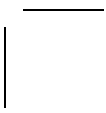
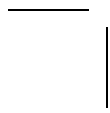
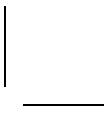
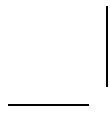


PART I

Elementary Cognitive
Mechanisms



Multidimensional Signal Detection Theory

F. Gregory Ashby and Fabian A. Soto

Abstract

Multidimensional signal detection theory is a multivariate extension of signal detection theory that makes two fundamental assumptions, namely that every mental state is noisy and that every action requires a decision. The most widely studied version is known as general recognition theory (GRT). General recognition theory assumes that the percept on each trial can be modeled as a random sample from a multivariate probability distribution defined over the perceptual space. Decision bounds divide this space into regions that are each associated with a response alternative. General recognition theory rigorously defines and tests a number of important perceptual and cognitive conditions, including perceptual and decisional separability and perceptual independence. General recognition theory has been used to analyze data from identification experiments in two ways: (1) fitting and comparing models that make different assumptions about perceptual and decisional processing, and (2) testing assumptions by computing summary statistics and checking whether these satisfy certain conditions. Much has been learned recently about the neural networks that mediate the perceptual and decisional processing modeled by GRT, and this knowledge can be used to improve the design of experiments where a GRT analysis is anticipated.

Key Words: signal detection theory, general recognition theory, perceptual separability, perceptual independence, identification, categorization

Introduction

Signal detection theory revolutionized psychophysics in two different ways. First, it introduced the idea that trial-by-trial variability in sensation can significantly affect a subject's performance. And second, it introduced the field to the then-radical idea that every psychophysical response requires a decision from the subject, even when the task is as simple as detecting a signal in the presence of noise. Of course, signal detection theory proved to be wildly successful and both of these assumptions are now routinely accepted without question in virtually all areas of psychology.

The mathematical basis of signal detection theory is rooted in statistical decision theory, which

itself has a history that dates back at least several centuries. The insight of signal detection theorists was that this model of statistical decisions was also a good model of sensory decisions. The first signal detection theory publication appeared in 1954 (Peterson, Birdsall, & Fox, 1954), but the theory did not really become widely known in psychology until the seminal article of Swets, Tanner, and Birdsall appeared in *Psychological Review* in 1961. From then until 1986, almost all applications of signal detection theory assumed only one sensory dimension (Tanner, 1956, is the principal exception). In almost all cases, this dimension was meant to represent sensory magnitude. For a detailed description of this standard univariate

theory, see the excellent texts of either Macmillan and Creelman (2005) or Wickens (2002). This chapter describes multivariate generalizations of signal detection theory.

Multidimensional signal detection theory is a multivariate extension of signal detection to cases in which there is more than one perceptual dimension. It has all the advantages of univariate signal detection theory (i.e., it separates perceptual and decision processes) but it also offers the best existing method for examining interactions among perceptual dimensions (or components). The most widely studied version of multidimensional signal detection theory is known as general recognition theory (GRT; Ashby & Townsend, 1986). Since its inception, more than 350 articles have applied GRT to a wide variety of phenomena, including categorization (e.g., Ashby & Gott, 1988; Maddox & Ashby, 1993), similarity judgment (Ashby & Perrin, 1988), face perception (Blaha, Silbert, & Townsend, 2011; Thomas, 2001; Wenger & Ingvalson, 2002), recognition and source memory (Banks, 2000; Rotello, Macmillan, & Reeder, 2004), source monitoring (DeCarlo, 2003), attention (Maddox, Ashby, & Waldron, 2002), object recognition (Cohen, 1997; Demeyer, Zaenen, & Wagemans, 2007), perception/action interactions (Amazeen & DaSilva, 2005), auditory and speech perception (Silbert, 2012; Silbert, Townsend, & Lentz, 2009), haptic perception (Giordano et al., 2012; Louw, Kappers, & Koenderink, 2002), and the perception of sexual interest (Farris, Viken, & Treat, 2010).

Extending signal detection theory to multiple dimensions might seem like a straightforward mathematical exercise, but, in fact, several new conceptual problems must be solved. First, with more than one dimension, it becomes necessary to model interactions (or the lack thereof) among those dimensions. During the 1960s and 1970s, a great many terms were coined that attempted to describe perceptual interactions among separate stimulus components. None of these, however, were rigorously defined or had any underlying theoretical foundation. Included in this list were perceptual independence, separability, integrality, performance parity, and sampling independence. Thus, to be useful as a model of perception, any multivariate extension of signal detection theory needed to provide theoretical interpretations of these terms and show rigorously how they were related to one another.

Second, the problem of how to model decision processes when the perceptual space is multidimensional is far more difficult than when there is only one sensory dimension. A standard signal-detection-theory lecture is to show that almost any decision strategy is mathematically equivalent to setting a criterion on the single sensory dimension, then giving one response if the sensory value falls on one side of this criterion, and the other response if the sensory value falls on the other side. For example, in the normal, equal-variance model, this is true regardless of whether subjects base their decision on sensory magnitude or on likelihood ratio. A straightforward generalization of this model to two perceptual dimensions divides the perceptual plane into two response regions. One response is given if the percept falls in the first region and the other response is given if the percept falls in the second region. The obvious problem is that, unlike a line, there are an infinite number of ways to divide a plane into two regions. How do we know which of these has the most empirical validity?

The solution to the first of these two problems—that is, the sensory problem—was proposed by Ashby and Townsend (1986) in the article that first developed GRT. The GRT model of sensory interactions has been embellished during the past 25 years, but the core concepts introduced by Ashby and Townsend (1986) remain unchanged (i.e., perceptual independence, perceptual separability). In contrast, the decision problem has been much more difficult. Ashby and Townsend (1986) proposed some candidate decision processes, but at that time they were largely without empirical support. In the ensuing 25 years, however, hundreds of studies have attacked this problem, and today much is known about human decision processes in perceptual and cognitive tasks that use multidimensional perceptual stimuli.

Box 1 Notation

$A_i B_j$ = stimulus constructed by setting component A to level i and component B to level j

$a_i b_j$ = response in an identification experiment signaling that component A is at level i and component B is at level j

X_1 = perceived value of component A

X_2 = perceived value of component B

Box 1 Continued

$f_{ij}(x_1, x_2)$ = joint likelihood that the perceived value of component A is x_1 and the perceived value of component B is x_2 on a trial when the presented stimulus is $A_i B_j$

$g_{ij}(x_1)$ = marginal pdf of component A on trials when stimulus $A_i B_j$ is presented

r_{ij} = frequency with which the subject responded R_j on trials when stimulus S_i was presented

$P(R_j|S_i)$ = probability that response R_j is given on a trial when stimulus S_i is presented

General Recognition Theory

General recognition theory (see the Glossary for key concepts related to GRT) can be applied to virtually any task. The most common applications, however, are to tasks in which the stimuli vary on two stimulus components or dimensions. As an example, consider an experiment in which participants are asked to categorize or identify faces that vary across trials on gender and age. Suppose there are four stimuli (i.e., faces) that are created by factorially combining two levels of each dimension. In this case we could denote the two levels of the gender dimension by A_1 (male) and A_2 (female) and the two levels of the age dimension by B_1 (teen) and B_2 (adult). Then the four faces are denoted as $A_1 B_1$ (male teen), $A_1 B_2$ (male adult), $A_2 B_1$ (female teen), and $A_2 B_2$ (female adult).

As with signal detection theory, a fundamental assumption of GRT is that all perceptual systems are inherently noisy. There is noise both in the stimulus (e.g., photon noise) and in the neural systems that determine its sensory representation (Ashby & Lee, 1993). Even so, the perceived value on each sensory dimension will tend to increase as the level of the relevant stimulus component increases. In other words, the distribution of percepts will change when the stimulus changes. So, for example, each time the $A_1 B_1$ face is presented, its perceived age and maleness will tend to be slightly different.

General recognition theory models the sensory or perceptual effects of a stimulus $A_i B_j$ via the joint probability density function (pdf) $f_{ij}(x_1, x_2)$ (see Box 1 for a description of the notation used in this article). On any particular trial when stimulus $A_i B_j$ is presented, GRT assumes that the subject's percept can be modeled as a random sample from this

joint pdf. Any such sample defines an ordered pair (x_1, x_2) , the entries of which fix the perceived value of the stimulus on the two sensory dimensions. General recognition theory assumes that the subject uses these values to select a response.

In GRT, the relationship of the joint pdf to the marginal pdfs plays a critical role in determining whether the stimulus dimensions are perceptually integral or separable. The marginal pdf $g_{ij}(x_1)$ simply describes the likelihoods of all possible sensory values of X_1 . Note that the marginal pdfs are identical to the one-dimensional pdfs of classical signal detection theory.

Component A is *perceptually separable* from component B if the subject's perception of A does not change when the level of B is varied. For example, age is perceptually separable from gender if the perceived age of the adult in our face experiment is the same for the male adult as for the female adult, and if a similar invariance holds for the perceived age of the teen. More formally, in an experiment with the four stimuli, $A_1 B_1$, $A_1 B_2$, $A_2 B_1$, and $A_2 B_2$, component A is perceptually separable from B if and only if

$$g_{11}(x_1) = g_{12}(x_1) \quad \text{and} \quad g_{21}(x_1) = g_{22}(x_1) \\ \text{for all values of } x_1. \quad (1)$$

Similarly, component B is perceptually separable from A if and only if

$$g_{11}(x_2) = g_{21}(x_2) \quad \text{and} \quad g_{12}(x_2) = g_{22}(x_2), \quad (2)$$

for all values of x_2 . If perceptual separability fails then A and B are said to be *perceptually integral*. Note that this definition is purely perceptual since it places no constraints on any decision processes.

Another purely perceptual phenomenon is *perceptual independence*. According to GRT, components A and B are perceived independently in stimulus $A_i B_j$ if and only if the perceptual value of component A is statistically independent of the perceptual value of component B on $A_i B_j$ trials. More specifically, A and B are perceived independently in stimulus $A_i B_j$ if and only if

$$f_{ij}(x_1, x_2) = g_{ij}(x_1) g_{ij}(x_2) \quad (3)$$

for all values of x_1 and x_2 . If perceptual independence is violated, then components A and B are perceived dependently. Note that perceptual independence is a property of a single stimulus, whereas perceptual separability is a property of groups of stimuli.

