



# Using a multinomial tree model for detecting mixtures in perceptual detection

Richard A. Chechile \*

Psychology Department, Tufts University, Medford, MA, USA

**Edited by:**

Joseph W. Houpt, Wright State University, USA

**Reviewed by:**

Richard Schweickert, Purdue University, USA

Noah H. Silbert, University of Cincinnati, USA

**\*Correspondence:**

Richard A. Chechile, Psychology Department, Tufts University, Psychology Building, 490 Boston Av., Medford, MA 02155, USA  
e-mail: richard.chechile@tufts.edu

In the area of memory research there have been two rival approaches for memory measurement—signal detection theory (SDT) and multinomial processing trees (MPT). Both approaches provide measures for the quality of the memory representation, and both approaches provide for corrections for response bias. In recent years there has been a strong case advanced for the MPT approach because of the finding of stochastic mixtures on both target-present and target-absent tests. In this paper a case is made that perceptual detection, like memory recognition, involves a mixture of processes that are readily represented as a MPT model. The Chechile (2004) 6P memory measurement model is modified in order to apply to the case of perceptual detection. This new MPT model is called the Perceptual Detection (PD) model. The properties of the PD model are developed, and the model is applied to some existing data of a radiologist examining CT scans. The PD model brings out novel features that were absent from a standard SDT analysis. Also the topic of optimal parameter estimation on an individual-observer basis is explored with Monte Carlo simulations. These simulations reveal that the mean of the Bayesian posterior distribution is a more accurate estimator than the corresponding maximum likelihood estimator (MLE). Monte Carlo simulations also indicate that model estimates based on only the data from an individual observer can be improved upon (in the sense of being more accurate) by an adjustment that takes into account the parameter estimate based on the data pooled across all the observers. The adjustment of the estimate for an individual is discussed as an analogous statistical effect to the improvement over the individual MLE demonstrated by the James–Stein shrinkage estimator in the case of the multiple-group normal model.

**Keywords:** signal detection theory, multinomial processing tree models, perceptual learning, mixture detection, shrinkage estimators

## 1. INTRODUCTION

The title of this special issue implies two very different questions. The first question is: how should perceptual decision-making be modeled? The second question is: how should individual differences be estimated? This paper addresses both of these questions from a perspective that has been informed by research in the area of model-based memory measurement. The recommendations from this perspective result in some novel techniques for examining perceptual detection data.

Signal detection theory (SDT) is the classic method for measuring the perceived strength of a stimulus (Tanner and Swets, 1954; Green and Swets, 1966). The original applications of SDT typically dealt with cases of detecting the presence of a slight intensity increase on a single sensory dimension such as the loudness of white noise or an increase in the brightness of a color patch. The data from these studies are multinomial frequencies that are used to estimate either a signal sensitivity measure ( $d'$ ) associated with the separation between two presumed distributions on a psychological strength continuum, or a non-parametric measure such as  $A'$  associated with the area under the receiver-operator characteristic (ROC) curve. For such

applications there has been a general consensus that SDT is valid, accurate and useful. SDT has also been extended to the case of multiple dimensions (e.g., Ashby and Townsend, 1986).

Egan (1958) first noted that the target-present versus target-absent test trials used in a yes/no recognition memory study correspond to the signal-present versus signal-absent tests used in a sensory-based signal detection task. It therefore followed that SDT provided a method for measuring memory strength. In fact Macmillan and Creelman (2005) observed that contemporary applications of SDT in the memory area outnumbered the psychophysical applications. Malmberg (2008) and Yonelinas (2002) provide extensive reviews of recognition memory from the perspective of strength-based SDT models. Yet despite the widespread use of the SDT approach toward recognition memory measurement, there also has been substantial criticism of this approach (Chechile, 1978, 2013; Bröder and Schütz, 2009; Kellen et al., 2013). These critics argue instead for the use of multinomial process tree (MPT) models for a variety of reasons. MPT models have a number of desirable statistical properties and can result in measurements of important latent cognitive processes. For example Chechile and Meyer (1976) first used MPT models

for recognition memory data as well as recall data in order to obtain separate probability measures for trace storage and for the retrieval of stored traces, because forgetting was more suitably described in terms of either storage failures or retrieval failures rather than simply a change in “memory strength.” The implicit-explicit separation (IES) model is another example of a MPT model rather than a SDT model for memory (Chechile et al., 2012). With the IES model separate probability measures are estimated for explicit storage, implicit storage, fractional storage and non-storage. In these examples, the MPT modeler deliberately prefers to measure cognitive processes other than a SDT strength measure. See Erdfelder et al. (2009) and Batchelder and Riefer (1999) for additional examples of MPT models in psychology.

MPT models are mixture models because with this approach it is assumed that there are possibly different knowledge states that have differential consequence for behavior. For example, sometimes there is enough information stored in memory that the individual can reproduce the target event entirely, provided that the information is accessible at the time of test. But for other tests, the requisite information is either incomplete or totally missing. In the Chechile (2004) 6P model there are separate tree pathways for these two different knowledge states. The overall proportion of traces that are sufficiently stored is defined as the storage probability  $\theta_S$ . The  $\theta_S$  parameter is thus a mixture component. Similarly the other parameters in the 6P model are also probabilities and can be regarded as conditional mixture probabilities. Chechile (2013) provided strong evidence for the necessity of considering mixtures for both target-present memory tests as well as for target-absent tests. Evidence was also provided that mixtures are difficult to detect, i.e., data can be generated where a mixture is present but where conventional density plots or quantile–quantile plots fail to detect the mixture. In contrast MPT models are an excellent method for detecting mixtures. Moreover, the absence of a mixture is a special case of a MPT model where the tree paths have probabilities of either 0 or 1<sup>1</sup>.

While there is an ongoing debate about SDT and MPT models in the memory literature, there has not been a corresponding contemporary debate in perceptual psychology about the relative merits of SDT and MPT approaches. Yet the possibility of stochastic mixtures is quite plausible for perceptual detection studies, so there are reasons for considering MPT models for perceptual detection.

One rationale for suspecting that there are mixtures comes from the Stevens (1957, 1961) distinction between prothetic and metathetic continua. Stevens (1961); Stevens (p. 41) illustrated a prothetic dimension with loudness and distinguished it from pitch, which is regarded as 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 Stevens distinction stresses the difference between changes in intensity on a single dimension and changes in qualities. A homogeneous process (as opposed to a mixture) is more likely when dealing with a prothetic continuum; although DeCarlo (2002, 2007) has pointed out that trial-by-trial shifts in attention or phasic alertness can produce a stochastic mixture even in a perceptual detection task on a single dimension. However, if the stimuli are complex and possess qualitative features, then stochastic mixtures are even more likely. Consider, for example, a sonar operator attempting to detect any enemy threats. The operator might detect a clear auditory pattern that is a prototypical signal of a particular class of an enemy submarine. With training and experience the sonar operator can be highly skilled in detecting the complex set of features that are associated with an enemy threat; after all perceptual learning is a well established fact (Kellman, 2002). From this framework, the operator might confidently detect a target, not because of a greater strength or intensity, but because the metathetic pattern exhibited by the stimulus is linked through training to a particular type of target. Yet there might be other cases when a threat is present, but the sonar signal is too poorly defined to be identified as a threat. The operator has to guess in these cases. Hence, from this perspective targets stimuli can be considered a mixture of occasions where the target is confidently and correctly identified and other occasion where the operator guesses. A mixture is also possible over all the target-absent cases. For example, a sonar operator might decide that the stimulus is something other than an enemy threat (e.g., a party boat, or a whale), but for other target-absent events the signal might be too poorly defined for the sonar operator to confidently identify. In this paper, a variation of a MPT model will be advanced for perceptual-detection applications in order to capture the possibility that there are mixtures reflected in the data.

