Journal of Mathematical Psychology **44**, 310–329 (2000)
doi:10.1006/jmps.1998.1249, available online at http://www.idealibrary.com on IDEAL®

A Stochastic Version of General Recognition Theory

F. Gregory Ashby

University of California, Santa Barbara

General recognition theory (GRT) is a multivariate generalization of signal detection theory. Past versions of GRT were static and lacked a process interpretation. This article presents a stochastic version of GRT that models moment-by-moment fluctuations in the output of perceptual channels via a multivariate diffusion process. A decision stage then computes a linear or quadratic function of the outputs from the perceptual channels, which drives a univariate diffusion process that determines the subject's response. Conditions are established under which the stochastic and static versions of GRT make identical accuracy predictions. These equivalence relations show that traditional estimates of perceptual noise may often be corrupted by decisional influences. © 2000 Academic Press

General recognition theory (GRT), which was first introduced by Ashby and Townsend (1986), is a multivariate generalization of signal detection theory (e.g., Green & Swets, 1966; Tanner & Swets, 1954). GRT has been used successfully to model perceptual and decisional processing in stimulus identification (Ashby & Lee, 1991, 1992), categorization (Ashby, 1992; Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992, 1993; Maddox & Ashby, 1993), similarity judgment (Ashby & Perrin, 1988; Perrin & Ashby, 1991), preference judgment (Perrin, 1992), same-difference judgment (Thomas, 1996), and speeded classification (Ashby & Maddox, 1994; Maddox & Ashby, 1996). When GRT is used to model categorization behavior, it is often called decision bound theory.

GRT assumes that the percept elicited by a stimulus on any single trial of a categorization or identification experiment can be represented as a point in a multidimensional perceptual space. Over trials, however, stimulus and perceptual noise are assumed to induce variability in the percept associated with every specific stimulus (e.g., Ashby & Lee, 1993; Green & Swets, 1966: Tanner, 1956; Tanner & Swets, 1954). As a result, the appropriate experiment-level representation of a stimulus is a

This research was supported by Grant SBR-9514427 from the National Science Foundation. I thank Wolfgang Schwarz, who contributed greatly to the ideas developed in this article. I am also grateful to J. D. Balakrishnan, W. K. Estes, Jean-Claude Falmagne, Todd Maddox, Roger Ratcliff, and Elliot Waldron for their helpful comments and suggestions.

Address correspondence and reprint requests to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106. E-mail: ashby@psych.ucsb.edu.

0022-2496/00 $35.00

multivariate probability distribution. The subject is assumed to divide the perceptual space into regions and to associate a response label with each region. On each trial, the subject notes which region the percept is in and then emits the associated response. Ashby, Alfonso-Reese, Turken, and Waldron (1998) proposed a neuropsychological version of the theory and described how the decision bounds might be learned.

Although GRT is quite general, in the sense that it can be used to account for data from a great variety of experiments, it is a static theory and does not make response time (RT) predictions. This article develops a stochastic version of GRT that makes RT predictions, as well as predictions about response accuracy. The dynamic version of GRT replaces the static point representation of the percept, assumed by classical GRT, with a multivariate stochastic diffusion process. Among other results, I will establish conditions under which the static and dynamic versions of GRT make identical accuracy predictions. This result is important because it shows that traditional estimates of perceptual noise (i.e., obtained either from GRT or from signal detection theory) may often be corrupted by decisional influences.

GENERAL RECOGNITION THEORY

As mentioned above, GRT is a multivariate generalization of signal detection theory that has been used successfully to account for data from a wide variety of different experimental tasks. This article focuses on categorization or identification tasks in which there are two response alternatives. Under these conditions, GRT assumes that on any single trial, the percept elicited by stimulus i can be represented by the vector $\mathbf{x}_i' = [x_{1i}, x_{2i}, ..., x_{ni}]$, where x_{ji} is the perceived value of stimulus i on dimension or component j. Because of stimulus and neural noise, $\mathbf{x}_i$ is assumed to be a random vector that varies across trials. In most applications of GRT, the distribution of $\mathbf{x}_i$ is assumed to be multivariate normal. Let $\boldsymbol{\mu}_{Si}$ denote the vector of mean percepts elicited by stimulus i, where the S indicates that this is the static or classic version of the model, and let Σ_{Si} denote the variance–covariance matrix of the stimulus i perceptual distribution.