A third important construct from GRT is *decisional separability*. In our hypothetical experiment

with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 , and two perceptual dimensions X_1 and X_2 , decisional separability holds on dimension X_1 (for example), if the subject's decision about whether stimulus component A is at level 1 or 2 depends only on the perceived value on dimension X_1 . A decision bound is a line or curve that separates regions of the perceptual space that elicit different responses. The only types of decision bounds that satisfy decisional separability are vertical and horizontal lines.

The Multivariate Normal Model

So far we have made no assumptions about the form of the joint or marginal pdfs. Our only assumption has been that there exists some probability distribution associated with each stimulus and that these distributions are all embedded in some Euclidean space (e.g., with orthogonal dimensions). There have been some efforts to extend GRT to more general geometric spaces (i.e., Riemannian manifolds; Townsend, Aisbett, Assadi, & Busemeyer, 2006; Townsend & Spencer-Smith, 2004), but much more common is to add more restrictions to the original version of GRT, not fewer. For example, some applications of GRT have been distribution free (e.g., Ashby & Maddox, 1994; Ashby & Townsend, 1986), but most have assumed that the percepts are multivariate normally distributed. The multivariate normal distribution includes two assumptions. First, the marginal distributions are all normal. Second, the only possible dependencies are pairwise linear relationships. Thus, in multivariate normal distributions, uncorrelated random variables are statistically independent.

A hypothetical example of a GRT model that assumes multivariate normal distributions is shown in Figure 2.1. The ellipses shown there are contours of equal likelihood; that is, all points on the same ellipse are equally likely to be sampled from the underlying distribution. The contours of equal likelihood also describe the shape a scatterplot of points would take if they were random samples from the underlying distribution. Geometrically, the contours are created by taking a slice through the distribution parallel to the perceptual plane and looking down at the result from above. Contours of equal likelihood in multivariate normal distributions are always circles or ellipses. Bivariate normal distributions, like those depicted in Figure 2.1 are each characterized by five parameters: a mean on each dimension, a variance on each dimension, and a covariance or correlation between the values on

the two dimensions. These are typically catalogued in a mean vector and a variance-covariance matrix. For example, consider a bivariate normal distribution with joint density function $f(x_1, x_2)$. Then the mean vector would equal

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad (4)$$

and the variance-covariance matrix would equal

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \text{cov}_{12} \\ \text{cov}_{21} & \sigma_2^2 \end{bmatrix} \quad (5)$$

where cov_{12} is the covariance between the values on the two dimensions (i.e., note that the correlation coefficient is the standardized covariance—that is, the correlation $\rho_{12} = \frac{\text{cov}_{12}}{\sigma_1\sigma_2}$).

The multivariate normal distribution has another important property. Consider an identification task with only two stimuli and suppose the perceptual effects associated with the presentation of each stimulus can be modeled as a multivariate normal distribution. Then it is straightforward to show that the decision boundary that maximizes accuracy is always linear or quadratic (e.g., Ashby, 1992). The optimal boundary is linear if the two perceptual distributions have equal variance-covariance matrices (and so the contours of equal likelihood have the same shape and are just translations of each other) and the optimal boundary is quadratic if the two variance-covariance matrices are unequal. Thus, in the Gaussian version of GRT, the only decision bounds that are typically considered are either linear or quadratic.

In Figure 2.1, note that perceptual independence holds for all stimuli except A_2B_2 . This can be seen in the contours of equal likelihood. Note that the major and minor axes of the ellipses that define the contours of equal likelihood for stimuli A_1B_1 , A_1B_2 , and A_2B_1 are all parallel to the two perceptual dimensions. Thus, a scatterplot of samples from each of these distributions would be characterized by zero correlation and, therefore, statistical independence (i.e., in the special Gaussian case). However, the major and minor axes of the A_2B_2 distribution are tilted, reflecting a positive correlation and hence a violation of perceptual independence.

Next, note in Figure 2.1 that stimulus component A is perceptually separable from stimulus component B, but B is not perceptually separable from A. To see this, note that the marginal distributions for stimulus component A are the same, regardless of the level of component B [i.e., $g_{11}(x_1) =$

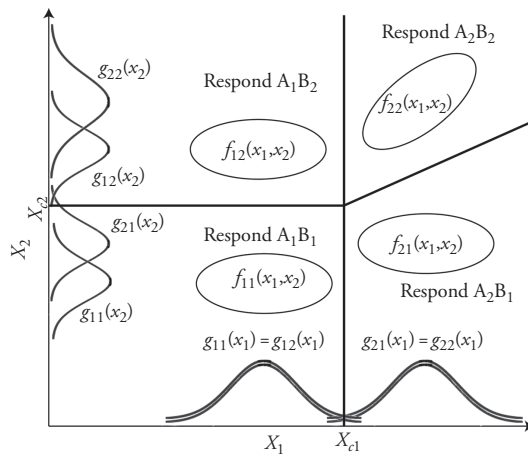


Fig. 2.1 Contours of equal likelihood, decision bounds, and marginal perceptual distributions from a hypothetical multivariate normal GRT model that describes the results of an identification experiment with four stimuli that were constructed by factorially combining two levels of two stimulus dimensions.

$g_{12}(x_1)$ and $g_{21}(x_1) = g_{22}(x_1)$, for all values of x_1 . Thus, the subject's perception of component A does not depend on the level of B and, therefore, stimulus component A is perceptually separable from B. On the other hand, note that the subject's perception of component B does change when the level of component A changes [i.e., $g_{11}(x_1) \neq g_{21}(x_1)$ and $g_{12}(x_1) \neq g_{22}(x_1)$ for most values of x_1]. In particular, when A changes from level 1 to level 2 the subject's mean perceived value of each level of component B increases. Thus, the perception of component B depends on the level of component A and therefore B is not perceptually separable from A.

Finally, note that decisional separability holds on dimension 1 but not on dimension 2. On dimension 1 the decision bound is vertical. Thus, the subject has adopted the following decision rule:

Component A is at level 2 if $x_1 > X_{c1}$; otherwise component A is at level 1.

Where X_{c1} is the criterion on dimension 1 (i.e., the x_1 intercept of the vertical decision bound). Thus, the subject's decision about whether component A is at level 1 or 2 does not depend on the perceived value of component B. So component A is decisionally separable from component B. On the other hand, the decision bound on dimension x_2 is not horizontal, so the criterion used to judge whether component B is at level 1 or 2 changes with the perceived value of component A (at least for

larger perceived values of A). As a result, component B is not decisionally separable from component A.

Applying GRT to Data

The most common applications of GRT are to data collected in an identification experiment like the one modeled in Figure 2.1. The key data from such experiments are collected in a confusion matrix, which contains a row for every stimulus and a column for every response (Table 2.1 displays an example of a confusion matrix, which will be discussed and analyzed later). The entry in row i and column j lists the number of trials on which stimulus S_i was presented and the subject gave response R_j . Thus, the entries on the main diagonal give the frequencies of all correct responses and the off-diagonal entries describe the various errors (or confusions). Note that each row sum equals the total number of stimulus presentations of that type. So if each stimulus is presented 100 times then the sum of all entries in each row will equal 100. This means that there is one constraint per row, so an $n \times n$ confusion matrix will have $n \times (n - 1)$ degrees of freedom.

General recognition theory has been used to analyze data from confusion matrices in two different ways. One is to fit the model to the entire confusion matrix. In this method, a GRT model is constructed with specific numerical values of all of its parameters and a predicted confusion matrix is computed. Next, values of each parameter are found that make the predicted matrix as close as possible to the empirical confusion matrix. To test various assumptions about perceptual and decisional processing—for example, whether perceptual independence holds—a version of the model that assumes perceptual independence is fit to the data as well as a version that makes no assumptions about perceptual independence. This latter version contains the former version as a special case (i.e., in which all covariance parameters are set to zero), so it can never fit worse. After fitting these two models, we assume that perceptual independence is violated if the more general model fits significantly better than the more restricted model that assumes perceptual independence. The other method for using GRT to test assumptions about perceptual processing, which is arguably more popular, is to compute certain summary statistics from the empirical confusion matrix and then to check whether these satisfy certain conditions that are characteristic of perceptual separability or

independence. Because these two methods are so different, we will discuss each in turn.

It is important to note however, that regardless of which method is used, there are certain nonidentifiabilities in the GRT model that could limit the conclusions that are possible to draw from any such analyses (e.g., Menneer, Wenger, & Blaha, 2010; Silbert & Thomas, 2013). The problems are most severe when GRT is applied to 2×2 identification data (i.e., when the stimuli are A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2). For example, Silbert and Thomas (2013) showed that in 2×2 applications where there are two linear decision bounds that do not satisfy decisional separability, there always exists an alternative model that makes the exact same empirical predictions and satisfies decisional separability (and these two models are related by an affine transformation). Thus, decisional separability is not testable with standard applications of GRT to 2×2 identification data (nor can the slopes of the decision bounds be uniquely estimated). For several reasons, however, these nonidentifiabilities are not catastrophic.

First, the problems don't generally exist with 3×3 or larger identification tasks. In the 3×3 case the GRT model with linear bounds requires at least 4 decision bounds to divide the perceptual space into 9 response regions (e.g., in a tic-tac-toe configuration). Typically, two will have a generally vertical orientation and two will have a generally horizontal orientation. In this case, there is no affine transformation that guarantees decisional separability except in the special case where the two vertical-tending bounds are parallel and the two horizontal-tending bounds are parallel (because parallel lines remain parallel after affine transformations). Thus, in 3×3 (or higher) designs, decisional separability is typically identifiable and testable.

Second, there are simple experimental manipulations that can be added to the basic 2×2 identification experiment to test for decisional separability. In particular, switching the locations of the response keys is known to interfere with performance if decisional separability fails but not if decisional separability holds (Maddox, Glass, O'Brien, Filoteo, & Ashby, 2010; for more information on this, see the section later entitled "Neural Implementations of GRT"). Thus, one could add 100 extra trials to the end of a 2×2 identification experiment where the response key locations are randomly interchanged (and participants are informed of this change). If accuracy drops

significantly during this period, then decisional separability can be rejected, whereas if accuracy is unaffected then decisional separability is supported.