The second focus for this paper concerns the relative accuracy of various statistical procedures for modeling individual differences in terms of the key parameters of a perceptual detection MPT model. There is a widespread belief that the maximum likelihood estimates (MLE) of model parameters, done on an individual basis, is the optional method for obtaining estimates of individual differences. This belief is mistaken; there is now considerable evidence that the MLE can be non-optimal and biased for a number of important practical cases. Even in the case of the Gaussian model with more than two conditions, the MLE estimates are known to be biased and “inadmissible” due to the Stein paradox (Stein, 1956; James and Stein, 1961; Efron and Morris, 1977). These insights have led to empirical Bayes, James–Stein estimators, and other shrinkage estimators as improvements to the MLE (Efron and Morris, 1973; Gruber, 1998). Moreover, based on Monte Carlo simulations of multinomial data,

<sup>1</sup>Some MPT models have been characterized as threshold models by the authors of the model (e.g., the two high-threshold model of Snodgrass and Corwin, 1988). A threshold is an activation level on an underlying strength continuum that triggers the memory to be in a given state. The assumption of thresholds in MPT models has been vigorously challenged by researchers who prefer a SDT perspective (viz. Dube and Rotello, 2012). However, the concept of a mixture over different knowledge states does not require the assumption of a threshold. For example in the Chechile (2004) 6P model, the knowledge states discussed above are not driven by an underlying strength, but rather it is based simply on the existence or not of specific memory content.

Chechile (2009) found that the averaging of individual parameter estimates resulted in greater error than pooling the multinomial data across individuals and fitting the MPT model once. This finding foreshadows a relatively surprising result that is similar to the James–Stein shrinkage estimate for individual model parameter estimates.

**2. THE PERCEPTUAL-DETECTION (PD) MPT MODEL**

**2.1. DATA STRUCTURE AND TREE MODEL**

The Perceptual-Detection (PD) model is essentially the Chechile (2004) 6P model for old/new recognition test trials. The 6P model for storage and retrieval components of memory also has a recall test that is not a part of the perceptual-detection task. The data categories for target-present and target-absent trials as well as the notation for the corresponding population proportions for each response category are shown in **Figure 1**. The PD tree is displayed in **Figure 2**. The MPT model has five parameters; the 6P model had an additional retrieval parameter that is not relevant for perceptual detection. The subscripts for the five parameters have been labeled differently in order to better match the perceptual detection context. The  $\theta_d$  parameter is the proportion of target-present tests when the operator clearly and confidently detects the target stimulus; this parameter corresponds to the sufficient storage parameter  $\theta_S$  in the 6P model. The  $\theta_{nt}$  parameter is the proportion of the target-absent trials when the operator can confidently identify a stimulus that is different than the target; this parameter corresponds to the knowledge-based foil rejection parameter  $\theta_k$  in the 6P model.

The  $\theta_d$  and  $1 - \theta_d$  parameters are mixing rates for target-present trials. When the target is not clearly detected, the observer can still decide that the stimulus is a target (with conditional probability  $\theta_g$ ) by a secondary process that is simply labeled as a guessing process. Similarly on target-absent tests, the operator (with probability  $1 - \theta_{nt}$ ) fails to confidently identify a non-target but can still guess (with probability  $\theta_{g'}$ ) that the stimulus is more likely a non-target than a target. The two guessing parameters in the PD model are the same as the guessing parameters in the 6P model. Finally the  $\theta_h$  parameter is a “nuisance” parameter because it is a conditional probability that is only important as a correction for overly confident guessing. This parameter corresponds to the  $\theta_1$  parameter in the 6P model.

**2.2. PARAMETER ESTIMATION AND A RADIOLOGY EXAMPLE**

A great deal is known about the 6P model, and this information directly transfers to the PD model. For example, Chechile (2004)

Response Categories for Target-Present Trials			
No (high conf.)	No (low conf.)	Yes (low conf.)	Yes (high conf.)
$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$

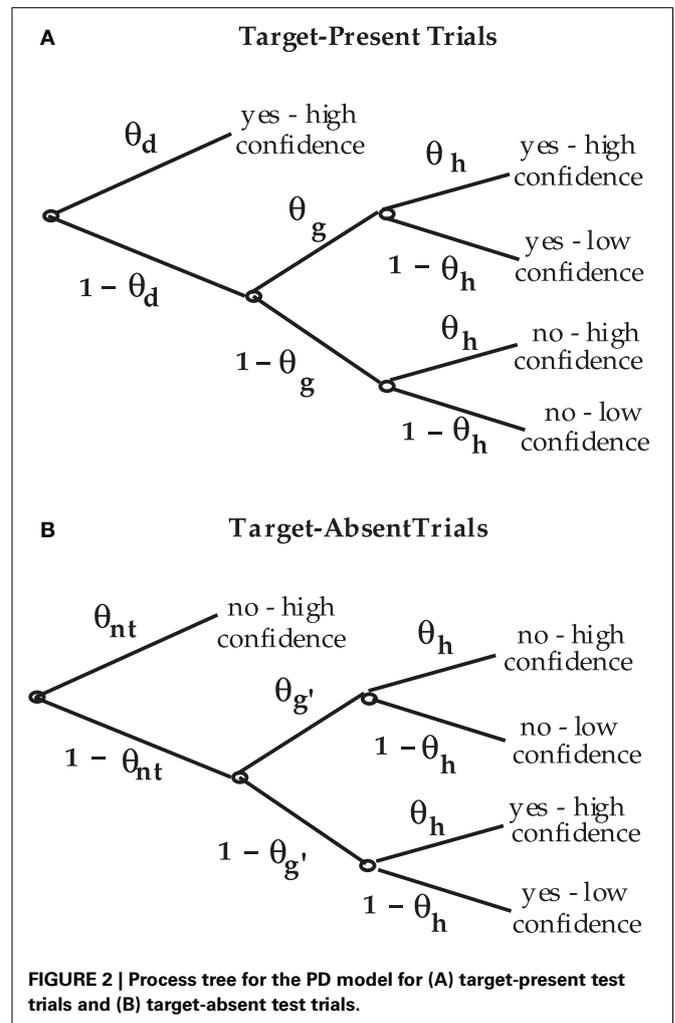
Response Categories for Target-Absent Tests			
No (high conf.)	No (low conf.)	Yes (low conf.)	Yes (high conf.)
$\phi_5$	$\phi_6$	$\phi_7$	$\phi_8$

**FIGURE 1 |** Data categories and population proportions for the PD model.

formally proved that the model is likelihood identifiable, i.e., each configuration of the model parameters results in a unique multinomial likelihood function<sup>2</sup>. Chechile (2004) also showed how the maximum likelihood estimates (MLE) are obtained for the model parameters. In that same paper, an exact Bayesian method for drawing random vectors of values from the posterior distribution was described; the method is called the population parameter mapping (PPM) method (see Chechile, 1998, 2010a). With the PPM method there is a full probability distribution for each model parameter, and there is a probability for the coherence of the model itself. Software also exists for obtaining random vectors from an approximate Bayesian posterior distribution by means of a Markov chain Monte Carlo (MCMC) sampling system<sup>3</sup>. For both the PPM method and the MCMC method, there is a point estimate for each parameter along with

<sup>2</sup>See Chechile (1977, 1998, 2004) for a more detailed discussion of model identifiability.

<sup>3</sup>The MCMC method is an implementation of the Metropolis–Hastings algorithm after an initial “burn in” period of 300,000 cycles for sampling each model parameter.



**FIGURE 2 |** Process tree for the PD model for (A) target-present test trials and (B) target-absent test trials.

a Bayesian posterior probability distribution<sup>4</sup>. The PPM method has several advantages over the MCMC method. First, it does not require a “burn in” period. Second, the posterior distribution is exact as opposed to asymptotically exact. Third, the samples from the posterior distribution are not autocorrelated. Fourth, the PPM method has a probability for the coherence of the model itself.

