are estimated on the boundary defined by those constraints, but that would violate the regularity conditions mentioned above and render the test invalid. For the same reason, LRT cannot be employed to compare two nested models with the same number of parameters but different functional forms, such as IHTM-6a and 6b in Figure 2.

Besides the above issues, it is important to note that the goal of LRT is to assess the descriptive adequacy of a given null model in the null hypothesis significance test framework, but not to choose among a set of candidate models the one that best captures the regularities underlying the data (Myung & Pitt, 1997). As such, LRT does not necessarily help identify the best approximating model to the truth, which is what model selection is about. This latter criterion is known as generalizability in statistics (e.g., Myung, 2000; Myung & Pitt, 1997). In the rest of this section we discuss various model selection criteria proposed as generalizability measures and the importance of model complexity in determining a model's
Generalizability of a model refers to how well the conclusion from the current observed data can be applied to future, not yet observed, data (Myung, 2000). By definition, the model with best generalizability gives the closest approximation to the underlying mechanism of the data and therefore should be preferred in model selection. Models that generalize well should first provide a good fit to the current data; however, generalizability is more than goodness-of-fit and is significantly affected by model complexity.

Model complexity or flexibility has to do with a model’s intrinsic capability to fit a wide range of data patterns. Generally speaking, a model with many parameters is more complex than a model with fewer parameters. Further, models with the same number of parameters but different equation forms can also differ in complexity. This is called the “functional form” dimension of model complexity (Myung & Pitt, 1997). To give an example, two psychophysics models, $y = ax^b + \varepsilon$ and $y = a \log(x + b) + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$, may differ in complexity, despite the fact that they both have two parameters. Because of its flexibility, complex models tend to overfit the current data, thereby generalizing poorly to future observations and should therefore be penalized in model selection.

In the case of MPT models, differences in complexity due to functional form can arise in a variety of ways: from different tree structures, different parametric constraints, and/or different category assignments to the leaves of a tree. For example, consider 1HTM-5a, 5b and 5c shown in Figure 2, each of which imposes a distinct set of equality constraints on the parameters of the largest model 1HTM-7. Although all three models have five
parameters, their complexity may be quite different from one another. This is in fact what Wu, Myung and Batchelder (submitted) found. Their results showed that the difference in complexity between 1HTM-5a and 5b is larger than that between 1HTM-5b and 4. An implication is that the complexity difference due to functional form of MPT models can be even greater than that due to the number of parameters. The finding such as this points to the importance of accounting for the functional form dimension of complexity, in particular, in model selection with MPT models.

Minimum Description Length

Various model selection methods that estimates a model’s generalizability have been proposed in statistics. Among them, Akaike information criterion (AIC: Akaike, 1973) and Bayesian information criterion (BIC: Schwartz, 1978) have been used in MPT modeling, though minimum description length (MDL: Rissanen, 1996, 2001; Grünwald, 2007) has not, to our knowledge. In what follows, we briefly review AIC and BIC before turning our discussion to MDL.

Unlike LRT, AIC (Akaike, 1973) and BIC (Schwartz, 1978; Wagenmakers, 2007; Raftery, 1999; Weakliem, 1999) can be used to compare multiple, nested or nonnested, models. They are defined as

\[ AIC = -2LML + 2S \]  \hspace{1cm} (4)

\[ BIC = -2LML + S \ln N \]  \hspace{1cm} (5)

The interested reader is directed to two recent *Journal of Mathematical Psychology* special issues on model selection for discussion and example applications of these and other methods of model selection (Myung, Forster & Browne, 2000; Wagenmakers & Waldorp, 2006).
where $LML (= \ln f(x|\hat{\theta}(x)))$ denotes the natural logarithm of the maximized likelihood (ML), $x$ is the current data, $S$ is the number of parameters and $N$ is the sample size.

For both model selection criteria, among a set of competing models, the model with the smallest criterion value is judged to best generalize and is thus preferred. We can see that in both equations the first term is related to the fit of the model, while the second term represents a complexity penalty, thereby formalizing the Occam's razor (Myung & Pitt, 1997). However, both AIC and BIC penalize complex models only by their number of parameters, neglecting their functional form complexity. Consequently, both criteria are not appropriate for selecting among models with inequality constraints.

In what follows, we discuss MDL-based model selection, which presents itself as an attractive alternative method because it overcomes all the aforementioned problems of LRT, AIC and BIC and is particularly appropriate for MPT model selection given its ability to fully capture model complexity.

The principle of minimum description length (MDL) originates from algorithmic coding theory in computer science. According to this principle, statistical modeling is viewed as data compression, and the best model is the one that compresses the data as tightly as possible. A model’s ability to compress the data is measured by the shortest code length with which the data can be coded with the help of the model. The resulting code length is related to generalizability such that the shorter the code length, the better the model generalizes (Grünwald, 2007; Grünwald, Myung & Pitt, 2005; Grünwald, 2000; Myung, Navarro & Pitt, 2006).