GRT assumes that to select a response, the subject partitions the perceptual space into two (not necessarily contiguous) regions and associates a different response with each region. On each trial, the subject determines which region the percept is in and then emits the associated response. The decision bound is the set of all points that separate the two regions. In most applications of GRT, it is possible to define a function $y_i = h(\mathbf{x}_i)$ with the property that

$$y_i = h(\mathbf{x}_i) \begin{cases} > 0, & \text{for all } \mathbf{x}_i \text{ in one response region} \\ = 0, & \text{for all } \mathbf{x}_i \text{ on the decision bound} \\ < 0, & \text{for all } \mathbf{x}_i \text{ in the other response region.} \end{cases}$$

As a consequence, the decision rule: "Respond A to all points on one side of the bound and B to all points on the other side," is equivalent to the rule:

$$\text{Respond A if } y_i > 0, \text{ and respond B if } y_i < 0.$$

Because the function $y_i = h(\mathbf{x}_i)$ discriminates between stimuli on either side of the decision bound, it is often called a *discriminant function*. Thus, the probability of responding A on trials when stimulus i was presented equals

$$\mathrm{P}(A\,|\,i) = \mathrm{P}[\,y_i > 0\,|\,i\,].$$

For example, if the subject uses a linear decision bound, then there exists some vector of constants $\mathbf{b}_S$ and some scalar c_S such that

$$y_i = \mathbf{b}_S'\mathbf{x}_i + c_S + \varepsilon_S,$$

where ε_S is a random variable (independent of $\mathbf{x}_i$) with zero mean and variance σ_S^2 that represents criterial noise. Therefore, the mean of y_i is $\mathbf{b}_S'\boldsymbol{\mu}_{Si} + c_S$ and the variance is $\mathbf{b}_S'\,\Sigma_{Si}\,\mathbf{b}_S + \sigma_S^2$. So, if the distribution of $\mathbf{x}_i$ is multivariate normal and the distribution of ε_S is normal, then

$$P(A\,|\,i) = \Phi\left(\frac{-\mathbf{b}_S'\mu_{Si} - c_S}{\sqrt{\mathbf{b}_S'\Sigma_{Si}\mathbf{b}_S + \sigma_S^2}}\right),$$

where Φ is the cumulative standard normal (i.e., Z) distribution function (see, e.g., Ashby, 1982, for more details).

In the absence of criterial noise, it turns out that with linear bounds, $|y_i|$ is monotonic with the distance between the percept and the decision bound (e.g., Ashby & Maddox, 1992). We could, therefore, create a new variable D_i, defined as the signed distance from the percept to the decision bound—that is, for any point above the bound, D_i is distance-to-the-bound, but for any point below the bound,

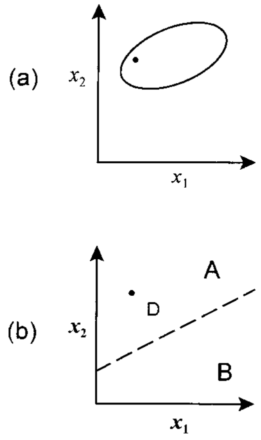

FIG. 1. The classic GRT model. Figure 1a shows the perceptual representation of stimulus i, which in this example is assumed to vary on two perceptual dimensions, x_1 and x_2. The ellipse is a contour of equal likelihood from the multivariate perceptual distribution that is the experiment-level perceptual representation, and the point is an example of the perceptual representation from a single trial. Figure 1b illustrates the decision process. See the text for more details.

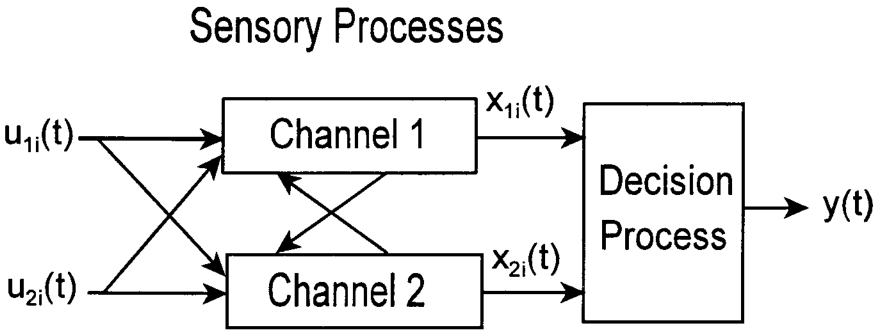

FIG. 2. A process interpretation of GRT for the special case in which there are only two perceptual dimensions (and two stimulus dimensions).