As an example of parameter estimation for the PD model, let us consider the actual case of the detection characteristics of a single radiologist who was assessing 109 CT scans in order to detect abnormal versus normal scans. Hanley and McNeil (1982) provided the frequencies in four response categories. The categories were labeled as (1) “definitely normal,” (2) “probably normal,” (3) “probably abnormal,” and (4) “definitely abnormal.” There were a total of 58 patients who were later determined to be normal, and 51 patients who were determined later to have an abnormality. The frequencies in these four respective categories for the normals (target-absent) are (33, 9, 14, 2)<sup>5</sup>. The corresponding frequencies for the abnormal (target-present) are (3, 3, 12, 33)<sup>6</sup>. The PPM, MCMC, and MLE point estimates for each parameter in the PD model are displayed in **Table 1**.

The PD model point estimates fit the multinomial frequencies very well as indicated by a non-significant goodness-of-fit difference between the observed and predicted frequencies, i.e.,  $G^2(1) = 0.262$ . In addition to the point estimates, the two Bayesian methods have a posterior probability distribution for each model parameter, and these distributions provide a method for testing some important questions about the radiologist. One of the central ideas in the PD model is the concept that there is

a mixture of states for both target-present cases (abnormals) and for target-absent cases (normals). From the posterior distribution of the  $\theta_d$  parameter, it can be stated that the probability exceeds 0.95 that the  $\theta_d$  parameter is at least 0.39, i.e.,  $P(\theta_d > 0.39) > 0.95$ . Similarly the posterior distribution for the  $\theta_{nt}$  parameter results in the high probability statement that  $\theta_{nt}$  is at least 0.37, i.e.,  $P(\theta_{nt} > 0.37) > 0.95$ .

Using a standard SDT model analysis of the radiological data results in an estimate of  $d' = 2.332$  and a ratio of the standard deviations between the signal and noise conditions of  $\frac{\sigma_S}{\sigma_N} = 1.409$ . This model also fits the data well as indicated by a non-significant difference between the observed and expected frequencies,  $G^2(1) = 0.220$ . However, the SDT model does not posit that there are mixtures, so the finding that the  $\theta_d$  and  $\theta_{nt}$  parameters are reliably different than zero demonstrates that the conventional signal detection model is missing an important feature exhibited by the radiologist. If there were an absence of mixtures, then the PD model would have estimated the  $\theta_d$  and  $\theta_{nt}$  parameters as approximately 0.

For MPT models, the mean of the Bayesian posterior distribution for a parameter is usually a different value than the MLE. Chechile (2004) conducted a series of Monte Carlo simulations to see which of these estimates is more accurate for the 6P model; these simulations directly apply to the PD model. For each Monte Carlo run, a random configuration of the model parameters was selected. These parameter values became the true values that are compared later to the estimated values. Also based on the true values, there is a corresponding set of true multinomial cell proportions, i.e., the  $\phi_i$  values in **Figure 1**. From the multinomial likelihood distributions,  $n$  random “observations” were drawn for the target-present frequencies and another  $n$  random observations were drawn for the target-absent frequencies<sup>7</sup>. Using the cell frequencies, the PPM and MLE parameter estimates are computed. For each estimate there is thus an error score based on the absolute value difference between the estimated value and the true value for that particular Monte Carlo run. For each sample size there was a total of 10,000 Monte Carlo runs. The mean absolute value across the 10,000 runs for PPM and MLE methods are denoted respectively as MAE(ppm) and MAE(mle). The standard deviation of the absolute value errors was also found for both estimation methods. Representative results from these Monte Carlo simulations are shown in **Table 2** for the  $\theta_d$  parameter.

The Bayesian PPM estimates are more accurate for all the sample sizes. Although the MLE and PPM errors are approaching each other, the rate of approach is relatively slow. Notice that even for the case of  $n = 1000$ , there is still a smaller standard deviation of the errors for the PPM estimates. The greater accuracy for

<sup>4</sup>There is a difference in the prior distributions used for the MCMC method and for the PPM method. For the MCMC approach, a flat prior is assumed for each of the PD model parameters, i.e., the  $(\theta_d, \theta_{nt}, \theta_g, \theta_{g'}, \theta_h)$  parameters. However, for the PPM method the prior is a flat distribution for the multinomial cell proportions shown in **Figure 1**, i.e., the  $(\phi_i)$  parameters. The joint posterior distribution for the  $(\phi_i)$  parameters is a product of two Dirichlet distributions. With the PPM method, random samples of  $(\phi_i)$  values are taken from the posterior distribution, and each vector of  $(\phi_i)$  values is mapped to a corresponding vector of the PD model parameters.

<sup>5</sup>There were six cases for the normals where the radiologist used another category called questionable. Three of these cases are assigned here to the second category (probably normal), and three cases were assigned here to the third category (probably abnormal).

<sup>6</sup>There were two CT scans for the abnormal that the radiologist gave the response of questionable. One of these cases was assigned here to the second category, and one was assigned here to the third response category.

**Table 1 | PPM, MCMC, and MLE values for the PD model parameters from 109 CT scans by one radiologist reported in the Hanley and McNeil (1982) study.**

Parameter	PPM	MCMC	MLE
$\theta_d$	0.552	0.555	0.578
$\theta_{nt}$	0.496	0.507	0.523
$\theta_g$	0.734	0.711	0.721
$\theta_{g'}$	0.405	0.438	0.421
$\theta_h$	0.250	0.259	0.227

<sup>7</sup>Given the values for  $p_1 = \phi_1$ ,  $p_2 = \phi_1 + \phi_2$ , and  $p_3 = \phi_1 + \phi_2 + \phi_3$  there are three decision points for randomly assigning a simulated “observation” to one of the four cells. For each simulated observation, a random score is sampled from a uniform distribution on the (0, 1) interval. If the random score is less than  $p_1$ , then the observation is for cell 1. If the random score is in the  $[p_1, p_2)$  interval, then it is an observation for cell 2. If the random score is in the  $[p_2, p_3)$  interval, then the observation is for cell 3. If the random score is greater or equal to  $p_3$ , then it is an observation in cell 4.

**Table 2 | The mean absolute value error (MAE) for the  $\theta_d$  parameter for both the PPM and MLE methods.**

$n$	MAE(ppm)	MAE(mle)	SDE(ppm)	SDE(mle)
10	0.129	0.198	0.090	0.174
20	0.102	0.143	0.076	0.135
30	0.090	0.124	0.070	0.123
40	0.082	0.112	0.064	0.115
50	0.075	0.099	0.061	0.104
100	0.059	0.071	0.050	0.074
300	0.039	0.043	0.035	0.047
600	0.029	0.030	0.027	0.031
1000	0.023	0.023	0.022	0.025

Also shown are the standard deviations of the errors (SDE). Each entry is based on 10,000 Monte Carlo runs from Chechile (2004).

the Bayesian PPM estimates has been also demonstrated for other MPT models (Chechile, 2009, 2010a).

### 2.3. INTERPRETING THE GUESSING PARAMETERS

The  $\theta_g$  and  $\theta_{g'}$  parameters have actually been used in memory applications since the original storage-retrieval separation paper by Chechile and Meyer (1976). In the memory context it was hypothesized that the guessing parameters involve a mixture of processes that include the possibility of partial storage as well as response bias factors. For memory applications, these parameters are both typically greater than  $\frac{1}{2}$ , (viz. Chechile and Ehrensbeck, 1983; Chechile and Meyer, 1976; Chechile, 1987, 2004, 2010b; Chechile and Roder, 1998). If the guessing parameters were strictly response bias, then both parameters should not exceed  $\frac{1}{2}$ , but if there is sometimes partial storage, then that information can be helpful and result in the two guessing parameters exceeding  $\frac{1}{2}$ . Although the possibility of partial storage was likely, it was not possible to estimate fractional storage with only the yes/no recognition data along with confidence ratings. Later Chechile and Soraci (1999) and Chechile et al. (2012) used different test protocols that enabled the measurement of partial storage. These other MPT models did find evidence for partial storage on some test trials; consequently, the finding of both guessing parameters being greater than  $\frac{1}{2}$  is a reasonable outcome.