The Fisher Information Approximation (FIA: Rissanen, 1996) represents a formal implementation of the MDL principle for model selection. It gives the shortest code length
with which a model can code the data.\footnote{More precisely, it is a large sample approximation to normalized maximum likelihood (NML: Myung, Navarro & Pitt, 2006; Rissanen, 2001), which can be considered the shortest code length a model can achieve in coding the current data in the worst case of its true distribution. NML can be computationally intensive and will not be discussed in this paper.} This criterion is defined as

\[ FIA = -LML + C_{FIA} \]  

with

\[ C_{FIA} = \frac{S}{2} \ln \frac{N}{2\pi} + \ln \int \sqrt{|\mathbf{I}(\theta)|} \, d\theta \]  

where \( \mathbf{I}(\theta) \) is the Fisher information matrix (e.g. Casella & Berger, 2001) of sample size one with its elements given by \( (\mathbf{I}(\theta))_{ij} = -E \left[ \frac{\partial^2 \ln f(x_1|\theta)}{\partial \theta_i \partial \theta_j} \right] \).\footnote{It should be noted that Equation 7 is valid only when the model is globally identified, i.e., if different parameter values generate different category probabilities. For a non-identified model, one should reparameterize it into an equivalent but identified model before applying this formula.} A smaller criterion value indicates better generalization, and thus, the model that minimizes the criterion should be chosen.

There are several observations one can make about FIA. First, in this selection criterion, generalizability is measured as a trade-off between goodness of fit (LML) and complexity (\( C_{FIA} \)). Second, regarding the complexity measure of FIA, \( C_{FIA} \), its first term captures the effects of the number of parameters (\( S \)) and its second term captures the functional form effects through the Fisher information matrix (\( \mathbf{I}(\theta) \)). Especially, note that the functional form complexity in \( C_{FIA} \) is expressed as an integral over the parameter space. As such, it would therefore be straightforward to represent inequality constraints on parameters in \( C_{FIA} \) as the constraints simply reduce the size of the parameter space. Third, FIA is related to BIC in that mathematically both criteria can be viewed as approximations to minus...
twice the log marginal likelihood in Bayesian statistics when Jeffreys’ prior $\pi(\theta) \propto \sqrt{|I(\theta)|}$ is assumed, but FIA gives a better approximation than BIC (Grünwald, 2007, Chapter 8).

To summarize, FIA overcomes the practical and theoretical shortcomings of LRT in that the former is based on generalizability and can be applied to multiple, nested or non-nested models. Furthermore, the ability of MDL to capture functional form complexity and also to account for the effects of inequality parametric constraints provide MDL with unique advantages over LRT as well as AIC and BIC. Given the great variability in functional form complexity among MPT models, MDL is ideally suited for model selection among these models.

MDL Model Selection of MPT Models

A Computer Program for $C_{FIA}$ Computation

Applying the FIA criterion to MPT models requires the computation of $C_{FIA}$ in (7). As no analytic solution is available in general for the integral in the expression of $C_{FIA}$, the solution must be sought by numerical integration, which may be too cumbersome for most researchers who are interested in applying FIA. To help ease some of the computational burden, we have written a computer program that can be used to compute the complexity measure.

The general purpose MatLab program for computing the quantity $C_{FIA}$ for BMPT models will be available for download from the Psychonomic journal archive web site (http://www.psychnomi.prg/archive; see Archived Materials). The program evaluates the integral using a Monte Carlo algorithm. The technical details of this numerical integration algorithm are described in Wu, Myung and Batchelder (submitted).
program is general enough to compute the complexity of any BMPT model. Given that
every MPT model can be reparameterized into an equivalent BMPT model, the program is
applicable to all MPT models. The program can also incorporate inequality constraints on
parameters in so far as the constraints are of the form \( \theta_1 < \theta_2 < \cdots < \theta_k \) or its combinations.\(^5\)

The program assumes that the BMPT model has a single tree structure. When
\( K \) trees are present in one model, these trees should be combined into one single tree
with multinomial probabilities \( p_k = \frac{N_k}{N} \), where \( N_k \) is the sample size for tree \( k \) in the
experimental design and \( N \) is the total sample size. To illustrate how this is done, consider
model 1HTM shown in Figure 1 and suppose that the sample sizes for items from sources A,
B and N are 250, 250 and 500, respectively. The three trees should then be joined together
to form a single tree with numerical probabilities \( p_A = p_B = 0.25 \) and \( p_N = 0.5 \). Although
all three trees in 1HTM are BMPT models, the new tree we obtained by joining them is
not because there are three branches from the root node. It needs to be turned into a
BMPT model through reparameterization. To reparameterize, one first joins the two trees
for sources A and B to a single node with branching probabilities 0.5, and then joins the
resulting binary tree with the tree for new items with branching probabilities 0.5. This is
shown in Figure 3.