Third, one could analyze the 2×2 data using the newly developed GRT model with individual differences (GRT-wIND; Soto, Vucovich, Musgrave, & Ashby, in press), which was patterned after the INDSCAL model of multidimensional scaling (Carroll & Chang, 1970). GRT-wIND is fit to the data from all individuals simultaneously. All participants are assumed to share the same group perceptual distributions, but different participants are allowed different linear bounds and they are assumed to allocate different amounts of attention to each perceptual dimension. The model does not suffer from the identifiability problems identified by Silbert and Thomas (2013), even in the 2×2 case, because with different linear bounds for each participant there is no affine transformation that simultaneously makes all these bounds satisfy decisional separability.

Fitting the GRT Model to Identification Data

COMPUTING THE LIKELIHOOD FUNCTION

When the full GRT model is fit to identification data, the best-fitting values of all free parameters must be found. Ideally, this is done via the method of maximum likelihood—that is, numerical values of all parameters are found that maximize the likelihood of the data given the model. Let S_1, S_2, \dots, S_n denote the n stimuli in an identification experiment and let R_1, R_2, \dots, R_n denote the n responses. Let r_{ij} denote the frequency with which the subject responded R_j on trials when stimulus S_i was presented. Thus, r_{ij} is the entry in row i and column j of the confusion matrix. Note that the r_{ij} are random variables. The entries in each row have a multinomial distribution. In particular, if $P(R_j|S_i)$ is the true probability that response R_j is given on trials when stimulus S_i is presented, then the probability of observing the response frequencies $r_{i1}, r_{i2}, \dots, r_{in}$ in row i equals

$$\frac{n_i!}{r_{i1}!r_{i2}!\dots r_{in}!}P(R_1|S_i)^{r_{i1}}P(R_2|S_i)^{r_{i2}}\dots P(R_n|S_i)^{r_{in}} \quad (6)$$

where n_i is the total number of times that stimulus S_i was presented during the course of the experiment. The probability or joint likelihood of observing the entire confusion matrix is the product

of the probabilities of observing each row; that is,

$$L = \prod_{i=1}^n \frac{n_i}{\prod_{j=1}^n r_{ij}} \prod_{j=1}^n P(R_j|S_i)^{r_{ij}} \quad (7)$$

General recognition theory models predict that $P(R_j|S_i)$ has a specific form. Specifically, they predict that $P(R_j|S_i)$ is the volume in the R_j response region under the multivariate distribution of perceptual effects elicited when stimulus S_i is presented. This requires computing a multiple integral. The maximum likelihood estimators of the GRT model parameters are those numerical values of each parameter that maximize L . Note that the first term in Eq. 7 does not depend on the values of any model parameters. Rather it only depends on the data. Thus, the parameter values that maximize the second term also maximize the whole expression. For this reason, the first term can be ignored during the maximization process. Another common practice is to take logs of both sides of Eq. 7. Parameter values that maximize L will also maximize any monotonic function of L (and log is a monotonic transformation). So, the standard approach is to find values of the free parameters that maximize

$$\sum_{i=1}^n \sum_{j=1}^n r_{ij} \log P(R_j|S_i) \quad (8)$$

ESTIMATING THE PARAMETERS

In the case of the multivariate normal model, the predicted probability $P(R_j|S_i)$ in Eq. 8 equals the volume under the multivariate normal pdf that describes the subject's perceptual experiences on trials when stimulus S_i is presented over the response region associated with response R_j . To estimate the best-fitting parameter values using a standard minimization routine, such integrals must be evaluated many times. If decisional separability is assumed, then the problem simplifies considerably. For example, under these conditions, Wickens (1992) derived the first and second derivatives necessary to quickly estimate parameters of the model using the Newton-Raphson method. Other methods must be used for more general models that do not assume decisional separability. Ennis and Ashby (2003) proposed an efficient algorithm for evaluating the integrals that arise when fitting any GRT model. This algorithm allows the parameters of virtually any GRT model to be estimated via standard minimization software. The remainder of this section describes this method.

The left side of Figure 2.2 shows a contour of equal likelihood from the bivariate normal

distribution that describes the perceptual effects of stimulus S_i , and the solid lines denote two possible decision bounds in this hypothetical task. In Figure 2.2 the bounds are linear, but the method works for any number of bounds that have any parametric form. The shaded region is the R_j response region. Thus, according to GRT, computing $P(R_j|S_i)$ is equivalent to computing the volume under the S_i perceptual distribution in the R_j response region. This volume is indicated by the shaded region in the figure. First note that any linear bound can be written in discriminant function form as

$$h(x_1, x_2) = h(\mathbf{x}) = \mathbf{b}'\mathbf{x} + c = 0 \quad (9)$$

where (in the bivariate case) \mathbf{x} and \mathbf{b}' are the vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{b}' = [b_1 \quad b_2]$$

and c is a constant. The discriminant function form of any decision bound has the property that positive values are obtained if any point on one side of the bound is inserted into the function, and negative values are obtained if any point on the opposite side is inserted. So, for example, in Figure 2.2, the constants b_1 , b_2 , and c can be selected so that $h_1(\mathbf{x}) > 0$ for any point \mathbf{x} above the h_1 bound and $h_1(\mathbf{x}) < 0$ for any point below the bound. Similarly, for the h_2 bound, the constants can be selected so that $h_2(\mathbf{x}) > 0$ for any point to the right of the bound and $h_2(\mathbf{x}) < 0$ for any point to the left. Note that under these conditions, the R_j response region is defined as the set of all \mathbf{x} such that $h_1(\mathbf{x}) > 0$ and $h_2(\mathbf{x}) > 0$. Therefore, if we denote the multivariate normal (mvn) pdf for stimulus S_i as $\text{mvn}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, then

$$P(R_j|S_i) = \iint_{\substack{h_1(\mathbf{x}) > 0; \\ h_2(\mathbf{x}) > 0}} \text{mvn}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) dx_1 dx_2 \quad (10)$$

Ennis and Ashby (2003) showed how to quickly approximate integrals of this type. The basic idea is to transform the problem using a multivariate form of the well-known z transformation. Ennis and Ashby proposed using the Cholesky transformation. Any random vector \mathbf{x} that has a multivariate normal distribution can always be rewritten as

$$\mathbf{x} = \mathbf{P}\mathbf{z} + \boldsymbol{\mu}, \quad (11)$$

where $\boldsymbol{\mu}$ is the mean vector of \mathbf{x} , \mathbf{z} is a random vector with a multivariate \mathbf{z} distribution (i.e., a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix equal to the identity

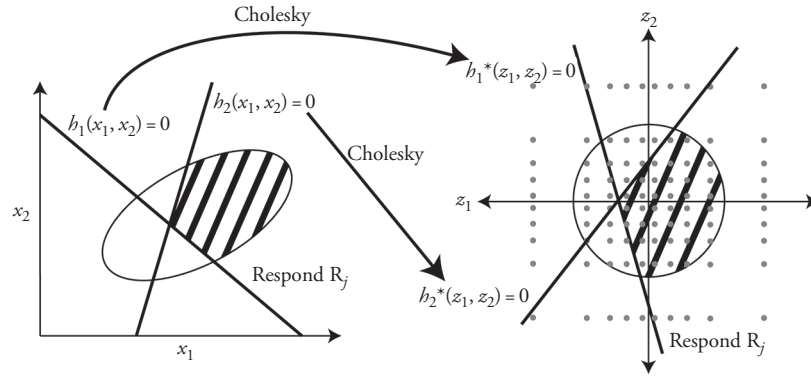


Fig. 2.2 Schematic illustration of how numerical integration is performed in the multivariate normal GRT model via Cholesky factorization.

matrix \mathbf{I}), and \mathbf{P} is a lower triangular matrix such that $\mathbf{P}\mathbf{P}' = \mathbf{\Sigma}$ (i.e., the variance-covariance matrix of \mathbf{x}). If \mathbf{x} is bivariate normal then

$$\mathbf{P} = \begin{bmatrix} \sigma_1 & 0 \\ \frac{\text{cov}_{12}}{\sigma_1} & \sqrt{\sigma_2^2 - \frac{\text{cov}_{12}^2}{\sigma_1^2}} \end{bmatrix} \quad (12)$$

The Cholesky transformation is linear (see Eq. 11), so linear bounds in \mathbf{x} space are transformed to linear bounds in \mathbf{z} space. In particular,

$$h_k(\mathbf{x}) = \mathbf{b}'\mathbf{x} + c = 0$$

becomes

$$h_k(\mathbf{P}\mathbf{z} + \boldsymbol{\mu}) = \mathbf{b}'(\mathbf{P}\mathbf{z} + \boldsymbol{\mu}) + c = 0$$

or equivalently

$$h_k^*(\mathbf{z}) = (\mathbf{b}'\mathbf{P})\mathbf{z} + (\mathbf{b}'\boldsymbol{\mu} + c) = 0 \quad (13)$$

Thus, in this way we can transform the Eq. 10 integral to

$$\begin{aligned} P(R_j | S_i) &= \iint_{\substack{h_1(\mathbf{x}) > 0; \\ h_2(\mathbf{x}) > 0}} \text{mvn}(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i) dx_1 dx_2 \\ &= \iint_{\substack{h_1^*(\mathbf{z}) > 0; \\ h_2^*(\mathbf{z}) > 0}} \text{mvn}(\mathbf{0}, \mathbf{I}) dz_1 dz_2 \quad (14) \end{aligned}$$

The right and left panels of Figure 2.2 illustrate these two integrals. The key to evaluating the second of these integrals quickly is to preload z -values that are centered in equal area intervals. In Figure 2.2 each gray point in the right panel has a z_1 coordinate that is the center¹ of an interval with area 0.10 under the z distribution (since there

are 10 points). Taking the Cartesian product of these 10 points produces a table of 100 ordered pairs (z_1, z_2) that are each the center of a rectangle with volume 0.01 (i.e., 0.10×0.10) under the bivariate z distribution. Given such a table, the Eq. 14 integral is evaluated by stepping through all (z_1, z_2) points in the table. Each point is substituted into Eq. 13 for $k = 1$ and 2 and the signs of $h_1^*(z_1, z_2)$ and $h_2^*(z_1, z_2)$ are determined. If $h_1^*(z_1, z_2) > 0$ and $h_2^*(z_1, z_2) > 0$ then the Eq. 14 integral is incremented by 0.01. If either or both of these signs are negative, then the value of the integral is unchanged. So the value of the integral is approximately equal to the number of (z_1, z_2) points that are in the R_j response region divided by the total number of (z_1, z_2) points in the table. Figure 2.2 shows a 10×10 grid of (z_1, z_2) points, but better results can be expected from a grid with higher resolution. We have had success with a 100×100 grid, which should produce approximations to the integral that are accurate to within 0.0001 (Ashby, Waldron, Lee, & Berkman, 2001).