D_i is negative distance-to-the-bound. In this way, rather than basing the decision rule on y_i, an equivalent alternative is to use the rule:

$$\text{Respond A if } D_i > 0, \text{ and respond B if } D_i < 0.$$

This static version of GRT is illustrated in Fig. 1.

Ashby (1989) proposed the process interpretation of GRT described in Fig. 2 (illustrated for the special case in which the number of stimulus dimensions, n, equals two). In this interpretation, there is a different sensory channel tuned to each stimulus component or dimension. For example, when the stimuli are lines that vary in length and orientation, channel 1 would be a size (i.e., spatial frequency) sensitive channel, whereas channel 2 would be orientation sensitive. The crossing lines on the left allow for the possibility that channel 2, for example, might respond to the channel 1 component. Ashby (1989) showed that whether or not these lines cross is closely related to whether the stimulus components are perceived integrally (i.e., crossed lines) or separably (i.e., no crossing). The crossed lines in the middle of Fig. 2 allow for possible lateral interactions between the channels (i.e., lateral inhibition or excitation). Ashby showed that this more central type of perceptual interaction is closely associated with the phenomenon of perceptual dependence.

The output of the sensory channels is the percept $\mathbf{x}$, which serves as input to the decision process. The decision process is assumed to compute the discriminant value $y = h(\mathbf{x})$. For mathematical equivalence with GRT, the output of each sensory channel is a single numerical value on each trial, as is the output of the decision process. Although mathematically convenient, this assumption is biologically implausible. In real sensory channels, we expect a continuous output throughout the duration of the trial or at least for as long as the stimulus is displayed. In the stochastic version of GRT developed in the next section, this limitation of the Fig. 2 model is corrected.

STOCHASTIC GRT

A dynamic version of GRT can be created by making the more reasonable assumption that the output of the Fig. 2 sensory channels changes continuously throughout the course of a trial. Let $x_{ji}(t)$ denote the output of channel j at time t, on a trial when stimulus i is presented. Then the stimulus percept at time t is represented by the vector $\mathbf{x}_i(t)' = [x_{1i}(t), x_{2i}(t), ..., x_{ni}(t)]$. Thus, in the stochastic version of GRT, $\mathbf{x}_i(t)$ is a multidimensional stochastic process. In the particularly simple version considered here, it is specifically assumed that $\mathbf{x}_i(t)$ is a multivariate normal diffusion (i.e., Wiener) process (e.g., Cox & Miller, 1965)—that is, each $x_{ji}(t)$ is a standard Wiener process, and the pair $x_{ji}(t)$ and $x_{ki}(t)$ may be correlated for any values of j and k.

In a multivariate normal diffusion process, for any fixed value of t, $\mathbf{x}_i(t)$ has a multivariate normal distribution with mean vector $t\boldsymbol{\mu}_{Di}$ and variance–covariance matrix $t\Sigma_{Di}$. The vector $\boldsymbol{\mu}_{Di}$ and the matrix Σ_{Di} are called the infinitesimal mean vector and variance–covariance matrix, respectively. The subscript D, which stands for dynamic, distinguishes these structures from the mean vector and the variance–covariance matrix in the static version of GRT. From a process perspective, the assumption that the output of the channels can be modeled by a multivariate normal diffusion process is equivalent to assuming that the input to each channel is a pure Gaussian noise process and that each channel is a perfect integrator of its input (e.g., Cox & Miller, 1965). For example, the input on channel j (i.e., on trials when stimulus i is presented) is a random process $u_{ji}(t)$, which has the following properties: (i) the distribution of $u_{ji}(t)$ is identical for all values of t, (ii) $u_{ji}(t)$ and $u_{ji}(t+\Delta t)$ are independent for all values of t and Δt, and (iii) for any fixed value of t, $u_{ji}(t)$ is normally distributed with mean μ_{Dji} and variance σ^2_{Dji}. The inputs to all channels form a vector $\mathbf{u}_i(t)$ that, for any fixed value of t, has a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{Di}$ and the variance–covariance matrix Σ_{Di}. Thus, the infinitesimal parameters of the multivariate diffusion process can be interpreted as the mean vector and the variance–covariance matrix of the vector of inputs to the sensory channels. Since both the input and the output of each channel are normally distributed for any fixed value of t, and since the inputs and outputs at any two time points are independent, the infinitesimal parameters completely describe the multivariate diffusion process.