The \textit{MatLab} program involves a function \textsc{BMPTFIA} with six input argu-
ments and six output arguments: \texttt{function [CFIA,CI,lnInt,CI1,lnconst,CI2] =}
\texttt{BMPTFIA(s,parameters,ineq0,category,N,Sample)}. The input and output arguments

\(^5\)Knapp and Batchelder (2004) has shown that BMPT models with such inequality constraints can be
reparameterized into BMPT models without inequality constraints. Our program computes the integral of
the original model directly without invoking such reparameterization.
are described below.

The first input argument *s* is related to the string representation of the BMPT model as discussed earlier. It can be obtained by replacing all categories in the string by the capital letter “C” and all branching probabilities, including parameters and fixed constants, by the lower case letter “p”. For example, for model 1HTM shown in Figure 1, this argument should be *s* = “ppppCpCCppCCCcppCCppCCCcppCCC”.

The second input argument *parameters* is a row vector that assigns parameters or constants to the p’s in the string *s*. Its length should be the same as the number of p’s in *s*, and its elements correspond to the p’s according to their order in *s*. Positive integer elements in *parameters* assign parameters to the corresponding p’s, with the same integer denoting the same parameter. Constants are assigned to the p’s using the negation of their values. For model 5c with multinomial probabilities .25, .25 and .5 for sources A, B and N, respectively, this input argument should be *parameters*= [−0.5, −0.5, 1, 3, 5, 4, 5, 2, 3, 5, 4, 5, 4, 5]. In this vector, the five parameters (*D*1, *D*2, *d*, *b*, *g*) are coded using integers 1 through 5, respectively, and the first two elements of the vector (−0.5’s) are the probability constants we used to join the three trees into a single tree.

The third input argument *ineq0* assigns inequality constraints imposed on the parameters. It is a matrix with two columns. Each element denotes a parameter coded in the same way as in *parameters*. For each row, the parameter on the left column is constrained to be *smaller* than that on the right column. The number of rows is determined by the total number of simple inequality constraints of the form θ1 < θ2 in the model. For example, if we were to impose an inequality constraint *D*1 > *D*2 for model 5c, the matrix would be set to *ineq0* = [2, 1]. If no inequality constraints are assumed, we set it to an “empty matrix”,...
The fourth input argument `category` assigns categories to the C’s in the string `s` in the same way `parameters` assigns branching probabilities, except that only positive consecutive integers from 1 to `J`, the total number of categories, are allowed. For model 1HTM, this argument should be set to `category = [1, 1, 2, 1, 2, 3, 5, 4, 5, 4, 5, 6, 7, 8, 9]`. Note that with 3 different responses in each of the 3 conditions, there are 9 different categories in total.

Finally, the fifth input argument `N` specifies the total sample size and the last input argument `Sample` specifies the number of random samples to be drawn in the Monte Carlo algorithm.

The first output argument `CFIA` gives $C_{FIA}$. Given the stochastic nature of the Monte Carlo algorithm, the output value `CFIA` changes from one to another run of the program.

The second output argument `CI` gives the Monte Carlo confidence interval for `CFIA`. The rest four output arguments are optional. They are described in detail in the program file.

We now provide an example of the application of the computer program. Consider again 1HTM-5c in Figure 2 with an inequality constraint of $D_1 > D_2$ and sample sizes 250, 250 and 500 for the three kinds of stimuli. In a Monte Carlo run with `Sample = 200,000`, we obtained $CFIA = 12.6182$, $CI = [12.6113, 12.6251]$.

The `MatLab` program described above gives only the complexity term, $C_{FIA}$. As shown in (6), to obtain the value of FIA for a given MPT model, one must also compute the goodness of fit term, $-LML$. This term can be obtained from a user-friendly program called `GPT.EXE` (Hu & Phillips, 1999) that is available for free download from `http://www.xiangenhu.info/`. This program performs maximum likelihood estimation and
output best-fit parameter values and the value of $-LML$ for any MPT model with and without inequality constraints.

In what follows, we demonstrate the use of the MatLab program for model selection of MPT models in two different experimental paradigms: source monitoring and pair clustering.

Modeling Source Monitoring Data

Our first example concerns the source monitoring experiment of Rose et al. (1975, Experiment 1). The purpose of their study was to examine whether accurate source memory of language could occur at the semantic level of language processing. In their experiment, subjects studied a mixed list of English and Spanish sentences before being tested on recognition and source memory performance. Contextual relationships between the sentences were manipulated in the experiment such that in one condition, the sentences were semantically related to a common topic, whereas in the other condition, all sentences were semantically unrelated. Rose et al. (1975) reasoned that if language source information is available at the semantic level of processing, because the contextual relationship among the sentences makes them semantically less distinguishable, language source memory would be poorer for related sentences than for unrelated sentences.