For the PD model, there is a counterpart to the educated guessing based on partial storage. For the perceptual detection task, there might be occasions where a stimulus is judged more likely a target than not but the quality of the perception is not good enough to constitute a confident classification. On other occasions, the stimulus might be judged more likely a particular “non-target” than a target, but again because the stimulus quality is degraded, the observer is uncertain. For both cases the stimulus is not in a clear detection state, but nonetheless, the person is still able to make informed decisions above a random guessing level.

An interesting special case is when the guessing in both target-present and target-absent conditions are purely response bias, i.e., when  $\theta_g = 1 - \theta_{g'}$ . However, if there is something like the partial storage found for some memory studies, then the stimulus is more likely to yield a yes response in the target-present condition than in the target-absent condition. Note that the radiologist

measured with the PD model exhibited guessing better than pure response bias because  $\theta_g = 0.734 > 1 - \theta_{g'} = 0.595$ . These results are consistent with the interpretation that the radiologist was relatively conservative because the doctor guessed that the patient had an abnormality at a rate of 0.595 for the subset of difficult scans from healthy patients. Nonetheless for the subset of difficult scans from patients with an abnormality, the rate for deciding on the abnormal categorization increased to 0.734. Consequently on these more challenging CT scans the physician did have some differential tendency to use the abnormal classification when in fact the CT scan came from a patient with an abnormality.

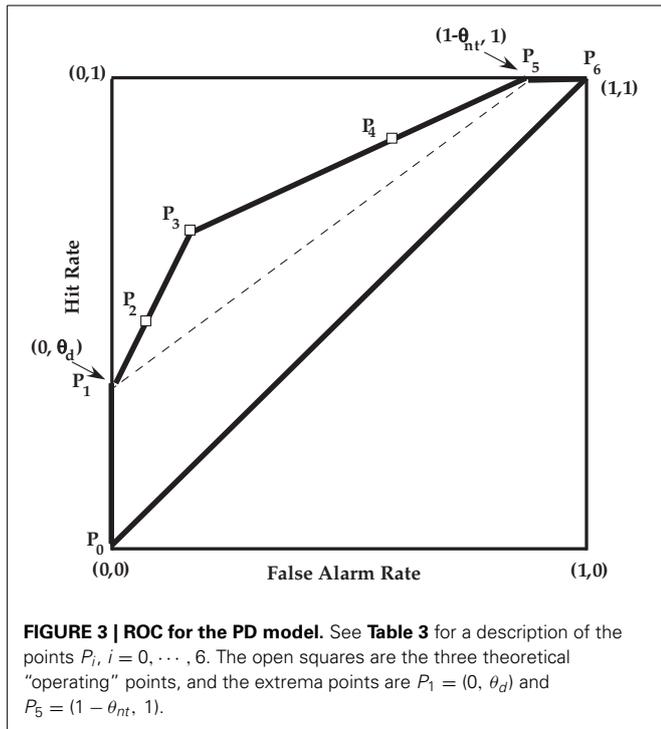
### 2.4. PROPERTIES OF THE ROC FOR THE PD MODEL

The Receiver Operator Characteristic (ROC) in SDT is a curved plot of the hit rate versus the false alarm rate. In standard SDT, any point on the ROC is a possible operating point depending on the decision criterion used by the subject. Hence in standard SDT, the ROC is an iso-sensitivity curve. In standard SDT, the points (0, 0) and (1, 1) are on the ROC curve; these points are the extrema. If the subject had no ability to detect the target, and the data are identical in the target-absent and target-present conditions, then the ROC would be the line of slope 1 connecting the extrema. If there is some greater tendency to detect the target in the target-present condition, then in standard SDT the ROC is a smooth curve in the region of the unit square where  $y \geq x$ .

Empirical ROC plots have been used in numerous experimental papers as a method for comparing theories, but it is challenging to statistically discriminate between models based on only a few points on the empirical ROC. However, given the historical interest in the ROC in psychology, it is instructive to consider the theoretical ROC for the PD model. See **Figure 3** for a general ROC illustration for the PD model. Also see **Table 3** for the PD model equations that are linked to key operating points. The table caption describes the definition of the three discrete points illustrated by the open squares in **Figure 3**, i.e., points  $P_2$ ,  $P_3$ , and  $P_4$ . These three points and the two extreme points for the PD model,  $P_1$  and  $P_5$  are a function of the five parameters in the PD model. If  $0 < \theta_d < 1$ ,  $0 < \theta_{nt} < 1$ , and  $\theta_g > 1 - \theta_{g'}$ , then the ROC path is along two linear segments. Note that the single-high threshold model discussed by Macmillan and Creelman (2005) is the special case of the PD model when  $\theta_{nt} = 0$  and  $\theta_g = 1 - \theta_{g'}$ . The double-high threshold model also discussed in Macmillan and Creelman (2005) is another special case of the PD model when  $\theta_{nt} = \theta_d$  and  $\theta_g = 1 - \theta_{g'}$ .

To better understand the PD ROC, consider points  $P_2$  and  $P_3$ . If we were to define an affirmative response as strictly a “yes” with high confidence, then the corresponding false alarm rate and hit rate would be illustrated by  $P_2$  and have the values corresponding to the prediction equation shown in **Table 3** for that point. Next we redefine an affirmative response as any “yes” response, then the false alarm rate and hit would be illustrated by  $P_3$  and the corresponding prediction equation in **Table 3**. The slope between  $P_2$  and  $P_3$  is denoted as  $s_{23}$  and is given as

$$s_{23} = \frac{(1 - \theta_d) \theta_g}{(1 - \theta_{nt})(1 - \theta_{g'})}, \quad (1)$$



**Table 3 | The PD model equations for the key points shown in Figure 3.**

Point	$x = \text{False alarm}$	$y = \text{Hit}$
$P_0$	0	0
$P_1$	0	$\theta_d$
$P_2$	$(1 - \theta_{nt})(1 - \theta_{g'})\theta_h$	$\theta_d + (1 - \theta_d)\theta_g\theta_h$
$P_3$	$(1 - \theta_{nt})(1 - \theta_{g'})$	$\theta_d + (1 - \theta_d)\theta_g$
$P_4$	$(1 - \theta_{nt})(1 - \theta_{g'}\theta_h)$	$\theta_d + (1 - \theta_d)(1 - (1 - \theta_g)\theta_h)$
$P_5$	$1 - \theta_{nt}$	1
$P_6$	1	1

Point  $P_2$  corresponds to the case where a positive response is considered as a high confident yes, but for point  $P_3$  a positive is regarded as any yes response. For point  $P_4$  a positive is considered as any response that is not a high confident no.

and the slope between points  $P_1$  and  $P_2$  is also equal to  $s_{23}$ . The linear path from points  $P_1$  and  $P_3$  can be described in terms of a hypothetical variable  $v$  that varies on the  $[0, 1]$  interval. The false alarm rate  $x$  and hit rate  $y$  on this path is described by the following equations:

$$x = (1 - \theta_{nt})(1 - \theta_{g'})v, \tag{2}$$

$$y = \theta_d + (1 - \theta_d)\theta_g v. \tag{3}$$

The least risky point  $P_1$  corresponds to when  $v = 0$ . Point  $P_2$  corresponds to the more risky case when  $v = \theta_h$ . Point  $P_3$  corresponds to the even more risky case of  $v = 1$ . Of course the only observable points on this path from  $P_1$  to  $P_3$  are  $P_2$  and  $P_3$ .

