Journal of Mathematical Psychology 57 (2013) 196–214

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

# A novel method for assessing rival models of recognition memory

Richard A. Chechile

*Psychology Department, Tufts University, Medford, MA 02155, United States*

## HIGHLIGHTS

- Contrast multinomial processing tree (MPT) and signal detection models for recognition.
- Provide evidence of mixtures being present on foil test trials.
- Develop a new MPT model for recognition memory.

## ARTICLE INFO

*Article history:*
Received 19 September 2012
Received in revised form
8 July 2013
Available online 16 August 2013

*Keywords:*
Models of recognition memory
Mixture models
Hazard functions
Signal detection
Multinomial-processing tree modeling

## ABSTRACT

A general comparison is made between the multinomial processing tree (MPT) approach and a strength-based approach for modeling recognition memory measurement. Strength models include the signal-detection model and the dual-process model. Existing MPT models for recognition memory and a new generic MPT model, called the Multistate (MS) model, are contrasted with the strength models. Although the ROC curves for the MS model and strength model are similar, there is a critical difference between existing strength models and MPT models that goes beyond the assessment of the ROC. This difference concerns the question of stochastic mixtures for foil test trials. The hazard function and the reverse hazard function are powerful methods for detecting the presence of a probabilistic mixture. Several new theorems establish a novel method for obtaining information about the hazard function and reverse hazard function for the latent continuous distributions that are assumed in the strength approach to recognition memory. Evidence is provided that foil test trials involve a stochastic mixture. This finding occurred for both short-term memory procedures, such as the Brown–Peterson task, and long-term list-learning procedures, such as the paired-associate task. The effect of mixtures on foil trials is problematic for existing strength models but can be readily handled by MPT models such as the MS model. Other phenomena, such as the mirror effect and the effect of target-foil similarity, are also predicted accurately by the MPT modeling framework.



## 1. Introduction

When it comes to the measurement of recognition memory, there are two distinctly different theoretical frameworks; one framework is exemplified by strength-based recognition memory models, and the other framework is exemplified by multinomial processing tree (MPT) models. See Malmberg (2008) and Yonelinas (2002) for extensive reviews of recognition memory from the perspective of strength-based models, and see Batchelder and Riefer (1999) and Erdfelder et al. (2009) for general reviews of the MPT literature that include models of recognition memory. Strength models and MPT models have different goals and are based on different theoretical characterizations of event memory. Although the strength-based models are designed specifically for recognition memory tests, a subset of the MPT models have been designed for recognition memory, and in those cases the two frameworks are rival models for recognition memory. Both systems are based on categorical data, i.e., multinomial data. Both measurement frameworks have important parameters to measure the quality of memory at the time of test, and both frameworks provide a method for correcting for subject response biases. Although the receiver operator characteristic (ROC) curves are not equivalent for the two approaches, these curves can look so similar that it is challenging to discriminate between the two frameworks (more will be devoted to this topic later).[1] Despite the high degree of similarity of these two measurement frameworks, there is one potential difference between the frameworks, and this difference is linked to the theoretical representation of foil items. In this paper some new formal results from hazard function analysis and reverse hazard function analysis are used to assess the

[1] In this paper the ROC is taken to be the plot of the hit rate versus the false alarm rate, and the z-ROC is taken to be a plot of the linearizing transformations of the hit and false alarm rates.

*E-mail address:* Richard.Chechile@Tufts.edu.

0022-2496/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmp.2013.07.002

two theoretical frameworks for recognition memory by means of a detailed examination of foil items.

In this paper strength-based and selective MPT alternatives for recognition memory are briefly reviewed in Section 2. Also in this section, the topic of mixture processes is discussed and linked to some of the differences between the memory measurement systems. In Section 3, the general topic of mixture detection is discussed in terms of density and hazard methods. In Section 4, several new theorems are developed that provide a new technique for the detection of mixtures. Also in this section the statistical issues for applying the new methods are discussed. In Section 5, the results from a number of experiments are examined with the new theoretical tools for uncovering mixtures. Finally, in Section 6, the implications of the new results are discussed within the broader context of recognition memory phenomena.