Based on an analysis of variance of the data, Rose et al. (1975) concluded that there was no significant difference between the two treatment conditions. Batchelder and Riefer (1990, Tables 7 and 8) reanalyzed the same data with model 1HTM-4, as shown in Figure 2, that can distinguish the effect of item detection from that of source discrimination. Their LRT results suggested that recognition memory ($D$) was significantly poorer for related than
for unrelated sentences, but there was no significant difference in source monitoring (d).

Based on this finding, Batchelder and Riefer (1990) concluded that contextual relationship
is detrimental to item detection but not to source discrimination, indicating that language
source information is not available at the semantic level.

We re-analyze the data as reproduced in Batchelder and Riefer (1990, Table 7) in a
model selection framework using MDL, as well as AIC and BIC. The results are shown in
Table 1. The first model $M_0$ has 8 parameters, four ($D, d, b, g$) for each treatment condition,
without any equality constraints. A total of 15 MPT models were further created by impos-
ing various equality constraints on the parameters. For the rest of the models, a subscript
notation is used to indicate how the parameters are constrained to be equal across the two
treatment conditions, related and unrelated. For example, in $M_{Dd}$, both $D$ and $d$ are set to
equal across the two conditions whereas $b$ and $g$ are allowed to vary across the conditions. In
addition to such equality constraints, inequality constraints are also considered for models
where either of the two parameters, $D$ and $d$, is allowed to differ between the two conditions
because theories underlying those models predict a particular direction of difference: both
parameters are expected to be smaller for related sentences than for unrelated sentences.

No inequality constraints on the guessing parameters, $b$ and $g$, are imposed even if they
are allowed to differ across the two conditions. Because parameter estimates do not viol-
ate these inequality constraints in the data, incorporating inequality constraints does not
change the values of LML, AIC and BIC. In contrast, inequality constraints do change the
value of $C_{FIA}$ and thus the value of FIA. In Table 1, the latter two values obtained under
inequality constraints are denoted by $C'_{FIA}$ and $FIA'$.

From the table we can observe that as the number of parameters decreases, $-LML$
increases while $C_{FIA}$ decreases, as expected, exhibiting a tradeoff between goodness of fit and complexity. Of particular interest is the observation that among models with the same number of parameters, $C_{FIA}$ varies greatly, indicating the effects of functional form on complexity. The difference in complexity due to functional form between two models can sometimes be even greater than the difference in goodness of fit. The case in point is the comparison between $M_d$ and $M_g$. In terms of $-LML$, $M_d$ fits better the data than $M_g$ (36.18 vs 36.61) but is more complex ($C_{FIA} = 20.7 > C_{FIA} = 19.7$), thus yielding an overall larger FIA value than $M_g$ (56.9 vs. 56.3). As a result, $M_g$ is preferred to $M_d$ under FIA, though the latter would be selected if we were to treat the two models equally complex as in AIC and BIC. On the same token, inequality constraints can have similar effects on model complexity. For example, $M_{Ddbg}$ ($FIA = 55.2$) is preferred to $M_{bg}$ ($FIA = 55.6$) if no inequality constraints on $b$ and $g$ in $M_{bg}$ are considered, but the preference is reversed if the constraints are considered ($FIA' = 54.2$ for $M_{bg}$).

Turning the discussion to model selection, we first consider the results in Table 1 obtained when no inequality constraints are considered. FIA selects $M_{dbg}$ with $FIA = 53.8$ as the best model among the 16 models under consideration, so does BIC. On the other hand, AIC selects $M_{dg}$, a model with one additional parameter. Now let us consider the inequality constraints. Obviously, there would be no changes in the conclusion for AIC and BIC as inequality constraints do not change the fit of the models for this data set and the complexity measures in both criteria are “blind” to inequality constraints. FIA still favors $M_{dbg}$ with inequality constraints. In summary, among a total of 32 models compared including the ones with inequality constraints, we conclude that $M_{dbg}$ with inequality constraints is the best generalizing model of all.
The model selection analysis discussed so far is conducted for models obtained by considering all possible combinations of constraints, equality and inequality, on the four parameters \( (D, b, g, b) \). In addressing the theoretical issue raised in Rose et al. (1975) and Batchelder and Riefer (1990), however, one only needs to consider the two parameters \( D \) and \( d \) of main interest. Under this more focused scope, there are four relevant models to compare: \( M_0, M_D, M_d \) and \( M_{Dd} \), along with their inequality constraints. Among these four, \( M_d \) is favored under all three criteria, AIC, BIC and FIA. According to this best generalizing model, the two treatment conditions with related and unrelated sentences, differ in item recognition \( (D) \) but not in source monitoring \( (d) \). An implication of this conclusion is that semantic information is useful in the recognition task but does not include any information about language source.