EVALUATING GOODNESS OF FIT

As indicated before, one popular method for testing an assumption about perceptual or decisional processing is to fit two versions of a GRT model to the data using the procedures outlined in this section. In the first, restricted version of the model, a number of parameters are set to values that reflect the assumption being tested. For example, fixing all correlations to zero would test perceptual independence. In the second, unrestricted version of the model, the same parameters are free to vary. Once the restricted and unrestricted versions of the model have been fit, they can be compared through

a likelihood ratio test:

$$\Lambda = -2(\log L_R - \log L_U), \quad (15)$$

where L_R and L_U represent the likelihoods of the restricted and unrestricted models, respectively. Under the null hypothesis that the restricted model is correct, the statistic Λ has a Chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between the restricted and unrestricted models.

If several non-nested models were fitted to the data, we would usually want to select the best candidate from this set. The likelihood ratio test cannot be used to select among such non-nested models. Instead, we can compute the Akaike information criterion (AIC, Akaike, 1974) or the Bayesian information criterion (BIC, Schwarz, 1978):

$$AIC = -2\log L + 2m \quad (16)$$

$$BIC = -2\log L + m \log N \quad (17)$$

where m is the number of free parameters in the model and N is the number of data points being fit. When the sample size is small compared to the number of free parameters of the model, as in most applications of GRT, a correction factor equal to $2m(m+1)/(n^2 - m - 1)$ should be added to the *AIC* (see Burnham & Anderson, 2004). The best model is the one with the smallest *AIC* or *BIC*.

Because an $n \times n$ confusion matrix has $n(n-1)$ degrees of freedom, the maximum number of free parameters that can be estimated from any confusion matrix is $n(n-1)$. The origin and unit of measurement on each perceptual dimension are arbitrary. Therefore, without loss of generality, the mean vector of one perceptual distribution can be set to $\mathbf{0}$, and all variances of that distribution can be set to 1.0. Therefore, if there are two perceptual dimensions and n stimuli, then the full GRT model has $5(n-1) + 1$ free distributional parameters (i.e., $n-1$ stimuli have 5 free parameters—2 means, 2 variances, and a covariance—and the distribution with mean $\mathbf{0}$ and all variances set to 1 has 1 free parameter—a covariance). If linear bounds are assumed, then another 2 free parameters must be added for every bound (e.g., slope and intercept). With a factorial design (e.g., as when the stimulus set is A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2), there must be at least one bound on each dimension to separate each pair of consecutive component levels. So for stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 , at least two bounds are required (e.g., see Figure 2.1). If, instead, there are 3 levels of each component,

then at least 4 bounds are required. The confusion matrix from a 2×2 factorial experiment has 12 degrees of freedom. The full model has more free parameters than this, so it cannot be fit to the data from this experiment. As a result, some restrictive assumptions are required. In a 3×3 factorial experiment, however, the confusion matrix has 72 degrees of freedom (9×8) and the full model has 49 free parameters (i.e., 41 distributional parameters and 8 decision bound parameters), so the full model can be fit to identification data when there are at least 3 levels of each stimulus dimension. For an alternative to the GRT identification model presented in this section, see Box 2.

Box 2 GRT Versus the Similarity-Choice Model

The most widely known alternative identification model is the similarity-choice model (SCM; Luce, 1963; Shepard, 1957), which assumes that

$$P(R_j|S_i) = \frac{\eta_{ij}\beta_j}{\sum_k \eta_{ik}\beta_k},$$

where η_{ij} is the similarity between stimuli S_i and S_j and β_j is the bias toward response R_j . The SCM has had remarkable success. For many years, it was the standard against which competing models were compared. For example, in 1992 J. E. K. Smith summarized its performance by concluding that the SCM “has never had a serious competitor as a model of identification data. Even when it has provided a poor model of such data, other models have done even less well” (p. 199). Shortly thereafter, however, the GRT model ended this dominance, at least for identification data collected from experiments with stimuli that differ on only a couple of stimulus dimensions. In virtually every such comparison, the GRT model has provided a substantially better fit than the SCM, in many cases with fewer free parameters (Ashby et al., 2001). Even so, it is important to note that the SCM is still valuable, especially in the case of identification experiments in which the stimuli vary on many unknown stimulus dimensions.

The Summary Statistics Approach

The summary statistics approach (Ashby & Townsend, 1986; Kadlec and Townsend, 1992a, 1992b) draws inferences about perceptual independence, perceptual separability, and decisional separability by using summary statistics that are easily computed from a confusion matrix. Consider again the factorial identification experiment with 2 levels of 2 stimulus components. As before, we will denote the stimuli in this experiment as A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . In this case, it is convenient to denote the responses as a_1b_1 , a_1b_2 , a_2b_1 , and a_2b_2 . The summary statistics approach operates by computing certain summary statistics that are derived from the 4×4 confusion matrix that results from this experiment. The statistics are computed at either the macro- or micro-level of analysis.

MACRO-ANALYSES

Macro-analyses draw conclusions about perceptual and decisional separability from changes in accuracy, sensitivity, and bias measures computed for one dimension across levels of a second dimension. One of the most widely used summary statistics in macro-analysis is *marginal response invariance*, which holds for a dimension when the probability of identifying the correct level of that dimension does not depend on the level of any irrelevant dimensions (Ashby & Townsend, 1986). For example, marginal response invariance requires that the probability of correctly identifying that component A is at level 1 is the same regardless of the level of component B, or in other words that

$$P(a_1 | A_1B_1) = P(a_1 | A_1B_2)$$

Now in an identification experiment, A_1 can be correctly identified regardless of whether the level of B is correctly identified, and so

$$P(a_1 | A_1B_1) = P(a_1b_1 | A_1B_1) + P(a_1b_2 | A_1B_1)$$

For this reason, marginal response invariance holds on dimension X_1 if and only if

$$\begin{aligned} P(a_1b_1 | A_iB_1) + P(a_1b_2 | A_iB_1) \\ = P(a_1b_1 | A_iB_2) + P(a_1b_2 | A_iB_2) \end{aligned} \quad (18)$$

for both $i = 1$ and 2. Similarly, marginal response invariance holds on dimension X_2 if and only if

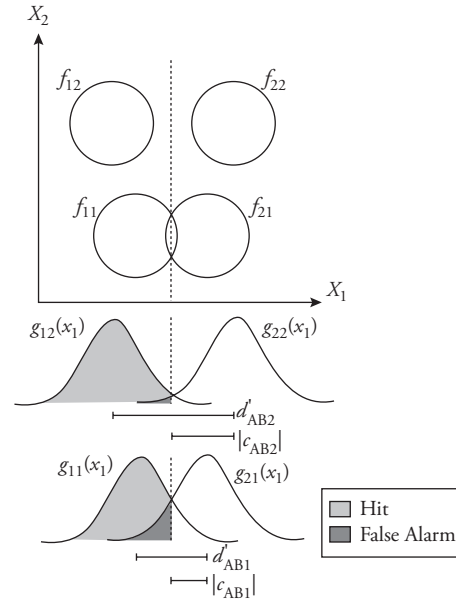


Fig. 2.3 Diagram explaining the relation between macro-analytic summary statistics and the concepts of perceptual and decisional separability.

$$\begin{aligned} P(a_1b_j | A_1B_j) + P(a_2b_j | A_1B_j) \\ = P(a_1b_j | A_2B_j) + P(a_2b_j | A_2B_j) \end{aligned} \quad (19)$$

for both $j = 1$ and 2.

Marginal response invariance is closely related to perceptual and decisional separability. In fact, if dimension X_1 is perceptually and decisionally separable from dimension X_2 , then marginal response invariance must hold for X_1 (Ashby & Townsend, 1986). In the later section entitled “Extensions to Response Time,” we describe how an even stronger test is possible with a response time version of marginal response invariance. Figure 2.3 helps to understand intuitively why perceptual and decisional separability together imply marginal response invariance. The top of the figure shows the perceptual distributions of four stimuli that vary on two dimensions. Dimension X_1 is decisionally but not perceptually separable from dimension X_2 ; the distance between the means of the perceptual distributions along the X_1 axis is much greater for the top two stimuli than for the bottom two stimuli. The marginal distributions at the bottom of Figure 2.3 show that the proportion of correct responses, represented by the light-grey areas under the curves, is larger in the second level of X_2 than in the first level. The result would be similar if perceptual separability held and decisional separability failed, as would be the case for X_2

if its decision bound was not perpendicular to its main axis.

To test marginal response invariance in dimension X_1 , we estimate the various probabilities in Eq. 18 from the empirical confusion matrix that results from this identification experiment. Next, equality between the two sides of Eq. 18 is assessed via a standard statistical test. These computations are repeated for both levels of component A and if either of the two tests is significant, then we conclude that marginal response invariance fails, and, therefore, that either perceptual or decisional separability are violated.