## 2. Strength and selective MPT recognition models

### 2.1. Differences in the mathematical representation

The memory for an event is represented differently in strength-based models and MPT models. The different characterizations of information parallel an important distinction drawn in psychophysics between *prothetic* and *metathetic* continua (Stevens, 1957, 1961). Stevens (1961, p. 41) observed the following distinction about loudness (a prothetic continuum) and pitch (a metathetic continuum):

> … it is interesting that some of the better known prothetic continua seem to be mediated by an *additive* mechanism at the physiological level, whereas the metathetic continua appear to involve *substitutive* processes at the physiological level. Thus we experience a change in loudness when excitation is added to excitation already present on the basilar membrane, but we note a change in pitch when new excitation is substituted for excitation that has been removed, i.e., the pattern of excitation is displaced

The Steven's distinction stresses the inherent difference between changes in intensity and changes in qualities. Clearly pitch and hue are a frequency issue as a physicist looks at it; whereas, loudness and brightness are an amplitude issue. There can be graduated changes in either frequency or amplitude from a physical perspective. Stevens, however, is stressing the point that the physiological responses are quite different in the two cases. For example, in the Békésy (1960) place theory for hearing, the frequency of a sound corresponds to the specific hair cell detectors that are activated—different frequencies activate different hair cell detectors. However, the loudness of a sound at some specific frequency is captured by the number of hair cells activated at that given frequency and by the firing rate of those hair cells. Hence, loudness has this additive property but frequency is characterized by the specific detectors activated. Also for some metathetic attributes, there is not an order associated with the attribute. For example, hues are not perceived as an ordered dimension, but brightness is clearly additive, and brightnesses can be ordered. The difference in the intensity of a prothetic dimension and the qualitative difference of metathetic stimuli parallels roughly the differences between a strength-based memory measurement framework and the MPT approach to memory measurement. The strength approach treats the memory of an encoded event in terms of a prothetic continuum; whereas MPT memory models choose to measure the probability that the metathetic, informational content of the event is sufficiently preserved in memory. In SDT, stimuli are characterized as an intensity change (e.g., foils are weak and targets are stronger by an additive amount). The differences among the various targets are largely neglected with the SDT approach. Any difference among the stimuli would have to be in terms of strength, so in principle with the SDT approach, the target stimuli have an ordering. However, with the MPT approach, there is specific and different informational content initially encoded for each stimulus. The stimuli on a target list are different from one another like the various colors are different.

#### 2.1.1. Selective MPT models and their probability measures

Let us consider first the rationale for the MPT framework. A memory target has rich informational content, which includes the perceptually detected information as well as the cognitive elaborations of the target event. Even within the narrow class of stimuli that are typical of a well controlled memory experiment, the encodings are widely varied and qualitatively different. For example, consider a consonant triad target, such as LRH. Clearly the individual letters and their position within the triad are distinctly different in terms of informational content. Moreover, individuals can cognitively elaborate the stimulus in very different ways. Based on post-experimental interviews, Chechile (1973) made note of some of these encoding strategies. For example, one subject reported that she pictured each briefly presented auditory triad as letters being written on a blackboard. Another subject associated each stimulus with an item in his company's stock inventory. One subject reported that LRH was encoded as "Linda Rides Horses" along with an image of a person on a horse. Other subjects kept the memory encoding at the auditory level. The subjects also reported some variability from trial-to-trial in their encoding strategy.

Because the target event is a complex pattern of information that is encoded in an idiosyncratic fashion, MPT modelers have chosen to measure the surviving memory at the time of test in terms of probabilities over some mutually exclusive categories (Batchelder, Hu, & Riefer, 1994; Batchelder & Riefer, 1980; Chechile, 2004; Chechile & Meyer, 1976; Chechile, Sloboda, & Chamberland, 2012; Chechile & Soraci, 1999). Although there are important differences among these specific MPT models, each model conceptualizes target storage into mutually exclusive categories.

For some of the above MPT models, the information about the target event is either sufficiently stored, or it is deficient in some fashion, so even if the subject has access to the deficient information, it would not be enough to support the correct recall of the entire target.[2] Targets in the sufficiently stored category could be encoded in many different ways, but in each case all the target information is still in the memory system. This two-part classification was used by Batchelder and Riefer (1980), Chechile (2004) and Chechile and Meyer (1976).[3] Items that are sufficiently stored might have variability in strength, but those differences are ignored with these MPT models because the storage is sufficient to support correct recall. Similarly, items that are insufficiently stored might have variability, but these differences are also ignored. The two mutually exclusive categories of sufficient storage versus insufficient storage are measured by the respective probabilities, i.e., $\theta_S$ and $1 - \theta_S$. For the Chechile (2004) and Chechile and Meyer (1976) models, it is assumed that on an old recognition test, the subject will be highly confident and correct if there is sufficient storage, but with insufficient storage there might be either a yes or no response, and the confidence might not be as high. These MPT

2 The recall test context is one where the subject must produce the item from memory without the benefit of seeing the item again. It is assumed that targets are complex, so the probability of a correct recall given incomplete storage is approximately zero.

3 The Batchelder and Riefer (1980) model is based on free recall data of clusterable pairs. Although this model discussed storage and retrieval for pairs of items, it is not a model for recognition memory. It is included here because storage for a cluster is categorized as opposed to being treated as a continuous variable of strength.