Essentially the same conclusion as ours was reached by both Rose et al. (1975) and Batchelder and Riefer (1990). Although this particular example may be somewhat disappointing, one should note that generally speaking, FIA-based model selection analysis allows one to entertain and evaluate all types of theoretical hypotheses of interest, and that the effort is in turn likely to generate much richer and deeper insights into the underlying cognitive processes than analyses based on traditional methods such as LRT and analysis of variance.

**Modeling Pair-clustering Data**

Our second example demonstrating the application of FIA based model selection is related to the pair clustering experiment of Batchelder and Riefer (1980, experiment 1a). The purpose of this experiment was to examine the effects of within-category spacing on
recall performance. They hypothesized that a small lag between a pair of categorically related words facilitates the formation and storage of a pair-cluster whereas a large lag facilitates the retrieval process. In the experiment, participants first studied a word list that consisted of both paired words and singletons and were then tested in a free recall task. Paired words were two words that were categorically related. In the word list, each pair of words occupied positions that were separated by $J = 0, 4, 12, 24$ words unrelated to the pair. Five trials of this study-recall cycle were repeated. The data set we reanalyze in this paper was from Batchelder and Riefer (1986, Table 1).

Batchelder and Riefer (1986) used LRTs to analyze the data with the pair-clustering MPT model. The original version of this model is shown in Figure 4. The model assumes three parameters: $c$, the probability of pairs being clustered and stored in memory; $r$, the (conditional) probability of a stored pair being retrieved from memory; $u$, the probability of a single item being stored and retrieved from memory for either pairs or singletons. Accordingly, response category $E_1$ indicates recalling adjacently both items of a studied pair, $E_2$ indicates recalling non-adjacently both items of a pair, $E_3$ indicates recalling only one item, $E_4$ indicates recalling neither items in a pair, and finally, $F_1$ and $F_2$ indicate successful and unsuccessful recall of a singleton, respectively. Because there were four conditions for category pairs in the experiment, the MPT model for each trial consists of four trees for category pairs, one for each lag condition, and another tree for singletons. If the parameter $u$ is assumed to be different for pairs and singletons and for different lag conditions, there will be 13 parameters for each trial, with 12 for category pairs and 1 for singletons, and 65 parameters for the entire five trials. A typical pair-clustering model assumes a single $u$, thereby reducing the number of parameters to 9 for each trial and 45
for the entire data set.

Results from separate LRTs for the five trials performed by Batchelder and Riefer (1986) showed that parameter \( c \) was significantly different across lag conditions for data from all five trials. However, \( r \) was significant only for trials 1 and 3, but not for the other trials. Another LRT with data from all five trials combined revealed that \( c \) was significantly different across lag conditions, but \( r \) was only marginally significant. These results, taken together, supported the hypothesis that small lag between category pairs facilitates the formation and storage of pair clusters, but offered only marginal support for the hypothesis that long lag facilitates the retrieval of pairs. Further and importantly, given that boundary maximum likelihood estimates have been obtained due to the inequality constraints when fitting the model, the LRTs results would be uninterpretable.

Now we re-analyze the data by FIA based model selection. Table 2 summarizes the results. There are eight models to be compared. \( M_0 \) is the model with 65 parameters described above. \( M_u \) with 45 parameters assumes a single \( u \) for each trial as in typical pair clustering models. From this model, various equality constraints on \( c \) and \( r \) are applied to form the rest of models. Models \( M_{ur} \) (\( M_{uc} \)) assumes the same \( r \) (\( c \)) across the four conditions. In \( M_{ucr} \), both \( c \) and \( r \) are assumed to be the same throughout the lag conditions. The three “primed” models, \( M_u' \), \( M_{ur}' \) and \( M_{uc}' \), differ from the un-primed ones in that in the former, additional inequality constraints are imposed on the relevant parameters across the four lag conditions, such as \( c_{J=0} > c_{J=4} > c_{J=12} > c_{J=24} \) or \( r_{J=0} < r_{J=4} < r_{J=12} < r_{J=24} \), or both, to embody the theoretical hypotheses concerning the order of those parameters. Such constraints do not change the number of parameters in the model but they may lead to a larger \(-LML\) value as the maximum likelihood is searched over the restricted and thus
smaller parameter space. This is indeed observed in Table 2 for all three pairs of models. For the same reason, inequality constraints reduce model complexity.

From Table 2, one can observe the trade-off between goodness of fit and model complexity; the smallest $-LML$ value of 141.3 is achieved by the most complex model $M_0$ with $C_{FIA} = 137.0$. At the other end of the complexity spectrum, $M_{ucr}$ with the fewest number of parameters (15) gives the largest $-LML$ value (206.2) and the smallest $C_{FIA}$ value (43.9). We also note that models with the same number of parameters can differ greatly in their complexity. For example, the four models, $M_{uc}$, $M'_{uc}$, $M_{ur}$ and $M'_{ur}$, all have 30 parameters yet their $C_{FIA}$ complexity varies from the lowest 59.9 to the largest 79.5, due to the combination of tree structure and inequality constraints on parameters. Such complexity difference between two models with the same number of parameters can be larger than the difference in LML, thus affecting model selection results. Such a pattern of result is indeed observed in the table. That is, among the same four models with 30 parameters, both AIC and BIC select $M_{ur}$ whereas FIA picks $M'_{ur}$. This is because although $M_{ur}$ fits the data better than $M'_{ur}$ ($-LML = 168.0$ vs 169.5), the model is more complex ($C_{FIA} = 79.5$ vs 63.6) and thus yields a larger FIA than its counterpart ($FIA = 247.5$ vs 233.1).