So far, we have replaced the static perceptual representation of GRT with a multidimensional stochastic process. The next question is: Given that the percept changes continuously over time, how can the subject select a response? It is important to realize that, despite the fluctuating percept, the decision bound of static GRT still provides important information about the correct response. When the percept is on one side of the decision bound, one response is favored, and when it crosses to the other side of the bound, the other response is favored. Thus, one plausible response rule is for the subject to monitor the discriminant value $y_i = h(\mathbf{x}_i)$ continuously (or perhaps, the signed distance to the bound D_i). Positive values of y_i still favor response A and negative values still favor response B. When the discriminant value is near zero, however, rather than choosing between two almost equally attractive responses, the subject can instead continue processing the stimulus, in hopes that one response will become clearly favored over the other.

In this model, the discriminant value changes randomly from moment to moment, so it too becomes a stochastic process, denoted now by $y_i(t) = h[\mathbf{x}_i(t)]$. Note, however, that $y_i(t)$ is univariate, no matter what the dimensionality of $\mathbf{x}_i(t)$.

Since positive values of $y_i(t)$ favor response A, and negative values favor response B, a plausible strategy would be to set thresholds at numerical values A and -B and use the decision rule:

$$\text{if} \quad y_i(t) = h[\mathbf{x}_i(t)] \begin{cases} > A, & \text{then respond A} \\ < -B, & \text{then respond B} \\ \text{otherwise, continue sampling.} \end{cases}$$

Of course, this defines the classic random walk or diffusion process with two absorbing barriers that has been used to account for two-choice response time data in psychology for more than 35 years (e.g., Laming, 1968; Link & Heath, 1975; Luce, 1986; Ratcliff, 1978; Ratcliff, Van Zandt, & McKoon, 1999; Stone, 1960; Townsend & Ashby, 1983).

The distribution of $y_i(t)$ depends on the nature of the decision bound (i.e.. discriminant function) used by the decision system. From a conceptual or computational perspective, it makes little difference what type of bound is assumed. Even so, the model is analytically tractable only if the decision bound is linear. Fortunately, linear bounds are known to provide good accounts of data when the optimal bound (i.e., the bound that maximizes accuracy) is linear (Ashby & Gott, 1988; Ashby & Maddox, 1990; Maddox & Ashby, 1993). However, it is also known that when the optimal bound is more complex than linear (e.g., quadratic), linear bounds provide poor fits to the resulting data (Ashby & Maddox. 1992; Maddox & Ashby, 1993). Thus, applications of the present model to these more complex category structures will likely require numerical methods to derive predictions.

The remainder of this article considers experimental designs for which subjects are likely to use linear decision bounds. With a linear decision bound, the decision system computes the discriminant process

$$y_i(t) = \mathbf{b}'_D \mathbf{x}_i(t) + c_D + \varepsilon_D(t),$$

where $\mathbf{b}_D$ is a vector of constants, c_D is a constant, and $\varepsilon_D(t)$ is a Wiener process with zero mean and infinitesimal variance σ_D^2 that represents criterial noise. Under these conditions, $y_i(t)$ is itself a Wiener process. For any fixed value of t, $y_i(t)$ is normally distributed with mean $(\mathbf{b}'_D \boldsymbol{\mu}_{Di} + c_D)\, t$ and variance $(\mathbf{b}'_D \Sigma_{Di} \mathbf{b}_D + \sigma_D^2)\, t$.

Figure 3 illustrates this stochastic generalization of GRT. In the sensory process, the point representation of the percept assumed by GRT is replaced by a multivariate diffusion process (i.e., compare Fig. 3 with Fig. 1). The percept is now assumed to change from moment to moment as more sensory information is collected. This causes the discriminant value $y_i(t)$, or alternatively the signed distance $D_i(t)$, also to change from moment to moment. The decision process is assumed to cumulate these values over time. A response is emitted when the cumulated discriminant value first crosses an absorbing barrier, at either A or -B.