Interestingly the slope from  $P_3$  to  $P_4$  is in general different than the slope from  $P_1$  to  $P_3$ . Let us denote the slope from  $P_3$  to  $P_4$  as  $s_{34}$ , and it is given as

$$s_{34} = \frac{(1 - \theta_d)(1 - \theta_g)}{(1 - \theta_{nt})\theta_{g'}}. \tag{4}$$

It is also the case that the slope from  $P_4$  to  $P_5$  is also equal to  $s_{34}$ . Moreover, the linear path from  $P_3$  to  $P_5$  can be described in terms of another hypothetical variable  $w$  that varies from 0 to 1 as the risk increases. The false alarms  $x$  and hits  $y$  on this path is characterized by the following equations:

$$x = (1 - \theta_{nt})(1 - \theta_{g'} + \theta_{g'} w), \tag{5}$$

$$y = \theta_d + (1 - \theta_d)\theta_g + (1 - \theta_d)(1 - \theta_g)w. \tag{6}$$

The  $P_3$  point corresponds to  $w = 0$ ; whereas the  $P_4$  point corresponds to  $w = 1 - \theta_h$  and  $P_5$  corresponds to  $w = 1$ .

Figure 4 illustrates the PD model ROC path from one extreme point to the other in terms of the  $v$  and  $w$  variables. As  $v$  varies from 0 to 1 it traces points on the  $P_1$  to  $P_3$  line as stipulated by Equations (2, 3). Similarly as  $w$  varies from 0 to 1, (Equation 5) and (Equation 6) traces points on the  $P_3$  to  $P_5$  line. Notice that  $\theta_h$  determines the separation from each of the two extreme ends. This feature is a property of the PD model because there is a common parameter of incorrectly using the high confidence rating when guessing regardless if the guessing is done in either the target-present condition or the target-absent condition. Chechile (2004) also presented another identifiable memory MPT model where there are separate parameters for over confidence when using the “yes” response ( $\theta_2$ ) versus over confidence when using the “no” response ( $\theta_1$ ). This model is the 7B model. Other than the difference in the handling of over confidence, the 7B and 6P models are identical, i.e., the 6P model is the special case of 7B where  $\theta_h = \theta_1 = \theta_2$ . Model 7B can also be applied to the perceptual detection task (lets denote that model as the PD\* model). In the PD\* model the  $\theta_2$  parameter determines the location for the  $v$  variable for the  $P_2$  point, and the  $\theta_1$  parameter determines the separation for the  $w$  variable from the maximum of 1. Hence, the spacing for the points on the  $v - w$  plot is different for the PD\* model than the spacing shown in Figure 4 for the PD model.

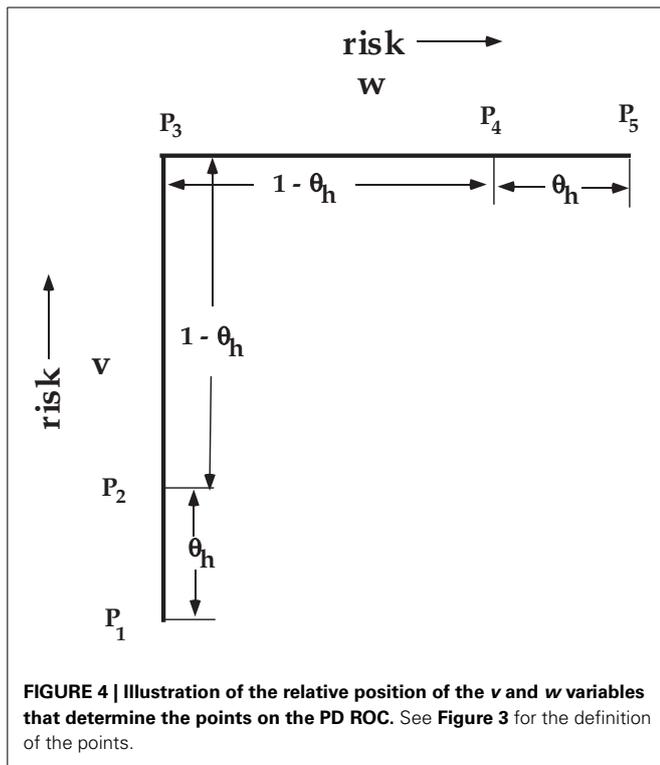
In general the slope from  $P_3$  to  $P_5$  is less than the slope from  $P_1$  to  $P_3$ . Given Equations (1), and (4) the ratio of the slopes can be written as

$$r = \frac{s_{35}}{s_{13}} = \frac{(1 - \theta_g)(1 - \theta_{g'})}{\theta_g\theta_{g'}}. \tag{7}$$

If there is some partial or degraded perception, then the tendency to respond “yes” is at least equal or greater in the target-present condition as it is in the target-absent condition. It follows that

$$\frac{\theta_g}{1 - \theta_g} \geq \frac{1 - \theta_{g'}}{\theta_{g'}}. \tag{8}$$

It also follows from Equations (7, 8) that  $r \leq 1$ . Consequently, if  $\theta_g > 1 - \theta_{g'}$ , then the slope from  $P_1$  to  $P_3$  is larger than the slope



from  $P_3$  to  $P_5$ . The case where  $r = 1$  corresponds to when  $\theta_g = 1 - \theta_{g'}$ , or when there is the same “yes” guessing in the target-present condition as in the target-absent condition. In this special case, there is no partial detection, and the ROC does not have two linear components, but there is instead a single line of slope  $\frac{1-\theta_d}{1-\theta_{nt}}$  between  $P_1$  and  $P_5$ .

The area under the ROC has been used as a measure of sensitivity in standard SDT. It is straightforward to show that area  $A_c$  between the  $P_1$ - $P_5$  dashed line in Figure 3 and the main diagonal line of  $y = x$  is  $\frac{1}{2}(\theta_d + \theta_{nt} - \theta_d\theta_{nt})^8$ . This region is a function of certain perceptual detection and does not depend on guessing. Because the total area in the upper half of the unit square where  $y > x$  is  $\frac{1}{2}$ , it is advantageous to multiply  $A_c$  by 2, so that the area measure of certain detection is placed on a 0 to 1 scale. This measure is defined as a certain detection  $D_c$ , and

$$D_c = \theta_d + \theta_{nt} - \theta_d\theta_{nt}. \tag{9}$$

The area of the  $P_1 P_3 P_5$  triangle is a function of guessing. This area is denoted as  $A_g$ , and it can be found from Heron’s formula, i.e.,  $A_g = \frac{1}{2}(1 - \theta_{nt})(1 - \theta_d)[\theta_g - (1 - \theta_{g'})]$ . We can put this measure of effective guessing on a 0 to 1 scale by defining  $D_g = 2A_g$  or

$$D_g = (1 - \theta_{nt})(1 - \theta_d)[\theta_g - (1 - \theta_{g'})]. \tag{10}$$

<sup>8</sup>Note that the total area above the main diagonal is  $\frac{1}{2}$ , and the area above the dashed line is  $\frac{1}{2}(1 - \theta_d)(1 - \theta_{nt})$ , so  $A_c$  can be determined by subtracting these quantities.

Thus the total detection measure can be defined as twice the area between the ROC and the main diagonal; this metric is  $D = D_c + D_g$  or

$$D = \theta_d + \theta_{nt} - \theta_d\theta_{nt} + (1 - \theta_{nt})(1 - \theta_d)[\theta_g - (1 - \theta_{g'})], \tag{11}$$

As an example, let us compute these area-based metrics for the radiological data discussed in section 2.2. Using PPM estimates for  $\theta_d$  and  $\theta_{nt}$ , it follows from Equation (9) that  $D_c = 0.774$ . The corresponding  $D_g$  measure from Equation (10) is 0.031, so the overall  $D$  metric is 0.805.