Turning the discussion to model selection, among the eight models compared, AIC favors the 30-parameter $M_{ur}$ and BIC favors the most restrictive model $M_{ucr}$ whereas FIA selects $M'_{u}$. Note that the model $M'_{u}$ imposes inequality constraints on both $c$ and $r$ in the directions consistent with the experimental hypotheses of Batchelder and Riefer (1980). In other words, our FIA-based reanalysis of the data supports the hypotheses in their ordered form. As discussed earlier, the LRT results by Batchelder and Riefer (1986) indicated that the hypothesized within-pair spacing effect on parameter $r$ was inconclusive while the
hypothesized effect on parameter $c$ was supported. The specific order relationships among the hypotheses were not, however, examined. Model $M_{ur}'$ embodies this interpretation of the data and interestingly, turns out to be the second best model after $M_{ur}$. Especially, if the three models with inequality constraints were not among the competing models, FIA would choose $M_{ur}$, leading to a different conclusion. This shows that the hypothesized order relationship of parameters, which may restrict the parameter space and reduce the complexity of the model, can lead to different model selection conclusions, and as such, should not be neglected.

To summarize, we demonstrated the application of the FIA-based model selection approach for selecting among MPT models of pair-clustering for the Batchelder and Riefer (1980) data set. The flexibility of the approach allowed us to construct and test a variety of MPT models including models with inequality constraints on parameters. We compared the results from our analysis to those from the LRT-based analysis of the same data reported in Batchelder and Riefer (1986). By and large, the same scientific conclusions were drawn from either analysis, though our FIA-based analysis provides more definitive support for the hypotheses in their ordered form originally formulated and tested in Batchelder and Riefer (1980,1989).

Conclusion

Multinomial processing tree modeling represents a theoretically motivated and statistical justified methodology for evaluating cognitive capacities for various experimental paradigms. Selecting among different MPT models is especially important both in addressing theoretical issues and in validating an MPT model in a particular experimental
paradigm. In this paper we have introduced the MDL based model selection method to the
practitioners of MPT modeling. Especially, to facilitate the use of this new methodology, we
provide a general purpose computer program in *MatLab* that can be exploited to compute
FIA for any MPT model. Two example applications of MDL model selection with real data
sets selected from different experimental paradigms are also discussed.

MDL’s flexibility of application to a wide range of model comparison situations that
may arise in MPT modeling makes it an attractive alternative to traditional methods such as
LRT, AIC, and BIC. First, instead of using a series of null hypothesis significance tests such
as LRTs, MDL represents a *model selection* approach in which the models in contention are
ranked by their generalizability, or, equivalently, predictive accuracy, which is the hallmark
of model selection. Also, unlike AIC and BIC, MDL considers the effects of the number
of parameters and sample size on model complexity but also importantly, the effect of
functional form, which alone can significantly contribute to complexity and sometimes even
more so than the number of parameters. As a result, another advantage factor of MDL
over the other three methods is that MDL allows one to incorporate inequality constraints
on model parameters. Last but not least, with the freely available *MatLab* program, FIA is
now entirely within the reach of everyday practitioners.

As it is usually the case with any new methodology, MDL as presented in this pa-
er is not without shortcomings. For example, it does not address the issues of model
misspecification and individual differences. To allow for model misspecification, exact
equality constraints should be replaced by fuzzy equality constraints as done in an ele-
gant sampling-based method known as Population-Parameter Mapping (PPM) proposed
by Chechile (1998). Regarding individual differences, one way of incorporating this im-
important factor is through hierarchical modeling in which parameter values corresponding
different individuals are assumed to be sampled from a common distribution (see, e.g.,
Klauer, 2006; Smith & Batchelder, in press). Although it is possible in theory to address
these issues within the MDL framework, it is beyond the scope of the present work.

In conclusion, model selection lies at the core of the scientific inference process. Ac-
cordingly, a theoretically well-justified and widely applicable methodology can help advance
science. We believe that MDL represents such a methodology that provides versatile yet
powerful tools for assessing the validity of MPT models in a way that goes beyond the
shortcomings of the current methods such as LRT, AIC, and BIC.

One final note. It is important to note that statistical model selection techniques
alone, however sophisticated, are not a panacea for all inference problems. Other non-
statistical means of model evaluation such as plausibility, interpretability, and explanatory
adequacy are equally, if not more, important. Instead of automatic tools implemented in
softwares, statistical model selection methods can be most useful if it is combined with the
judicious use of sound subjective but scientific judgement.