The left side of Eq. 18 equals $P(a_i|A_iB_1)$ and the right side equals $P(a_i|A_iB_2)$. These are the probabilities that component A_i is correctly identified and are analogous to “hit” rates in signal detection theory. To emphasize this relationship, we define the *identification hit rate* of component A_i on trials when stimulus A_iB_j is presented as

$$H_{a_i|A_iB_j} = P(a_i|A_iB_j) = P(a_ib_1|A_iB_j) + P(a_ib_2|A_iB_j) \quad (20)$$

The analogous false alarm rates can be defined similarly. For example,

$$F_{a_2|A_1B_j} = P(a_2|A_1B_j) = P(a_2b_1|A_1B_j) + P(a_2b_2|A_1B_j) \quad (21)$$

In Figure 2.3, note that the dark grey areas in the marginal distributions equal $F_{a_1|A_2B_2}$ (top) and $F_{a_1|A_2B_1}$ (bottom). In signal detection theory, hit and false-alarm rates are used to measure stimulus discriminability (i.e., d'). We can use the identification analogues to compute *marginal discriminabilities* for each stimulus component (Thomas, 1999). For example,

$$d'_{AB_j} = \Phi^{-1}(H_{a_2|A_2B_j}) - \Phi^{-1}(F_{a_2|A_1B_j}) \quad (22)$$

where the function Φ^{-1} is the inverse cumulative distribution function for the standard normal distribution. As shown in Figure 2.3, the value of d'_{AB_j} represents the standardized distance between the means of the perceptual distributions of stimuli A_1B_j and A_2B_j . If component A is perceptually separable from component B, then the marginal discriminabilities between the two levels of A must be the same for each level of B – that is, $d'_{AB_1} = d'_{AB_2}$ (Kadlec & Townsend, 1992a, 1992b). Thus, if this equality fails, then perceptual separability is

violated. The equality between two d' s can be tested using the following statistic (Marascuilo, 1970):

$$Z = \frac{d'_1 - d'_2}{\sqrt{s_{d'_1}^2 + s_{d'_2}^2}} \quad (23)$$

where

$$s_{d'}^2 = \frac{F(1-F)}{n_n \{ \phi [\Phi^{-1}(F)] \}^2} + \frac{H(1-H)}{n_s \{ \phi [\Phi^{-1}(H)] \}^2} \quad (24)$$

where ϕ is the standard normal probability density function, H and F are the hit and false-alarm rates associated with the relevant d' , n_n is the number of trials used to compute F , and n_s is the number of trials used to compute H . Under the null hypothesis of equal d' s, Z follows a standard normal distribution.

Marginal hit and false-alarm rates can also be used to compute a *marginal response criterion*. Several measures of response criterion and bias have been proposed (see Chapter 2 of Macmillan & Creelman, 2005), but perhaps the most widely used criterion measure in recent years (due to Kadlec, 1999) is:

$$c_{AB_j} = \Phi^{-1}(F_{a_1|A_2B_j}) \quad (25)$$

As shown in Figure 2.3, this measure represents the placement of the decision-bound relative to the center of the A_2B_j distribution. If component A is perceptually separable from component B, but $c_{AB_1} \neq c_{AB_2}$, then decisional separability must have failed on dimension X_1 (Kadlec & Townsend, 1992a, 1992b). On the other hand, if perceptual separability is violated, then examining the marginal response criteria provides no information about decisional separability. To understand why this is the case, note that in Figure 2.3 the marginal c values are not equal, even though decisional separability holds. A failure of perceptual separability has affected measures of both discriminability and response criteria.

To test the difference between two c values, the following test statistic can be used (Kadlec, 1999):

$$Z = \frac{c_1 - c_2}{\sqrt{s_{c_1}^2 + s_{c_2}^2}} \quad (26)$$

where

$$s_c^2 = \frac{F(1-F)}{n_n \{ \phi [\Phi^{-1}(F)] \}^2}. \quad (27)$$

MICRO-ANALYSES

Macro-analyses focus on properties of the entire stimulus ensemble. In contrast, micro-analyses test assumptions about perceptual independence and decisional separability by examining summary statistics computed for only one or two stimuli.

The most widely used test of perceptual independence is via *sampling independence*, which holds when the probability of reporting a combination of components $P(a_i b_j)$ equals the product of the probabilities of reporting each component alone, $P(a_i)P(b_j)$. For example, sampling independence holds for stimulus $A_1 B_1$ if and only if

$$\begin{aligned} P(a_1 b_1 | A_1 B_1) &= P(a_1 | A_1 B_1) \times P(b_1 | A_1 B_1) \\ &= [P(a_1 b_1 | A_1 B_1) + P(a_1 b_2 | A_1 B_1)] \\ &\quad \times [P(a_1 b_1 | A_1 B_1) + P(a_2 b_1 | A_1 B_1)] \end{aligned} \quad (28)$$

Sampling independence provides a strong test of perceptual independence if decisional separability holds. In fact, if decisional separability holds on both dimensions, then sampling independence holds if and only if perceptual independence holds (Ashby & Townsend, 1986). Figure 2.4A gives an intuitive illustration of this theoretical result. Two cases are presented in which decisional separability holds on both dimensions and the decision bounds cross at the mean of the perceptual distribution. In the distribution to the left, perceptual independence holds and it is easy to see that all four responses are equally likely. Thus, the volume of this bivariate normal distribution in response region $R_4 = a_2 b_2$ is 0.25. It is also easy to see that half of each marginal distribution lies above its relevant decision criterion (i.e., the two shaded regions), so $P(a_2) = P(b_2) = 0.5$. As a result, sampling independence is satisfied since $P(a_2 b_2) = P(a_2) \times P(b_2)$. It turns out that this relation holds regardless of where the bounds are placed, as long as they remain perpendicular to the dimension that they divide. The distribution to the right of Figure 2.4A has the same variances as the previous distribution, and, therefore, the same marginal response proportions for a_2 and b_2 . However, in this case, the covariance is larger than zero and it is clear that $P(a_2 b_2) > 0.25$.

Perceptual independence can also be assessed through discriminability and criterion measures computed for one dimension conditioned on the perceived value on the other dimension. Figure 2.4B shows the perceptual distributions of two stimuli that share the same level of component B (i.e., B_1) and have the same perceptual mean on

dimension X_2 . The decision bound perpendicular to X_2 separates the perceptual plane into two regions: percepts falling in the upper region elicit an incorrect response on component B (i.e., a miss for B), whereas percepts falling in the lower region elicit a correct B response (i.e., a hit). The bottom of the figure shows the marginal distribution for each stimulus conditioned on whether B is a hit or a miss. When perceptual independence holds, as is the case for the stimulus to the left, these conditional distributions have the same mean. On the other hand, when perceptual independence does not hold, as is the case for the stimulus to the right, the conditional distributions have different means, which is reflected in different d' and c values depending on whether there is a hit or a miss on B. If decisional separability holds, differences in the conditional d' s and c s are evidence of violations of perceptual independence (Kadlec & Townsend, 1992a, 1992b).

Conditional d' and c values can be computed from hit and false alarm rates for two stimuli differing in one dimension, conditioned on the reported level of the second dimension. For example, for the pair $A_1 B_1$ and $A_2 B_1$, conditioned on a hit on B, the hit rate for A is $P(a_1 b_1 | A_1 B_1)$ and the false alarm rate is $P(a_1 b_1 | A_2 B_1)$. Conditioned on a miss on B, the hit rate for A is $P(a_1 b_2 | A_1 B_1)$ and the false alarm rate is $P(a_1 b_2 | A_2 B_1)$. These values are used as input to Eqs. 22–27 to reach a statistical conclusion.

Note that if perceptual independence and decisional separability both hold, then the tests based on sampling independence and equal conditional d' and c should lead to the same conclusion. If only one of these two tests holds and the other fails, this indicates a violation of decisional separability (Kadlec & Townsend, 1992a, 1992b).

An Empirical Example

In this section we show with a concrete example how to analyze the data from an identification experiment using GRT. We will first analyze the data by fitting GRT models to the identification confusion matrix, and then we will conduct summary statistics analyses on the same data. Finally, we will compare the results from the two separate analyses. Imagine that you are a researcher interested in how the age and gender of faces interact during face recognition. You run an experiment in which subjects must identify four stimuli, the combination of two levels of age (teen and adult) and two levels of gender (male and female). Each stimulus is presented 250 times, for a total of 1,000

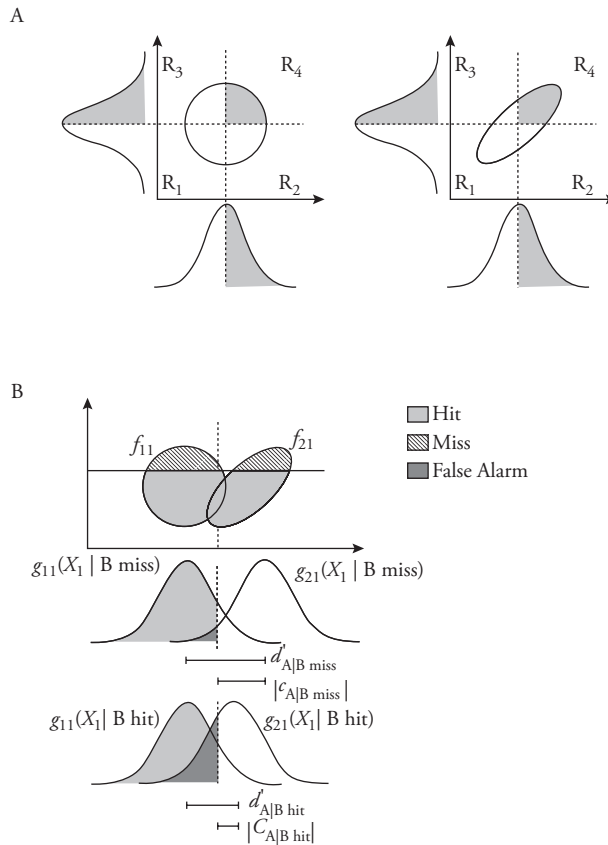


Fig. 2.4 Diagram explaining the relation between micro-analytic summary statistics and the concepts of perceptual independence and decisional separability. Panel A focuses on sampling independence and Panel B on conditional signal detection measures.

trials in the whole experiment. The data to be analyzed are summarized in the confusion matrix displayed in Table 2.1. These data were generated by random sampling from the model shown in Figure 2.5A.

The advantage of generating artificial data from this model is that we know in advance what conclusions should be reached by our analyses. For example, note that decisional separability holds in the Figure 2.5A model. Also, because the distance between the “male” and “female” distributions is larger for “adult” than for “teen,” gender is not perceptually separable from age. In contrast, the “adult” and “teen” marginal distributions are the same across levels of gender, so age is perceptually separable from gender. Finally, because all distributions show a positive correlation, perceptual independence is violated for all stimuli.