It is important to distinguish between the decision bound, which partitions the perceptual space into regions that favor the two competing responses, and the absorbing barriers of the decisional diffusion process, which set response thresholds

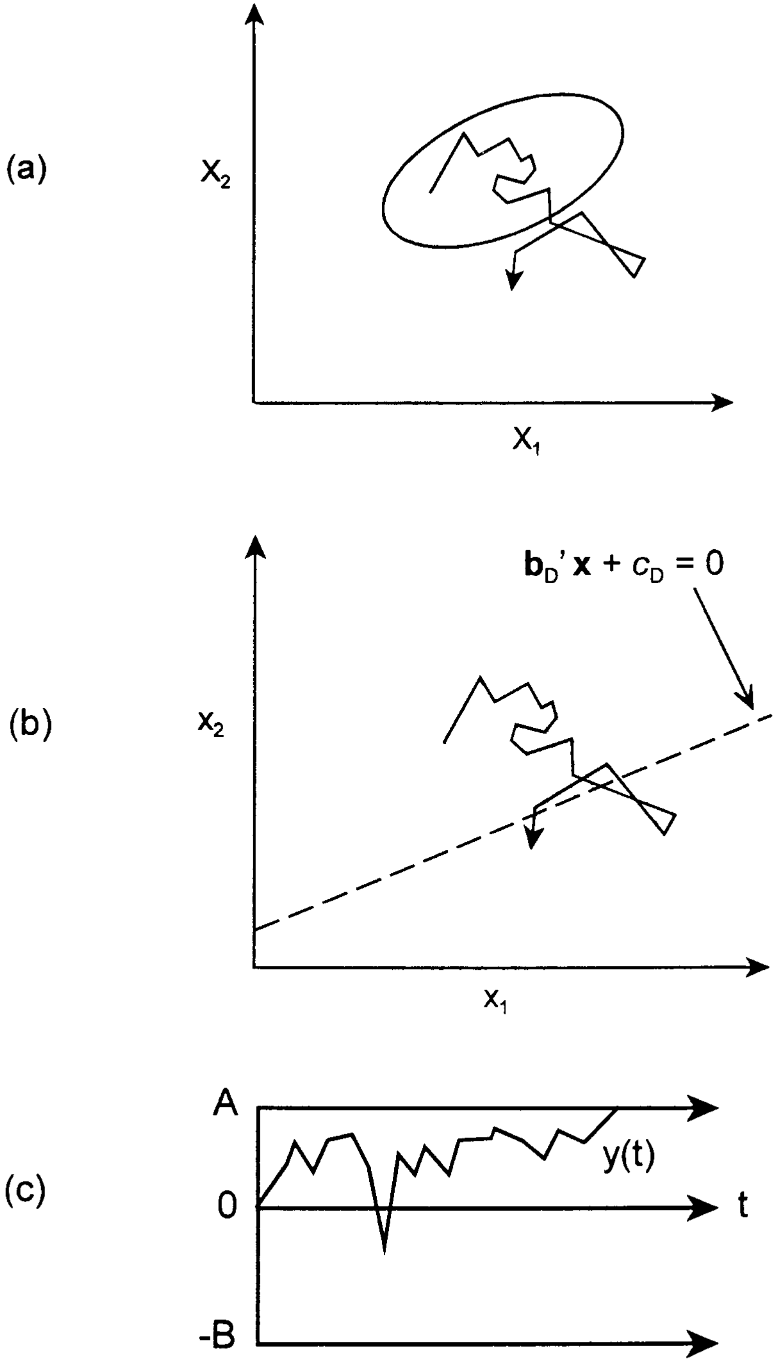

FIG. 3. The stochastic version of GRT. Figure 3a shows the perceptual representation of stimulus *i*. The ellipse is a contour of equal likelihood from the multivariate perceptual distribution that is the experiment-level infinitesimal perceptual representation, and the sample path is an example of the perceptual representation from a single trial. The decision process is illustrated in Figs. 3b and 3c. See the text for more details.

on the decision variable (e.g., on the discriminant value $y_i(t)$). The notion of a decision bound, although not a mathematical necessity to the developments of this article, is conceptually important because it ties the present stochastic GRT model to the rich and extensive GRT literature, in which the decision bound is a fundamental construct. To avoid confusion, I will be careful to refer to the set of all $\mathbf{x}_i$ for which $y_i = h(\mathbf{x}_i) = 0$ as the decision bound and the response thresholds, A and -B, as the absorbing barriers.