Archived Materials

The following materials [and links] associated with this article may be ac-
cessed through the Psychonomic Societys Norms, Stimuli, and Data archive,
www.psychonomic.org/archive.

To access these files [or links], search the archive for this article using the journal
name (Psychonomic Bulletin & Review), the first authors name (Wu), and the publication
year (to be filled in if accepted).
FILE: ????.zip

DESCRIPTION: The compressed archive file contains one file, BMPTFIA.m, a Mat-
Lab m-file that computes the FIA complexity $C_{FIA}$ for general BMPT models.
MULTINOMIAL PROCESSING TREE MODELS

References


MULTINOMIAL PROCESSING TREE MODELS


MULTINOMIAL PROCESSING TREE MODELS

1 Schweickert, R. & Chen, S. (2008). Tree inference with factors influencing processes in a processing
tree. *Journal of Mathematical Psychology*, 52, 158-183

2 Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate anal-

addressing individual differences. *Journal of Mathematical Psychology*


ogy*, 50, 99-100.

logical Methods & Research*, 27, 359-397

7 Wu, H., Myung, J. I. & Batchelder, W. H. (submitted). On the minimum description length com-
plexity of multinomial processing tree models. *Journal of Mathematical Psychology*

Acknowledgements

This paper is based on Hao Wu’s Master of the Arts thesis submitted to the Ohio State
University in July 2006. It is supported in part by National Institute of Health Grant R01-
MH57472 to JIM. Work on this paper by the third author was supported in part from a grant
from the National Science Foundation: SES-0616657 to X. Hu and W. H. Batchelder(Co-
PIs). We wish to thank Mark A. Pitt and Michael W. Browne for valuable feedbacks
provided for the project. Correspondence concerning this article should be addressed to
Hao Wu, Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus, OH 43210. E-mail: wu.498@osu.edu. Tel: 614-292-5510
Table 1: Summary of model selection results for source monitoring data from Rose et al. (1975, experiment 1). The data can be found in Batchelder and Riefer (1990, Table 7). See the main text for the description of the models. It should be noted that inequality constraints apply only to parameters $D$ and $d$ (but not to $b$ or $g$) when the corresponding equality constraints are not present. The row $FIA'$ shows FIA values if inequality constraints are taken into account, while the $FIA$ gives FIA values if those constraints are neglected. The total sample size is $N = 1920$.

<table>
<thead>
<tr>
<th>Models</th>
<th>$M_0$</th>
<th>$M_q$</th>
<th>$M_b$</th>
<th>$M_d$</th>
<th>$M_D$</th>
<th>$M_bg$</th>
<th>$M_dg$</th>
<th>$M_Dg$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$-LML$</td>
<td>36.17</td>
<td>36.61</td>
<td>38.23</td>
<td>36.18</td>
<td>41.16</td>
<td>38.66</td>
<td>36.61</td>
<td>41.60</td>
</tr>
<tr>
<td>$C_{AIC}$</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$AIC$</td>
<td>44.17</td>
<td>43.61</td>
<td>45.23</td>
<td>43.18</td>
<td>48.16</td>
<td>44.66</td>
<td>42.61</td>
<td>47.60</td>
</tr>
<tr>
<td>$C_{BIC}$</td>
<td>66.41</td>
<td>63.07</td>
<td>64.69</td>
<td>62.64</td>
<td>67.62</td>
<td>61.34</td>
<td>59.29</td>
<td>64.28</td>
</tr>
<tr>
<td>$BIC$</td>
<td>22.2</td>
<td>19.7</td>
<td>19.4</td>
<td>20.7</td>
<td>20.4</td>
<td>16.9</td>
<td>18.2</td>
<td>17.8</td>
</tr>
<tr>
<td>$FIA$</td>
<td>58.4</td>
<td>56.3</td>
<td>57.6</td>
<td>56.9</td>
<td>61.6</td>
<td>55.6</td>
<td>54.8</td>
<td>59.4</td>
</tr>
<tr>
<td>$FIA'$</td>
<td>20.9</td>
<td>18.4</td>
<td>18.0</td>
<td>20.0</td>
<td>19.7</td>
<td>15.5</td>
<td>17.5</td>
<td>17.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>$M_db$</th>
<th>$M_Db$</th>
<th>$M_Dd$</th>
<th>$M_dbg$</th>
<th>$M_Dbg$</th>
<th>$M_Ddb$</th>
<th>$M_Ddbg$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>$-LML$</td>
<td>38.30</td>
<td>41.84</td>
<td>41.55</td>
<td>38.73</td>
<td>42.28</td>
<td>41.98</td>
<td>42.31</td>
</tr>
<tr>
<td>$C_{AIC}$</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>$AIC$</td>
<td>44.30</td>
<td>47.84</td>
<td>47.55</td>
<td>43.73</td>
<td>47.28</td>
<td>46.98</td>
<td>47.31</td>
</tr>
<tr>
<td>$C_{BIC}$</td>
<td>22.68</td>
<td>22.68</td>
<td>22.68</td>
<td>18.90</td>
<td>18.90</td>
<td>18.90</td>
<td>18.90</td>
</tr>
<tr>
<td>$BIC$</td>
<td>60.98</td>
<td>64.52</td>
<td>64.23</td>
<td>57.63</td>
<td>61.18</td>
<td>60.88</td>
<td>61.21</td>
</tr>
<tr>
<td>$C_{FIA}$</td>
<td>17.8</td>
<td>17.3</td>
<td>18.5</td>
<td>15.1</td>
<td>14.6</td>
<td>15.8</td>
<td>15.3</td>
</tr>
<tr>
<td>$FIA$</td>
<td>56.1</td>
<td>59.1</td>
<td>60.1</td>
<td>53.8</td>
<td>56.9</td>
<td>57.8</td>
<td>57.6</td>
</tr>
<tr>
<td>$C_{FIA}'$</td>
<td>17.1</td>
<td>16.6</td>
<td>18.5</td>
<td>14.4</td>
<td>13.9</td>
<td>15.8</td>
<td>15.3</td>
</tr>
<tr>
<td>$FIA'$</td>
<td>55.4</td>
<td>58.4</td>
<td>60.1</td>
<td>53.1</td>
<td>56.2</td>
<td>57.8</td>
<td>57.6</td>
</tr>
</tbody>
</table>
Table 2: Summary of model selection results for pair-clustering data from Batchelder and Riefer (1980, experiment 1a). See the main text for the description of the models. The total sample size is $N = 3220$.