A hierarchy of models were fit to the data in Table 2.1 using maximum likelihood estimation (as in Ashby et al., 2001; Thomas, 2001). Because there are only 12 degrees of freedom in the

data, some parameters were fixed for all models. Specifically, all variances were assumed to be equal to one and decisional separability was assumed for both dimensions. Figure 2.5C shows the hierarchy of models used for the analysis, together with the number of free parameters m for each of them. In this figure, PS stands for perceptual separability, PI for perceptual independence, DS for decisional separability and 1_RHO describes a model with a single correlation parameter for all distributions. Note that several other models could be tested, depending on specific research goals and hypotheses, or on the results from summary statistics analysis.

The arrows in Figure 2.5C connect models that are nested within each other. The result of likelihood ratio tests comparing such nested models are displayed next to each arrow, with an asterisk representing significantly better fit for the more general model (lower in the hierarchy) and n.s. representing a nonsignificant difference in fit. Starting at the top of the hierarchy, it

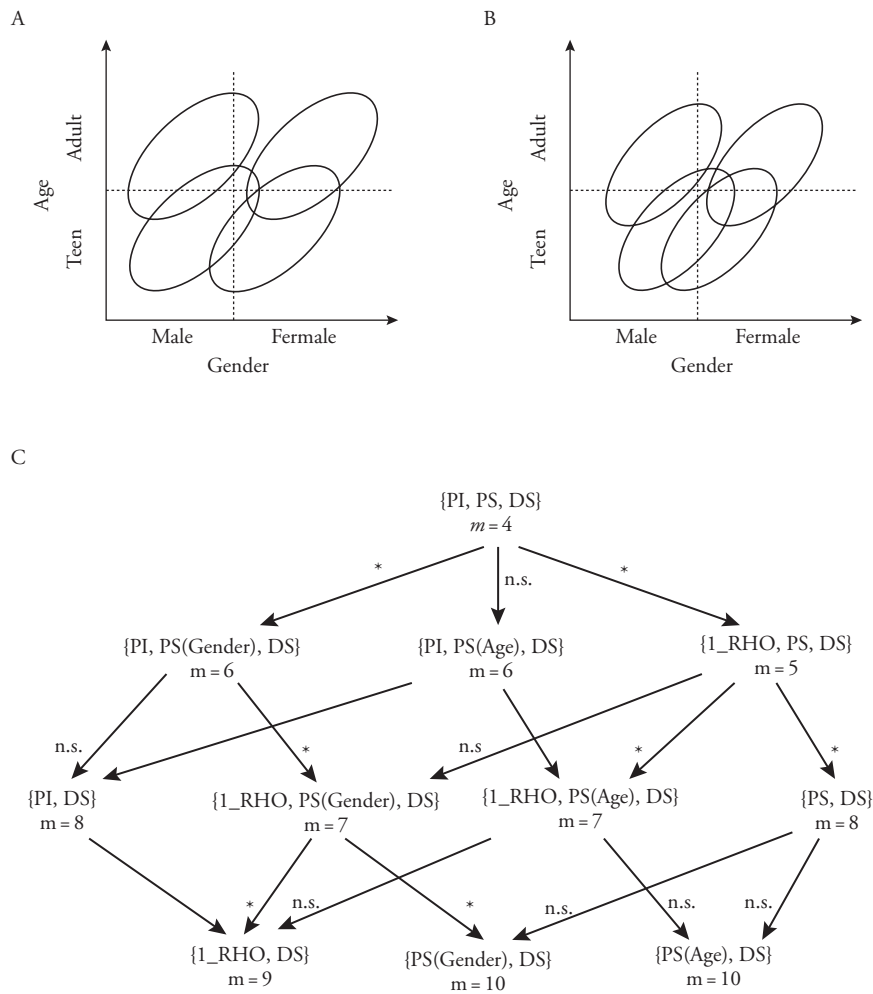


Fig. 2.5 Results of analyzing the data in Table 2.1 with GRT. Panel A shows the GRT model that was used to generate the data. Panel B shows the recovered model from the model fitting and selection process. Panel C shows the hierarchy of models used for the analysis and the number of free parameters (m) in each. PI stands for perceptual independence, PS for perceptual separability, DS for decisional separability and 1_RHO for a single correlation in all distributions.

Table 2.1. Data from a simulated identification experiment with four face stimuli, created by factorially combining two levels of gender (male and female) and two levels of age (teen and adult).

Stimulus	Response			
	Male/Teen	Female/Teen	Male/Adult	Female/Adult
Male/Teen	140	36	34	40
Female/Teen	89	91	4	66
Male/Adult	85	5	90	70
Female/Adult	20	59	8	163

Table 2.2. Results of the summary statistics analysis for the simulated Gender \times Age identification experiment.

Macroanalyses					
Marginal Response Invariance					
Test	Result	Conclusion			
Equal P (Gender=Male) across all Ages	$z = -0.09, p > .1$	Yes			
Equal P (Gender=Female) across all Ages	$z = -7.12, p < .001$	No			
Equal P (Age=Teen) across all Genders	$z = -0.39, p > .1$	Yes			
Equal P (Age=Adult) across all Genders	$z = -1.04, p > .1$	Yes			
Marginal d'					
Test	d' for level 1	d' for level 2	Result	Conclusion	
Equal d' for Gender across all Ages	0.84	1.74	$z = -5.09, p < .001$	No	
Equal d' for Age across all Genders	0.89	1.06	$z = -1.01, p > .1$	Yes	
Marginal c					
Test	c for level 1	c for level 2	Result	Conclusion	
Equal c for Gender across all Ages	-0.33	-1.22	$z = 6.72, p < .001$	No	
Equal c for Age across all Genders	-0.36	-0.48	$z = 1.04, p > .1$	Yes	
Microanalyses					
Sampling Independence					
Stimulus	Response	Expected Proportion	Observed Proportion	Result	Conclusion
Male/Teen	Male/Teen	0.49	0.56	$z = 1.57, p > .1$	Yes
Male/Teen	Female/Teen	0.21	0.14	$z = -2.05, p < .05$	No
Male/Teen	Male/Adult	0.21	0.14	$z = -2.24, p < .05$	No
Male/Teen	Female/Adult	0.09	0.16	$z = 2.29, p < .05$	No
Female/Teen	Male/Teen	0.27	0.36	$z = 2.14, p < .05$	No
Female/Teen	Female/Teen	0.45	0.36	$z = -2.01, p < .05$	No
Female/Teen	Male/Adult	0.23	0.02	$z = -7.80, p < .001$	No
Female/Teen	Female/Adult	0.39	0.26	$z = -3.13, p < .01$	No
Male/Adult	Male/Teen	0.25	0.34	$z = 2.17, p < .05$	No
Male/Adult	Female/Teen	0.11	0.02	$z = -4.09, p < .001$	No
Male/Adult	Male/Adult	0.21	0.36	$z = 3.77, p < .001$	No
Male/Adult	Female/Adult	0.09	0.28	$z = 5.64, p < .001$	No
Female/Adult	Male/Teen	0.04	0.08	$z = 2.15, p < .05$	No
Female/Adult	Female/Teen	0.28	0.24	$z = -1.14, p > .1$	Yes
Female/Adult	Male/Adult	0.10	0.03	$z = -3.07, p < .01$	No
Female/Adult	Female/Adult	0.79	0.65	$z = -3.44, p < .001$	No
Conditional d'					
Test	d' Hit	d' Miss	Result	Conclusion	
Equal d' for Gender when Age=Teen	0.84	1.48	$z = -2.04, p < .05$	No	
Equal d' for Gender when Age=Adult	1.83	2.26	$z = -1.27, p > .1$	Yes	
Equal d' for Age when Gender=Male	0.89	1.44	$z = -1.80, p > .05$	Yes	
Equal d' for Age when Gender=Female	0.83	1.15	$z = -0.70, p > .1$	Yes	

Table 2.2. Continued

Conditional c				
Test	c Hit	c Miss	Result	Conclusion
Equal c for Gender when Age=Teen	-0.014	-1.58	$z = 6.03, p < .001$	No
Equal c for Gender when Age=Adult	-1.68	-0.66	$z = -4.50, p < .001$	No
Equal c for Age when Gender=Male	-0.04	-1.50	$z = 6.05, p < .001$	No
Equal c for Age when Gender=Female	-0.63	0.57	$z = -4.46, p < .001$	No

is possible to find the best candidate models by following the arrows with an asterisk on them down the hierarchy. This leaves the following candidate models: {PS, DS}, {1_RHO, PS(Age), DS}, {1_RHO, DS}, and {PS(Gender), DS}. From this list, we eliminate {1_RHO, DS} because it does not fit significantly better than the more restricted model {1_RHO, PS(Age), DS}. We also eliminate {PS(Gender), DS} because it does not fit better than the more restricted model {PS, DS}. This leaves two candidate models that cannot be compared through a likelihood ratio test, because they are not nested: {PS, DS} and {1_RHO, PS(Age), DS}.

To compare these two models, we can use the *BIC* or *AIC* goodness-of-fit measures introduced earlier. The smallest corrected *AIC* was found for the model {1_RHO, PS(Age), DS} (2,256.43, compared to 2,296.97 for its competitor). This leads to the conclusion that the model that fits these data best assumes perceptual separability of age from gender, violations of perceptual separability of gender from age, and violations of perceptual independence. This model is shown in Figure 2.5B, and it perfectly reproduces the most important features of the model that was used to generate the data. However, note that the quality of this fit depends strongly on the fact that the assumptions used for all the models in Figure 2.5C (decisional separability and all variances equal) are correct in the true model. This will not be the case in many applications, which is why it is always a good idea to complement the model-fitting results with an analysis of summary statistics.

The results from the summary statistics analysis are shown in Table 2.2. The interested reader can directly compute all the values in this table from the data in the confusion matrix (Table 2.1). The macro-analytic tests indicate violations of marginal response invariance, and unequal marginal d' and c values for the gender dimension, both of which suggest that gender is not perceptually separable from age. These results are uninformative about decisional separability. Marginal response

invariance, equal marginal d' and c values all hold for the age dimension, providing some weak evidence for perceptual and decisional separability of age from gender.