Although the detection measure  $D$  is on a proportional basis, it is, nonetheless, a confounded measure because it does not delineate how the detection was achieved. For example suppose that  $\theta_{nt} = 0.805$  and  $\theta_d = 0$ , then the resulting  $D$  value would be the same as for the radiologist discussed above. Clearly the hypothetical observer with  $\theta_d = 0$  and  $\theta_{nt} = 0.805$  would be very good at recognizing a normal CT scan, but would not be capable of detecting an abnormal scan, which would be a rather serious problem for the diseased patients of that hypothetical radiologist! Consequently, the area-based  $D$  metric, along with its component metrics of  $D_c$  and  $D_g$ , is less informative as the original PD model parameters. The detection of the target increases with the value of the  $\theta_d$  parameter, and the identification of a non-target increases with the value of the  $\theta_{nt}$  parameter. Those two types of detection can be quite different. It is also informative to know how the observer does for the unclear cases where there is guessing. The  $D$  metric does not pull out the many different perceptual and decision-making characteristics of the observer’s behavior. Also the standard SDT metrics of  $d'$  and the ratio of the standard deviations do not extract the different properties of the observer’s perceptual-detection performance.

### 3. INDIVIDUAL DIFFERENCE ESTIMATION FOR THE PD MODEL

A fundamental issue that arises in mathematical psychology is the basis for fitting a model. One method is to fit the model separately for each individual and to average individual estimates for the group average. Another method is to aggregate the data across a group of individuals for a particular experimental condition and then fit the model once for that condition<sup>9</sup>. The estimates from these two approaches differ. Although there are applications where each of these pure approaches is reasonable, in this paper a hybrid of these two methods will be recommended. Consequently, the answer to the question as to how to fit a model depends on the purpose of the analysis.

There are several contexts that necessitate the fitting of the model on an individual basis. For example, if the model is a non-linear function of an independent variable, then many investigators have demonstrated that group-averaged data can result in biased fits (Estes, 1956; Sigler, 1987; Ashby et al., 1994). Also

<sup>9</sup>A third approach also exists for obtaining individual and group effects by means of a hierarchical Bayesian model similar to the analysis developed for MPT models by Klauer (2010). This method is computationally challenging, and it has not yet been assessed to see if it has improved accuracy relative to the simple model advanced in the present paper.

the theoretical issue being examined can require that the analysis be done on an individual basis. For example, Chechile (2013) examined the memory hazard function to see if there was evidence of a mixture over stimuli. Had that analysis been done on a grouped-data basis, then any results suggesting a mixture could have been a mixture over individuals with different memory properties instead of a mixture over stimuli.

There are also cases when pooling the data prior to the model fit is the preferred analysis (Cohen et al., 2008; Chechile, 2009). Chechile (2009), for example, studied four prototypic MPT models with an extensive series of Monte Carlo simulations in order to examine the relative accuracy of averaging versus data pooling. For any given Monte Carlo run, a group of  $n_g$  simulated “subjects” with slightly different true values for the model parameters was constructed, and for each artificial subject there were  $n_r$  “observations” that were randomly sampled from the appropriate multinomial likelihood distribution<sup>10</sup>. Based on this set of simulated outcome frequencies, the model was fit in two different ways: (1) the averaging method and (2) the data-pooling method. For the averaging method the MPT model was fit separately for each of the  $n_g$  subjects, and these estimates were averaged to obtain an estimate for each model parameter. For an arbitrary model parameter,  $\theta_x$ , the group average estimate is  $\bar{\theta}_x = \frac{1}{n_g} \sum_{i=1}^{n_g} \hat{\theta}_{xi}$  where  $\hat{\theta}_{xi}$  is the parameter estimate for the  $i$ th subject. For any Monte Carlo run, the absolute value difference was computed between  $\bar{\theta}_x$  and the true mean for that parameter  $\theta_x(true) = \frac{1}{n_g} \sum_{i=1}^{n_g} \theta_{xi}(true)$ . This difference is taken as the error for the averaging method for that one Monte Carlo run. The process was then repeated so that in total there were 1000 separate Monte Carlo runs for each combination of  $n_g$  and  $n_r$ . Across these separate Monte Carlo runs the model parameters were varied, so the model was simulated over a vast set of configurations of the parameters. The overall error for the averaging method is the mean error across the 1000 Monte Carlo data sets for each combination of  $n_g$  and  $n_r$ . For the identical data as described above, a corresponding error was also found for the pooling method. For the pooling method the frequencies in each multinomial response category was summed across the  $n_g$  subjects in a group, and the model was fit once with the pooled data. The estimate based on pooling for the  $j$ th simulated data set is denoted as  $\hat{\theta}_{xj}(pooled)$ . The absolute value difference between this estimate and the true value for that run is the pooling error for the  $j$ th Monte Carlo data set, and mean error across all 1000 data sets is the overall error for the pooling method<sup>11</sup>. For all four models reported in Chechile (2009) and for most combinations of  $n_g$  and  $n_r$ , the mean error for the pooling method was less than the corresponding error obtained for the averaging method<sup>12</sup>. Consequently, Chechile (2009) reported a pooling

advantage score that was the difference between the mean averaging error and the mean pooling error. For example, a positive value for the pooling advantage score of 0.07 means that the averaging mean error was larger by 0.07 than the corresponding pooling error. A negative pooling advantage score would mean that the averaging method had less error than the pooling method.

One of the models examined in Chechile (2009) was a four-cell MPT model that is identical to the structure of the process trees for either the target-present or the target-absent test conditions with the PD model. Consequently, those Monte Carlo simulations directly apply to the PD model. **Table 4** provides a condensed summary of the Monte Carlo results from Chechile (2009). The  $\theta_d$  parameter in **Table 4** corresponds to the  $\theta_s$  parameter in Model A; whereas  $\theta_g$  and  $\theta_h$ , respectively, correspond to the  $\theta_g$  and  $\theta_l$  parameters in Model A.

The pooling advantage scores in **Table 4** exhibit a number of interesting properties that were also found with the other MPT models. First, the pooling advantage scores are positive indicating that there is greater accuracy for the pooling method. Second, although the magnitude of the pooling advantage decreases with the number of observations per subject ( $n_r$ ), there is still a non-trivial advantage for pooling even when  $n_r = 400$ . It is challenging to do an experiment with large values for  $n_r$ . For example, a replication number of 50 is larger than all but two of the memory studies reported from my laboratory. Consequently, the idea of running a large number of replication trials per subject is not a practical option. Third, the size of the pooling advantage increases with group size  $n_g$ . This effect is due to the fact that the error for the pooling method decreases rapidly with increasing group size; whereas the error for the averaging method slowly decreases with increasing  $n_g$ ,

**Table 4 | The difference in mean error between averaging and pooling for  $n_g$  individuals in a group and for  $n_r$  trials in the target-present condition.**

$n_g$ $n_r$	Pooling Advan. score		
	$\theta_d$	$\theta_g$	$\theta_h$
20 20	0.069	0.076	0.078
20 50	0.045	0.050	0.051
20 100	0.034	0.035	0.034
20 400	0.015	0.013	0.013
40 20	0.078	0.086	0.087
40 50	0.054	0.059	0.059
40 100	0.037	0.040	0.040
40 400	0.017	0.015	0.015
80 20	0.087	0.096	0.098
80 50	0.059	0.065	0.064
80 100	0.043	0.043	0.045
80 400	0.020	0.016	0.016

*This difference is a pooling advantage score. Positive values indicate less error for the pooling method. Monte Carlo simulations from Chechile (2009).*

<sup>10</sup>Each individual was within  $\pm 0.03$  of the group mean.

<sup>11</sup>This whole procedure of estimating the model with both the averaging and pooling method was done for both PPM and MLE estimates for each of the four typical MPT models.