Analytic predictions for this stochastic GRT model are easy to derive from the predictions of the standard diffusion process with two absorbing barriers, which are readily obtained from a number of classic sources (e.g., Cox & Miller, 1965; Ratcliff, 1978; Townsend & Ashby, 1983). Let $T_{\mathbf{A}}$ and $T_{\mathbf{B}}$ denote the decision times on trials when the subject selects response A and B, respectively. Then $E(T_{\mathrm{A}} \mid i)$ and $f_{\mathrm{A}}(t \mid i)$ are the expected decision time and the probability density function of decision time on trials when stimulus i is presented and response A is given. Now consider a standard Wiener process with absorbing barriers at A and -B. Denote the drift on trials when stimulus i is presented by μ_i and the variance by σ_i^2. Then it is well known that the predicted response probabilities are given by

$$P(\mathrm{A} \mid i) = \frac{1 - \exp\left[-\dfrac{2\mu_i B}{\sigma_i^2}\right]}{1 - \exp\left[-\dfrac{2\mu_i (A+B)}{\sigma_i^2}\right]}, \; P(\mathrm{B} \mid i) = 1 - P(\mathrm{A} \mid i), \tag{1}$$

and the expected decision times equal

$$E(T_A \mid i) = \frac{1}{\mu_i}\left\{(A+B)\frac{1 + \exp\left[-\dfrac{2\mu_i(A+B)}{\sigma_i^2}\right]}{1 - \exp\left[-\dfrac{2\mu_i(A+B)}{\sigma_i^2}\right]} - B\frac{1 + \exp\left[-\dfrac{2\mu_i B}{\sigma_i^2}\right]}{1 - \exp\left[-\dfrac{2\mu_i B}{\sigma_i^2}\right]}\right\},$$

and

$$E(T_B \mid i) = \frac{1}{\mu_i}\left\{(A+B)\frac{1 + \exp\left[-\dfrac{2\mu_i(A+B)}{\sigma_i^2}\right]}{1 - \exp\left[-\dfrac{2\mu_i(A+B)}{\sigma_i^2}\right]} - A\frac{1 + \exp\left[-\dfrac{2\mu_i A}{\sigma_i^2}\right]}{1 - \exp\left[-\dfrac{2\mu_i A}{\sigma_i^2}\right]}\right\}.$$

Finally, the decision time probability density functions are given by (e.g., Ratcliff, 1978)

$$\begin{aligned} f_A(t \mid i) = {} & \frac{1}{P(A \mid i)} \frac{\sigma_i^2 \pi}{(A+B)^2} \exp\left[\frac{\mu_i A}{\sigma_i^2}\right] \sum_{n=1}^{\infty} n \sin\left(\frac{nA\pi}{A+B}\right) \\ & \times \exp\left\{-\frac{1}{2}\left[\left(\frac{\mu_i}{\sigma_i}\right)^2 + \left(\frac{n\sigma_i \pi}{A+B}\right)^2\right] t\right\}, \end{aligned}$$

and

$$f_B(t\,|\,i) = \frac{1}{P(B\mid i)}\frac{\sigma_i^2\pi}{(A+B)^2}\exp\left[-\frac{\mu_i B}{\sigma_i^2}\right]\sum_{n=1}^{\infty} n\sin\left(\frac{nB\pi}{A+B}\right)$$
$$\times\exp\left\{-\frac{1}{2}\left[\left(\frac{\mu_i}{\sigma_i}\right)^2+\left(\frac{n\sigma_i\pi}{A+B}\right)^2\right]t\right\}.$$

From these equations, predictions are readily derived for the stochastic GRT model by noting that

$$\mu_i = \mathbf{b}'_D\boldsymbol{\mu}_{Di} + c_D,$$

and

$$\sigma_i^2 = \mathbf{b}'_D\Sigma_{Di}\mathbf{b}_D + \sigma_D^2.$$

These latency predictions are for categorization (or identification) time. Observable RT also typically includes time for motor processes and perhaps also time for stimulus encoding. Therefore, to derive predictions for the expected value of observable RT, for example, one would typically use equations such as

$$E(RT_A\,|\,i) = E(T_A\,|\,i) + T_r$$

and

$$E(RT_B\,|\,i) = E(T_B\,|\,i) + T_r,$$

where T_r represents mean residual time.

EQUIVALENCE RELATIONS BETWEEN THE STATIC AND DYNAMIC VERSIONS OF GRT

Conceptually, the stochastic version of GRT is strikingly similar to the static version. For example, in both models, the parameters of the perceptual representation are mean vectors and variance–covariance matrices, and both models postulate a decision bound. Because of these similarities, it is natural to ask whether there are any conditions under which the models make identical response accuracy predictions. Also, what exactly is the relation between the perceptual parameters in the two models? To begin answering these questions, note that the stochastic model has two more free parameters than the static model—namely, the distance to the two absorbing barriers, A and B. If $A \neq B$, then the dynamic model will exhibit a response bias of a type that cannot be represented in the static model. Therefore, it is clear that a necessary condition for equivalence is that $A = B$ in the dynamic model. A complete set of sufficient conditions are given in Theorem 1.