The micro-analytic tests show violations of sampling independence for all stimuli, and conditional c values that are significantly different for all stimulus pairs, suggesting possible violations of perceptual independence and decisional separability. Note that if we assumed decisional separability, as we did to fit models to the data, the results of the micro-analytic tests would lead to the conclusion of failure of perceptual independence.

Thus, the results of the model fitting and summary statistics analyses converge to similar conclusions, which is not uncommon for real applications of GRT. These conclusions turn out to be correct in our example, but note that several of them depend heavily on making correct assumptions about decisional separability and other features of the perceptual and decisional processes generating the observed data.

Extensions to Response Time

There have been a number of extensions of GRT that allow the theory to account both for response accuracy and response time (RT). These have differed in the amount of extra theoretical structure that was added to the theory described earlier. One approach was to add the fewest and least controversial assumptions possible that would allow GRT to make RT predictions. The resulting model succeeds, but it offers no process interpretation of how a decision is reached on each trial. An alternative approach is to add enough theoretical structure to make RT predictions *and* to describe the perceptual and cognitive processes that generated that decision. We describe each of these approaches in turn.

The RT-Distance Hypothesis

In standard univariate signal detection theory, the most common RT assumption is that RT

decreases with the distance between the perceptual effect and the response criterion (Bindra, Donderi, & Nishisato, 1968; Bindra, Williams, & Wise, 1965; Emmerich, Gray, Watson, & Tanis, 1972; Smith, 1968). The obvious multivariate analog of this, which is known as the *RT-distance hypothesis*, assumes that RT decreases with the distance between the percept and the decision bound. Considerable experimental support for the RT-distance hypothesis has been reported in categorization experiments in which there is only one decision bound and where more observability is possible (Ashby, Boynton, & Lee, 1994; Maddox, Ashby, & Gottlob, 1998). Efforts to incorporate the RT-distance hypothesis into GRT have been limited to two-choice experimental paradigms, such as categorization or speeded classification, which can be modeled with a single decision bound.

The most general form of the RT-distance hypothesis makes no assumptions about the parametric form of the function that relates RT and distance to bound. The only assumption is that this function is monotonically decreasing. Specific functional forms are sometimes assumed. Perhaps the most common choice is to assume that RT decreases exponentially with distance to bound (Maddox & Ashby, 1996; Murdock, 1985). An advantage of assuming a specific functional form is that it allows direct fitting to empirical RT distributions (Maddox & Ashby, 1996).

Even without any parametric assumptions, however, monotonicity by itself is enough to derive some strong results. For example, consider a *filtering task* with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 , and two perceptual dimensions X_1 and X_2 , in which the subject's task on each trial is to name the level of component A. Let $P_{FA}(RT_i \leq t | A_iB_j)$ denote the probability that the RT is less than or equal to some value t on trials of a filtering task when the subject correctly classified the level of component A. Given this, then the RT analog of marginal response invariance, referred to as *marginal RT invariance*, can be defined as (Ashby & Maddox, 1994)

$$P_{FA}(RT_i \leq t | A_iB_1) = P_{FA}(RT_i \leq t | A_iB_2) \quad (29)$$

for $i = 1$ and 2 and for all $t > 0$. Now assume that the weak version of the RT-distance hypothesis holds (i.e., where no functional form for the RT-distance relationship is specified) and that decisional separability also holds. Then Ashby and Maddox (1994) showed that

perceptual separability holds *if and only if* marginal RT invariance holds for both correct and incorrect responses. Note that this is an if and only if result, which was not true for marginal response invariance. In particular, if decisional separability and marginal response invariance both hold, perceptual separability could still be violated. But if decisional separability, marginal RT invariance, and the RT-distance hypothesis all hold, then perceptual separability must be satisfied. The reason we get the stronger result with RTs is that marginal RT invariance requires that Eq. 29 holds for all values of t , whereas marginal response invariance only requires a single equality to hold. A similar strong result could be obtained with accuracy data if marginal response invariance were required to hold for all possible placements of the response criterion (i.e., the point where the vertical decision bound intersects the X_1 axis).

Process Models of RT

At least three different process models have been proposed that account for both RT and accuracy within a GRT framework. Ashby (1989) proposed a stochastic interpretation of GRT that was instantiated in a discrete-time linear system. In effect, the model assumed that each stimulus component provides input into a set of parallel (and linear) mutually interacting perceptual channels. The channel outputs describe a point that moves through a multidimensional perceptual space during processing. With long exposure durations the percept settles into an equilibrium state, and under these conditions the model becomes equivalent to the static version of GRT. However, the model can also be used to make predictions in cases of short exposure durations and when the subject is operating under conditions of speed stress. In addition, this model makes it possible to relate properties like perceptual separability to network architecture. For example, a sufficient condition for perceptual separability to hold is that there is no crossing of the input lines and no crosstalk between channels.

Townsend, Houpt, and Silbert (2012) considerably generalized the stochastic model proposed by Ashby (1989) by extending it to a broad class of parallel processing models. In particular, they considered (almost) any model in which processing on each stimulus dimension occurs in parallel and the stimulus is identified as soon as processing finishes on all dimensions. They began by extending definitions of key GRT concepts, such as perceptual

and decisional separability and perceptual independence, to this broad class of parallel models. Next, under the assumption that decisional separability holds, they developed many RT versions of the summary statistics tests considered earlier in this chapter.

Ashby (2000) took a different approach. Rather than specify a processing architecture, he proposed that moment-by-moment fluctuations in the percept could be modeled via a continuous-time multivariate diffusion process. In two-choice tasks with one decision bound, a signed distance is computed to the decision bound at each point in time; that is, in one response region simple distance-to-bound is computed (which is always positive), but in the response region associated with the contrasting response the negative of distance to bound is computed. These values are then continuously integrated and this cumulative value drives a standard diffusion process with two absorbing barriers—one associated with each response. This stochastic version of GRT is more biologically plausible than the Ashby (1989) version (e.g., see Smith & Ratcliff, 2004) and it establishes links to the voluminous work on diffusion models of decision making.

Neural Implementations of GRT

Of course, the perceptual and cognitive processes modeled by GRT are mediated by circuits in the brain. During the past decade or two, much has been learned about the architecture and functioning of these circuits. Perhaps most importantly, there is now overwhelming evidence that humans have multiple neuroanatomically and functionally distinct learning systems (Ashby & Maddox, 2005; Eichenbaum, & Cohen, 2004; Squire, 1992). And most relevant to GRT, the evidence is good that the default decision strategy of one of these systems is decisional separability.

The most complete description of two of the most important learning systems is arguably provided by the COVIS theory of category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby, Paul, & Maddox, 2011). COVIS assumes separate rule-based and procedural-learning categorization systems that compete for access to response production. The rule-based system uses executive attention and working memory to select and test simple verbalizable hypotheses about category membership. The procedural system gradually associates categorization responses with regions of perceptual space via reinforcement learning.

COVIS assumes that rule-based categorization is mediated by a broad neural network that includes the prefrontal cortex, anterior cingulate, head of the caudate nucleus, and the hippocampus, whereas the key structures in the procedural-learning system are the striatum and the premotor cortex.

Virtually all decision rules that satisfy decisional separability are easily verbalized. In fact, COVIS assumes that the rule-based system is constrained to use rules that satisfy decisional separability (at least piecewise). In contrast, the COVIS procedural system has no such constraints. Instead, it tends to learn decision strategies that approximate the optimal bound. As we have seen, decisional separability is optimal only under some special, restrictive conditions. Thus, as a good first approximation, one can assume that decisional separability holds if subjects use their rule-based system, and that decisional separability is likely to fail if subjects use their procedural system. A large literature establishes conditions that favor one system over the other. Critical features include the nature of the optimal decision bound, the instructions given to the subjects, and the nature and timing of the feedback, to name just a few (e.g., Ashby & Maddox, 2005, 2010). For example, Ashby et al. (2001) fit the full GRT identification model to data from two experiments. In both, 9 similar stimuli were constructed by factorially combining 3 levels of the same 2 stimulus components. Thus, in stimulus space, the nine stimuli had the same 3×3 grid configuration in both experiments. In the first experiment however, subjects were shown this configuration beforehand and the response keypad had the same 3×3 grid as the stimuli. In the second experiment, the subjects were not told that the stimuli fell into a grid. Instead, the 9 stimuli were randomly assigned responses from the first 9 letters of the alphabet. In the first experiment, where subjects knew about the grid structure, the best-fitting GRT model assumed decisional separability on both stimulus dimensions. In the second experiment, where subjects lacked this knowledge, the decision bounds of the best-fitting GRT model violated decisional separability. Thus, one interpretation of these results is that the instructions biased subjects to use their rule-based system in the first experiment and their procedural system in the second experiment.

As we have consistently seen throughout this chapter, decisional separability greatly simplifies applications of GRT to behavioral data. Thus, researchers who want to increase the probability

that their subjects use decision strategies that satisfy decisional separability should adopt experimental procedures that encourage subjects to use their rule-based learning system. For example, subjects should be told about the factorial nature of the stimuli, the response device should map onto this factorial structure in a natural way, working memory demands should be minimized (e.g., avoid dual tasking) to ensure that working memory capacity is available for explicit hypothesis testing (Waldron & Ashby, 2001), and the intertrial interval should be long enough so that subjects have sufficient time to process the meaning of the feedback (Maddox, Ashby, Ing, & Pickering, 2004).

Conclusions

Multidimensional signal detection theory in general, and GRT in particular, make two fundamental assumptions, namely that every mental state is noisy and that every action requires a decision. When signal detection theory was first proposed, both of these assumptions were controversial. We now know, however, that every sensory, perceptual, or cognitive process must operate in the presence of inherent noise. There is inevitable noise in the stimulus (e.g., photon noise, variability in viewpoint) at the neural level and in secondary factors, such as attention and motivation. Furthermore, there is now overwhelming evidence that every volitional action requires a decision of some sort. In fact, these decisions are now being studied at the level of the single neuron (e.g., Shadlen & Newsome, 2001). Thus, multidimensional signal detection theory captures two fundamental features of almost all behaviors. Beyond these two assumptions, however, the theory is flexible enough to model a wide variety of decision processes and sensory and perceptual interactions. For these reasons, the popularity of multidimensional signal detection theory is likely to grow in the coming decades.