<sup>12</sup>Only eight cases out of 640 cases reported in Chechile (2009) had greater error for the pooling method, and all of these exceptions were when the MLE was used. Generally the MLE was not the optimal estimator for the model parameters because the corresponding Bayesian PPM estimator had greater accuracy.

so the net effect is that the pooling advantage score increases with  $n_g$ .

It might not seem intuitive as to why the pooling of data results in superior estimates for the group mean. This result is more reasonable when viewed from a Bayesian perspective. From Bayes theorem it does not matter if the data are examined in aggregate or one observation at a time, provided that the same starting prior probability is used. Suppose we use a uniform distribution as the prior distribution for each combination of the parameters  $(\theta_d, \theta_g, \theta_h)$ . Let us call this prior the “vague” prior. Furthermore suppose we examine the model parameters for the first individual in the group via Bayes theorem to yield a posterior distribution. The posterior distribution after the first individual should then be the prior distribution for examining the data for the second subject, i.e., it is no longer appropriate to maintain the vague prior after examining the first subject. Similarly the prior distribution for Subject 3 should be the posterior distribution after considering the first two subjects. This one-subject-at-a-time method eventually yields a posterior distribution that is the same as the posterior distribution achieved by pooling the multinomial categories and applying Bayes theorem once. Had the Bayesian analyst used a vague prior for each of the  $n_g$  subjects and averaged the estimates, then the analysis would not be consistent in the application of Bayes theorem. The averaging of separate estimates is not an operation by which probability distributions are revised via Bayes theorem. In terms of this framework, the findings in **Table 4** are quite reasonable. The pooling method should be more accurate, and the pooling advantage should grow with the size of the group.

Despite the above demonstration of a pooling advantage for estimating the group mean, it is still an open question as to what should be the basis for estimating the model parameters for an individual. Two choices seem reasonable. One method is simply to use the data for just the individual, e.g., for the  $\theta_d$  parameter it would be  $\hat{\theta}_{di}$  for the  $i$ th observer. For the second method the data for the individual is used but there is a fixed correction so that the mean across all observers is equal to the pooled estimate for the group. For the  $\theta_d$  parameter this estimate is denoted as  $\hat{\theta}_{di}^{(a)}$  and is defined as

$$\hat{\theta}_{di}^{(a)} = \hat{\theta}_d(\text{pooled}) - \bar{\theta}_d + \hat{\theta}_{di}. \tag{12}$$

Note that the two methods have estimates that are perfectly correlated because the adjusted estimate  $\hat{\theta}_{di}^{(a)}$  is a constant plus the individual estimate  $\hat{\theta}_{di}$ . The constant correction term is equal to  $\hat{\theta}_d(\text{pooled}) - \bar{\theta}_d$ . The correction makes the mean of the adjusted estimates equal to the pooling method estimate because

$$\frac{1}{n_g} \sum_{i=1}^{n_g} \hat{\theta}_{di}^{(a)} = \hat{\theta}_d(\text{pooled}) - \bar{\theta}_d + \bar{\theta}_d = \hat{\theta}_d(\text{pooled}).$$

The estimate based on Equation (12) is similar in principle to a James–Stein estimator used for the linear model for Gaussian random variables because the estimate for the individual is shifted based on properties of the group.

Another Monte Carlo simulation was designed for a widely different group of simulated observers in order to assess the relative accuracy of the two methods for estimating the parameters for individuals. The group consisted of 10 observers for each of the  $3 \times 3$  combinations of values for  $\theta_d$  and  $\theta_{nt}$ . The three values were 0.2, 0.5, and 0.8. For each of the 90 simulated observers the values for  $\theta_h$  were randomly selected from a beta distribution with coefficients of 2 and 4, and the  $\theta_g$  and  $\theta_{g'}$  parameters were randomly selected from a beta distribution with coefficients of 28 and 14. Consequently true scores were established for each simulated observer. For each observer, 20 simulated observations were randomly sampled for the target-present condition, and another 20 observations were randomly sampled for the target-absent condition. These observations were based on the appropriate multinomial likelihood distribution for each subject. The PD model was then estimated by each method described above. Because  $\theta_d$  and  $\theta_{nt}$  are the two key parameters of interest in the PD model, the root mean square (rms) error was found between the true score point  $\{\theta_{di}(\text{true}), \theta_{nti}(\text{true})\}$  and the estimated point for the individual  $\{\hat{\theta}_{di}, \hat{\theta}_{nti}\}$ . The rms error for the adjusted score point  $\{\hat{\theta}_{di}^{(a)}, \hat{\theta}_{nti}^{(a)}\}$  was also found. The rms errors for the individual and the adjusted method are respectively 0.1671 and 0.1385. Thus, the adjusted estimates based on Equation (12) resulted in a 17% reduction in the rms error. This simulation illustrates the improvement in the accuracy of model estimation by the use of the adjusted score method.

#### 4. DISCUSSION

In this paper the Chechile (2004) 6P memory measurement model was modified and applied to perceptual detection. The resulting PD model is a MPT model that has two mixture rate parameters ( $\theta_d$  and  $\theta_{nt}$ ) that measure the proportion of times that the observer confidently detects something that belongs to an identifiable category. The categories are different for targets and non-targets, but in both cases something is being identified. The measurement of these detection rates is an important part of the psychometric assessment of perceptual performance. The PD model also has three other parameters that come into play when the observer is unable to confidently classify the stimulus.

The PD model differs from standard SDT on the issue of stochastic mixtures. MPT models, like the PD model, are essentially probability mixture models. In contrast, SDT developed in the context of assuming separate but homogeneous distributions for target-present and target-absent conditions. The success of the PD model in accounting for the radiological judgments described earlier in this paper occurred because the PD model was sensitive to the fact the radiologist was able to know sometimes that a CT scan was normal and to know at other times that a CT scan revealed an identifiable abnormality. This attribute of categorical and sophisticated perception is not an isolated property of experts. More than 120 years ago William James discussed the importance of perceptual learning; in fact perception according to James differed from a pure sensation because of the information that the person associates and adds to the sensation (James, 1890). There is now a vast literature describing the improvement in

perception with practice (Kellman, 2002). With experience people can develop refined perceptual categories that sharpen their ability to process and to interpret stimuli.

It is noteworthy that the prototypic experiments in the early history of SDT used stimuli that were designed to be featureless and varied on only a single prothetic intensity dimension. For example the stimulus-absent stimulus for some experiments was white noise; whereas the target-present stimulus was a louder white noise (Tanner et al., 1956). Perceptual categories and perceptual learning is limited for such impoverished stimuli. SDT is expected to be quite successful for such applications, but SDT is expected to be problematic when stimuli possess rich perceptual features and when the observer has some experience with the class of stimuli. For those applications, the PD model would be a more suitable cognitive psychometric tool for assessing the properties of the observer.

The PD model is a minimalistic model that intentionally eschews delineating any specific cognitive representation of the stimulus. Like other MPT models, there are probability measures for specific states. The states for the PD model are: (1) a state of certain target recognition, which occurs on  $\theta_d$  proportion of the target-present trials, and (2) the state of certain identification of something other than a target, which occurs on  $\theta_{nt}$  proportion of the target-absent trials. These probability measures provide for a characterization of the observer's detection ability.

MPT models have many desirable statistical properties and can be estimated by a variety of methods. Monte Carlo simulations with large sample sizes demonstrated that the MLE and the Bayesian posterior mean for the PD model were very close, but the accuracy of these estimates differed more substantially for smaller sample sizes. When the estimates differ, the Bayesian mean was found to be more accurate. In addition, an improved estimate was found for the individual observer when the estimate based on the individual's data was adjusted. The adjustment was a fixed amount for all observers, and it equated the mean of the adjusted scores to the mean of the estimate based on pooled data. This adjustment was discussed as an analogous adjustment to the James–Stein shrinkage improvements to the MLE found for the multiple-group Gaussian model.