<table>
<thead>
<tr>
<th>Models</th>
<th>$M_0$</th>
<th>$M_u$</th>
<th>$M'_u$</th>
<th>$M_{uc}$</th>
<th>$M'_{uc}$</th>
<th>$M_{ur}$</th>
<th>$M'_{ur}$</th>
<th>$M_{ucr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>65</td>
<td>45</td>
<td>45</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>$-LML$</td>
<td>141.3</td>
<td>155.2</td>
<td>157.7</td>
<td>196.5</td>
<td>200.5</td>
<td>168.0</td>
<td>169.5</td>
<td>206.2</td>
</tr>
<tr>
<td>$C_{AIC}$</td>
<td>65</td>
<td>45</td>
<td>45</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>$AIC$</td>
<td>206.3</td>
<td>200.2</td>
<td>202.7</td>
<td>226.5</td>
<td>230.5</td>
<td>198.0</td>
<td>199.5</td>
<td>221.2</td>
</tr>
<tr>
<td>$C_{BIC}$</td>
<td>262.51</td>
<td>181.74</td>
<td>181.74</td>
<td>121.16</td>
<td>121.16</td>
<td>121.16</td>
<td>121.16</td>
<td>60.58</td>
</tr>
<tr>
<td>$BIC$</td>
<td>403.8</td>
<td>337.0</td>
<td>339.5</td>
<td>317.6</td>
<td>321.7</td>
<td>289.1</td>
<td>290.7</td>
<td>266.7</td>
</tr>
<tr>
<td>$C_{FIA}$</td>
<td>137.0</td>
<td>106.1</td>
<td>73.7</td>
<td>75.8</td>
<td>59.9</td>
<td>79.5</td>
<td>63.6</td>
<td>43.9</td>
</tr>
<tr>
<td>$FIA$</td>
<td>278.3</td>
<td>261.3</td>
<td>$231.4$</td>
<td>272.3</td>
<td>260.4</td>
<td>247.5</td>
<td>233.1</td>
<td>250.1</td>
</tr>
</tbody>
</table>
Figure 1. The one-high-threshold (1HT) multinomial processing tree model of source monitoring. The parameters are defined as follows: $D_1$ (detectability of source A items); $D_2$ (detectability of source B items); $d_1$ (source discriminability of source A items); $d_2$ (source discriminability of source B items); $a$ (guessing that a detected but nondiscriminated item belongs to source A); $b$ (guessing “old” to a nondetected item); $g$ (guessing that a nondetected item biased as old belongs to source A category). Adapted from Batchelder and Riefer (1990).
Figure 2. The nested hierarchy of eight versions of the 1HTM model in Figure 1, created by imposing successive constraints on the parameters. In the figure, the model parameters for each model are listed and a directed arrow from one model to another means that the second model is nested in the first.
Figure 3. Example of combining the three processing trees in the 1HTM (shown in Figure 1) into one BMPT model. The sample sizes are assumed to be 250, 250 and 500 for source A items, source B items and new items, respectively.
Figure 4. Batchelder and Riefer’s (1999) multinomial processing tree model of pair-clustering.