Acknowledgments

Preparation of this chapter was supported in part by Award Number P01NS044393 from the National Institute of Neurological Disorders and Stroke, by grant FA9550-12-1-0355 from the Air Force Office of Scientific Research, and by support from the U.S. Army Research Office through the Institute for Collaborative Biotechnologies under grant W911NF-07-1-0072.

Note

1. This is true except for the endpoints. Since these endpoint intervals have infinite width, the endpoints are set at the z -value that has equal area to the right and left in that interval (0.05 in Figure 2.2).

Glossary

Absorbing barriers: Barriers placed around a diffusion process that terminate the stochastic process upon first contact. In most cases there is one barrier for each response alternative.

Affine transformation: A transformation from an n -dimensional space to an m -dimensional space of the form $y = Ax + b$, where A is an $m \times n$ matrix and b is a vector.

Categorization experiment: An experiment in which the subject's task is to assign the presented stimulus to the category to which it belongs. If there are n different stimuli then a categorization experiment must include fewer than n separate response alternatives.

d' : A measure of discriminability from signal detection theory, defined as the standardized distance between the means of the signal and noise perceptual distributions (i.e., the mean difference divided by the common standard deviation).

Decision bound: The set of points separating regions of perceptual space associated with contrasting responses.

Diffusion process: A stochastic process that models the trajectory of a microscopic particle suspended in a liquid and subject to random displacement because of collisions with other molecules.

Euclidean space: The standard space taught in high-school geometry constructed from orthogonal axes of real numbers. Frequently, the n -dimensional Euclidean space is denoted by \mathfrak{R}^n .

False Alarm: Incorrectly reporting the presence of a signal when no signal was presented.

Hit: Correctly reporting the presence of a presented signal.

Identification experiment: An experiment in which the subject's task is to identify each stimulus uniquely. Thus, if there are n different stimuli, then there must be n separate response alternatives. Typically, on each trial, one stimulus is selected randomly and presented to the subject. The subject's task is to choose the response alternative that is uniquely associated with the presented stimulus.

Likelihood ratio: The ratio of the likelihoods associated with two possible outcomes. If the two trial types are equally likely, then accuracy is maximized when the subject gives one response if the likelihood ratio is greater than 1 and the other response if the likelihood ratio is less than 1.

Multidimensional scaling: A statistical technique in which objects or stimuli are situated in a multidimensional space in such a way that objects that are judged or perceived as similar are placed close together. In most approaches, each object is represented as a single point and the space is constructed from some type of proximity data collected on the to-be-scaled objects. A common choice is to collect similarity ratings on all possible stimulus pairs.

Nested mathematical models: Two mathematical models are nested if one is a special case of the other in which the restricted model is obtained from the more general model by fixing one or more parameters to certain specific values.

Nonidentifiable models: The case where two seemingly different models make identical predictions.

Perceptual dimension: A range of perceived values of some psychologically primary component of a stimulus.

Procedural learning: Learning that improves incrementally with practice and requires immediate feedback after each response. Prototypical examples include the learning of athletic skills and learning to play a musical instrument.

Response bias: The tendency to favor one response alternative in the face of equivocal sensory information. When the frequencies of different trial types are equal, a response bias occurs in signal detection theory whenever the response criterion is set at any point for which the likelihood ratio is unequal to 1.

Response criterion: In signal detection theory, this is the point on the sensory dimension that separates percepts that elicit one response (e.g., Yes) from percepts that elicit the contrasting response (e.g., No).

Speeded classification: An experimental task in which the subject must quickly categorize the stimulus according to the level of a single stimulus dimension. A common example is the filtering task.

Statistical decision theory: The statistical theory of optimal decision-making.

Striatum: A major input structure within the basal ganglia that includes the caudate nucleus and the putamen.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Amazeen, E. L., & DaSilva, F. (2005). Psychophysical test for the independence of perception and action. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 170.
- Ashby, F. G. (1989). Stochastic general recognition theory. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms and models* (pp. 435–457). Amsterdam, Netherlands: Elsevier Science Publishers.
- Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Erlbaum.
- Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*, 310–329.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, *55*, 11–27.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.
- Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S.C. Masin (Ed.), *Foundations of perceptual theory* (pp. 369–399). Amsterdam, Netherlands: Elsevier Science.
- Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, *38*, 423–466.
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.
- Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.
- Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 65–87). New York, NY: Cambridge University Press.
- Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*, 124–150.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, *130*, 77–96.
- Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*, 267–273.
- Bindra, D., Donderi, D. C., & Nishisato, S. (1968). Decision latencies of “same” and “different” judgments. *Perception & Psychophysics*, *3*(2), 121–136.
- Bindra, D., Williams, J. A., & Wise, J. S. (1965). Judgments of sameness and difference: Experiments on decision time. *Science*, *150*, 1625–1627.
- Blaha, L., Silbert, N., & Townsend, J. (2011). A general recognition theory study of race adaptation. *Journal of Vision*, *11*, 567–567.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261–304.
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, *35*, 283–319.
- Cohen, D. J. (1997). Visual detection and perceptual independence: Assessing color and form. *Attention, Perception, & Psychophysics*, *59*, 623–635.
- DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*, 292–303.
- Demeyer, M., Zaenen, P., & Wagemans, J. (2007). Low-level correlations between object properties and viewpoint can cause viewpoint-dependent object recognition. *Spatial Vision*, *20*, 79–106.
- Eichenbaum, H., & Cohen, N. J. (2004). *From conditioning to conscious recollection: Memory systems of the brain* (No. 35). New York, NY: Oxford University Press.
- Emmerich, D. S., Gray, C. S., Watson, C. S., & Tanis, D. C. (1972). Response latency, confidence and ROCs in auditory signal detection. *Perception & Psychophysics*, *11*, 65–72.

- Ennis, D. M., & Ashby, F. G. (2003). Fitting decision bound models to identification or categorization data. Unpublished manuscript. Available at <http://www.psych.ucsb.edu/~ashby/cholesky.pdf>
- Farris, C., Viken, R. J., & Treat, T. A. (2010). Perceived association between diagnostic and non-diagnostic cues of women's sexual interest: General Recognition Theory predictors of risk for sexual coercion. *Journal of mathematical psychology*, *54*, 137–149.
- Giordano, B. L., Visell, Y., Yao, H. Y., Hayward, V., Cooperstock, J. R., & McAdams, S. (2012). Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions. *Journal of the Acoustical Society of America*, *131*, 4002–4012.
- Kadlec, H. (1999). MSDA_2: Updated version of software for multidimensional signal detection analyses. *Behavior Research Methods*, *31*, 384–385.
- Kadlec, H., & Townsend, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology*, *36*, 325–374.
- Kadlec, H., & Townsend, J. T. (1992b). Signal detection analyses of multidimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–231). Hillsdale, NJ: Erlbaum.
- Louw, S., Kappers, A. M., & Koenderink, J. J. (2002). Haptic discrimination of stimuli varying in amplitude and width. *Experimental brain research*, *146*, 32–37.
- Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–190). New York, NY: Wiley.
- Macmillan, N. A., & Creelman, D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*, 49–70.
- Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, *22*, 795–817.
- Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception & Psychophysics*, *60*, 620–637.
- Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*, 582–591.
- Maddox, W. T., Ashby, F. G., & Waldron, E. M. (2002). Multiple attention systems in perceptual categorization. *Memory & Cognition*, *30*, 325–339.
- Maddox, W. T., Glass, B. D., O'Brien, J. B., Filoteo, J. V., & Ashby, F. G. (2010). Category label and response location shifts in category learning. *Psychological Research*, *74*, 219–236.
- Marascuilo, L. (1970). Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrika*, *35*, 237–243.
- Menneer, T., Wenger, M., & Blaha, L. (2010). Inferential challenges for General Recognition Theory: Mean-shift integrality and perceptual configularity. *Journal of Vision*, *10*, 1211–1211.
- Murdock, B. B. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, *13*, 511–521.
- Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory, PGIT-4*, 171–212.
- Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model. *Psychological Review*, *111*, 588.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.
- Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.
- Silbert, N. H. (2012). Syllable structure and integration of voicing and manner of articulation information in labial consonant identification. *Journal of the Acoustical Society of America*, *131*, 4076–4086.
- Silbert, N. H., & Thomas, R. D. (2013). Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Bulletin & Review*, *20*(1), 1–20.
- Silbert, N. H., Townsend, J. T., & Lentz, J. J. (2009). Independence and separability in the perception of complex nonspeech sounds. *Attention, Perception, & Psychophysics*, *71*, 1900–1915.
- Smith, E. E. (1968). Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, *69*, 77–110.
- Smith, J. E. K. (1992). Alternative biased choice models. *Mathematical Social Sciences*, *23*, 199–219.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–168.
- Soto, F. A., Musgrave, R., Vucovich, L., & Ashby, F. G. (in press). General recognition theory with individual differences: A new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin & Review*.
- Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, *4*, 232–243.
- Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340.
- Tanner, W. P., Jr. (1956). Theory of recognition. *Journal of the Acoustical Society of America*, *30*, 922–928.
- Thomas, R. D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general d' . *Psychonomic Bulletin & Review*, *6*, 224–238.
- Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics*, *63*, 625–650.
- Townsend, J. T., Aisbett, J., Assadi, A., & Busemeyer, J. (2006). General recognition theory and methodology for dimensional independence on simple cognitive

- manifolds. In H. Colonius & E. N. Dzhafarov (Eds.), *Measurement and representation of sensations: Recent progress in psychophysical theory* (pp. 203–242). Mahwah, NJ: Erlbaum.
- Townsend, J. T., Houpt, J. W., & Silbert, N. H. (2012). General recognition theory extended to include response times: Predictions for a class of parallel systems. *Journal of Mathematical Psychology*, *56*, 476–494.
- Townsend, J. T., & Spencer-Smith, J. B. (2004). Two kinds of global perceptual separability and curvature. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 89–109). Mahwah, NJ: Erlbaum.
- Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*, 168–176.
- Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 872–892.
- Wickens, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology*, *36*, 213–234.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York, NY: Oxford University Press.