## REFERENCES

- Ashby, F. G., Maddox, W. T., and Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychol. Sci.* 5, 144–151. doi: 10.1111/j.1467-9280.1994.tb00651.x
- Ashby, F. G., and Townsend, J. T. (1986). Varieties of perceptual independence. *Psychol. Rev.* 93, 154–179. doi: 10.1037/0033-295X.93.2.154
- Batchelder, W. H., and Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.* 6, 57–86. doi: 10.3758/BF03210812
- Bröder, A., and Schütz, J. (2009). Recognition ROCs are curvilinear - or are they? On premature arguments against the two-high-threshold model of recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 587–606. doi: 10.1037/a0015279
- Chechile, R. A. (1977). Likelihood and posterior identification: implications for mathematical psychology. *Br. J. Math. Stat. Psychol.* 30, 177–184. doi: 10.1111/j.2044-8317.1977.tb00737.x
- Chechile, R. A. (1978). Is  $d'_c$  a suitable measure of recognition memory "strength"? *Bull. Psychon. Soc.* 12, 152–154. doi: 10.3758/BF03329655
- Chechile, R. A. (1987). Trace susceptibility theory. *J. Exp. Psychol. Gen.* 116, 203–222. doi: 10.1037/0096-3445.116.3.203
- Chechile, R. A. (1998). A new method for estimating model parameters for multinomial data. *J. Math. Psychol.* 42, 432–471. doi: 10.1006/jmps.1998.1210
- Chechile, R. A. (2004). New multinomial models for the Chechile-Meyer task. *J. Math. Psychol.* 48, 364–384. doi: 10.1016/j.jmp.2004.09.002
- Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *J. Math. Psychol.* 53, 562–576. doi: 10.1016/j.jmp.2009.06.005
- Chechile, R. A. (2010a). A novel Bayesian parameter mapping method for estimating the parameters of an underlying scientific model. *Commun. Stat. Theor. Methods* 39, 1190–1201. doi: 10.1080/03610920902859615
- Chechile, R. A. (2010b). Modeling storage and retrieval processes with clinical populations with applications examining alcohol-induced amnesia and Korsakoff amnesia. *J. Math. Psychol.* 54, 150–166. doi: 10.1016/j.jmp.2009.03.006
- Chechile, R. A. (2013). A novel method for assessing rival models of recognition memory. *J. Math. Psychol.* 57, 196–214. doi: 10.1016/j.jmp.2013.07.002
- Chechile, R. A., and Ehrensbeck, K. (1983). Long-term storage losses: a dilemma for multistore models. *J. Gen. Psychol.* 109, 15–30. doi: 10.1080/00221309.1983.9711505
- Chechile, R. A., and Meyer, D. L. (1976). A Bayesian procedure for separately estimating storage and retrieval components of forgetting. *J. Math. Psychol.* 13, 269–295. doi: 10.1016/0022-2496(76)90022-5
- Chechile, R. A., and Roder, B. (1998). "Model-based measurement of group differences: an application directed toward understanding the information-processing mechanism of developmental dyslexia," in *Perspectives on Fundamental Processes in Intellectual Functioning: Volume 1 - A Survey of Research Approaches*, eds S. Soraci and W. J. McIlvane (Stamford, CN: Ablex), 91–112.
- Chechile, R. A., Sloboda, L. N., and Chamberland, J. R. (2012). Obtaining separate measures for implicit and explicit memory. *J. Math. Psychol.* 56, 35–53. doi: 10.1016/j.jmp.2012.01.002
- Chechile, R. A., and Soraci, S. A. (1999). Evidence for a multiple-process account of the generation effect. *Memory* 7, 483–508. doi: 10.1080/741944921
- Cohen, A. L., Sanborn, A. N., and Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychon. Bull. Rev.* 15, 692–712. doi: 10.3758/PBR.15.4.692
- DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: theoretical developments with applications to recognition memory. *Psychol. Rev.* 109, 710–721. doi: 10.1037/0033-295X.109.4.710
- DeCarlo, L. T. (2007). The mirror effect and mixture signal detection theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 18–33. doi: 10.1037/0278-7393.33.1.18
- Dube, C., and Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Exp. Psychol. Learn. Mem. Cogn.* 38, 130–151. doi: 10.1037/a0024957
- Efron, B., and Morris, C. (1973). Stein's estimation rule and its competitors. *J. Am. Stat. Assoc.* 65, 117–130. doi: 10.1080/01621459.1973.10481350
- Efron, B., and Morris, C. (1977). Stein's paradox in statistics. *Sci. Am.* 236, 119–127. doi: 10.1038/scientificamerican0577-119
- Egan, J. P. (1958). *Recognition Memory and the Operating Characteristic*. Technical Note AFCRC-TN-58-51. Hearing and Communication Laboratory. Indiana University.
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., and Nadarevic, L. (2009). Multinomial processing tree models: a review of the literature. *J. Psychol.* 217, 108–124. doi: 10.1027/0044-3409.217.3.108
- Estes, W. K. (1956). The problem of inference from curves based on grouped data. *Psychol. Bull.* 53, 134–140. doi: 10.1037/h0045156
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.
- Gruber, M. H. J. (1998). *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators*. New York, NY: Marcel Dekker Inc.
- Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. *Radiology* 143, 29–36.
- James, W. (1890). *The Principles of Psychology*, Vol. 2. New York, NY: Holt and Company. doi: 10.1037/11059-000
- James, W., and Stein, C. (1961). "Estimation with quadratic loss," in *Proceeding of the Fourth Berkeley Symposium on Mathematics and Statistics* (Berkeley, CA), 361–379.
- Kellen, D., Klauer, K. C., and Bröder, A. (2013). Recognition memory models and binary-response ROCs: a comparison by minimum description length. *Psychon. Bull. Rev.* 20, 693–719. doi: 10.3758/s13423-013-0407-2

- Kellman, P. J. (2002). "Perceptual learning," in *Steven's Handbook of Experimental Psychology*. Learning, motivation, and emotion, Vol. 3, eds R. Gallistel and H. Pashler (New York: Wiley), 259–299.
- Klauer, K. C. (2010). Hierarchical multinomial processing tree models: a latent-trait approach. *Psychometrika* 75, 70–75. doi: 10.1007/s11336-009-9141-0
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide*, 2nd Edn. Mahwah, NJ: Erlbaum.
- Malmberg, K. J. (2008). Recognition memory: a review of the critical findings and an integrated theory for relating them. *Cogn. Psychol.* 57, 335–384. doi: 10.1016/j.cogpsych.2008.02.004
- Sigler, R. S. (1987). The perils of averaging data over strategies: an example from children's addition. *J. Exp. Psychol. Gen.* 121, 278–304.
- Snodgrass, J. G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50. doi: 10.1037/0096-3445.117.1.34
- Stein, C. (1956). "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability* (Berkeley, CA), 197–206.
- Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.* 64, 163–181. doi: 10.1037/h0046162
- Stevens, S. S. (1961). Toward a resolution of the Fechner-Thurstone legacy. *Psychometrika* 26, 35–47. doi: 10.1007/BF02289683
- Tanner, W. P., and Swets, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.* 61, 401–409. doi: 10.1037/h0058700
- Tanner, W. P., Swets, J. A., and Green, D. M. (1956). *Some General Properties of the Hearing Mechanism*. (Electronic Defense Group, Technical Report No. 30). Ann Arbor, MI: University of Michigan.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517. doi: 10.1006/jmla.2002.2864

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 April 2014; accepted: 05 June 2014; published online: 27 June 2014.

Citation: Chechile RA (2014) Using a multinomial tree model for detecting mixtures in perceptual detection. *Front. Psychol.* 5:641. doi: 10.3389/fpsyg.2014.00641

This article was submitted to *Quantitative Psychology and Measurement*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Chechile. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.