THEOREM 1. *Consider a task in which the subject uses a linear decision bound, and suppose there is no bias in the placement of the response barriers of the dynamic* *model (i. e., so that $B = A$) and that noise in the static model has a logistic distribution (rather than the usual assumption of normality). Then the static and dynamic models predict identical response probabilities if the following three conditions hold.*

(1) *The two models have identical decision bounds.*

(2) *The mean perceptual vector for each stimulus in the static model is equal to the analogous infinitesimal mean vector in the dynamic model; that is, for all i*

$$\boldsymbol{\mu}_{Si} = \boldsymbol{\mu}_{Di}.$$

(3) *The variance–covariance matrices that describe the stimulus and perceptual noise in the two models and the criterial noise variances are related as*

$$\Sigma_{Si} = \frac{\pi^2}{12(A^*)^2} \Sigma_{Di} \qquad \textit{and} \qquad \sigma_S^2 = \frac{\pi^2}{12(A^*)^2} \sigma_D^2,$$

where A^* is the standardized distance to the absorbing barriers in the dynamic model—that is,

$$A^* = \frac{A}{\sqrt{\mathbf{b}'_D \Sigma_{Di} \mathbf{b}_D + \sigma_D^2}}.$$

Proof. All proofs are given in the Appendix.

This result shows that the interpretation of the parameters of the static and dynamic versions of GRT is strikingly similar. In both models, the decision bounds have exactly the same interpretation, and the mean percept in the static model is identical to the mean input to the sensory channels in the dynamic model. The only difference between the models is in the interpretation of the noise variance parameters. The perceptual and criterial noise variance parameters in the two models are not equal. Instead, they are proportional, and in both cases the constant of proportionality is inversely related to $(A^*)^2$, the squared standardized distance to the absorbing barriers in the dynamic model. Since the distance to the barriers in the dynamic model is under the control of the decision process, Theorem 1 says that estimates of perceptual or criterial noise obtained from classic GRT, or from signal detection theory (which is a special case of GRT), may often be corrupted by decisional influences. Thus, Theorem 1 formalizes a long-held belief in signal detection theory (e.g., Green & Swets, 1966; Pachella, 1974).

To better understand Theorem 1, consider a categorization or identification experiment in which there are several different speed–accuracy conditions. For example, suppose that in one condition, the experimenter emphasizes accuracy over speed, and in another condition subjects are encouraged to respond as quickly as possible, even if this extra speed costs them a few errors. Naturally, we expect accuracy to be higher in the accuracy condition than in the speed condition. Therefore, if a static GRT or signal detection model was fit to the data, the noise variance estimates would be larger in the speed condition. But does it really make sense to argue, for example, that the variability in the percept is affected by speed–accuracy instructions? After all, the stimulus information is identical in the two conditions. According to the dynamic version of GRT, the subject world respond to an increased emphasis on speed by decreasing the distance to the absorbing barriers—that is, by decreasing A. The noise variance parameters need not change at all. In these circumstances, Theorem 1 tells us that the perceptual noise variance estimates obtained from the static model will increase with speed emphasis. In summary, Theorem 1 is an important result, because it tells us that estimates of perceptual (or criterial) noise obtained from classical static versions of signal detection theory or GRT may often be corrupted by decisional influences. In fact, stochastic GRT can be used to predict exactly how such estimates will change as a function of speed stress.

Another important consequence of Theorem 1 follows from the fact that equality holds between the static and the dynamic versions of GRT only if the absorbing barriers in the dynamic model are equally distant from the origin (i.e., only if $A = B$). In signal detection theory, a response bias occurs if the response criterion (i.e., X_c) is set at any point for which the likelihood ratio does not equal one (i.e., $\beta \neq 1$). The analogue of the response criterion in classical GRT is the intercept of the decision bound, so in classical GRT a response bias occurs if the intercept of the decision bound is set to a value for which the likelihood ratio is unequal to one at all points on the decision bound. For example, in Fig. 1, a bias toward response B is created by increasing the intercept of the decision bound. In the dynamic version of GRT, this same response bias mechanism is available, since Theorem 1 indicates that the notion of a decision bound is identical in the static and dynamic theories. However, in the dynamic theory, a second mechanism is available for modeling response bias. If the distance to the two absorbing barriers is unequal (i.e., if $A \neq B$), then there will be a bias in favor of the response associated with the nearer barrier.

These two separate sources of bias are also explicitly modeled in Ratcliff's (1978) diffusion model, and Ratcliff (1985) showed that they are identifiable, in the sense that separately manipulating these two biases leads to differential RT and accuracy predictions. It is also straightforward to show that changing the intercept of the decision bound has a qualitatively different effect than changing the distance to the two absorbing barriers. To see this, consider a stochastic GRT model in which the subject uses the unbiased intercept for the decision bound. Then at time t, the discriminant value computed by the decision system equals

$$y_{i,\,\text{unbiased}}(t) = \mathbf{b}'_D \mathbf{x}_i(t) + c_{\text{unbiased}} + \varepsilon_D(t),$$

where c_{unbiased} is the unbiased intercept. Now suppose we increase the intercept by some fixed amount Δc. In this case, the new discriminant value at time t equals

$$\begin{aligned} y_i(t) &= \mathbf{b}'_D \mathbf{x}_i(t) + c_{\text{unbiased}} + \Delta c + \varepsilon_D(t) \\ &= y_{i,\,\text{unbiased}}(t) + \Delta c. \end{aligned}$$

In other words, introducing a response bias into the intercept of the decision bound is equivalent to adding a constant to each discriminant value or, in other words, to

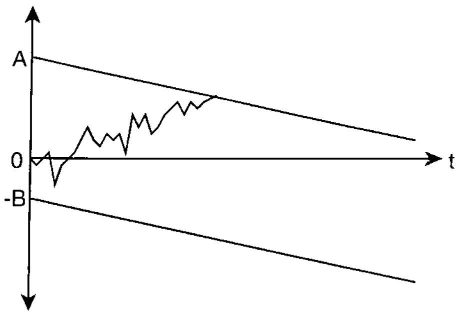

FIG. 4. The biased random walk model introduced by Ashby (1983).

the decisional diffusion process. Ashby (1983) showed that adding such a constant is in turn equivalent to rotating the absorbing barriers until they have a slope of $-\Delta c$ (illustrated in Fig. 4). Clearly, such a rotation is not equivalent to setting the barriers an unequal distance from the origin (i.e., to setting $A \neq B$). Balakrishnan (1998a, 1998b) developed a new nonparametric measure of response bias, which he applied to several data sets in which standard signal detection analysis had indicated a significant response bias. On the basis of these analyses, Balakrishnan (1998a, 1998b) effectively argued that when a response bias occurs, it is much more likely that $A \neq B$ than that the decision bound intercept is biased.

RT PREDICTIONS

This section investigates the RT predictions of the stochastic version of GRT. The first question to ask is how the RT predictions vary with location of the stimulus in the perceptual space. Theorem 2 answers this question.

Theorem 2. *Consider a categorization task in which the subject uses a linear decision bound. Then drift rate in the stochastic version of GRT increases with the distance between the infinitesimal mean percept and the decision bound. If the infinitesimal perceptual distributions of all stimuli have identical variance–covariance matrices (i.e.,* $\Sigma_{Di} = \Sigma_D$*, for all i), then the diffusion variance is identical for all stimuli.*

Thus, mean drift increases with the distance between the mean percept and the decision bound. The assumption that RT decreases with the distance between the mean percept and the response criterion has a long history in signal detection theory (e.g.. Murdock, 1985; Thomas, 1971; see also, Pike, 1973). When incorporated into GRT, Ashby and Maddox (1994) called the assumption the RT–distance hypothesis, and they argued that this simple assumption accounts for most of the variance in RT data from categorization and two alternative identification studies. More recent direct tests have yielded impressive empirical support for the RT–distance hypothesis (Ashby, Boynton, & Lee, 1994: Maddox, Ashby, & Gottlob, 1998). Theorem 2 indicates that the stochastic GRT model satisfies the RT–distance hypothesis if all infinitesimal perceptual distributions have identical variance–covariance matrices. However, it is straightforward to show that if at least some infinitesimal perceptual distributions have unequal variance–covariance matrices, then the stochastic GRT model sometimes predicts violations of the RT–distance hypothesis.

Maddox and Ashby (1996) incorporated the RT–distance hypothesis into the static GRT model by assuming that decision time decreases with the distance between the percept and the decision bound. Different versions of this RT–distance GRT model can be formulated by making different assumptions about the specific function that relates distance and decision time. Without the rich perceptual representation supplied by GRT, the RT–distance hypothesis predicts no difference between error and correct RTs, and it is unable to make any predictions about the form of the RT distributions. In contrast, the RT–distance GRT models make specific predictions about RT distributions, and they generally predict that error RTs will be greater than correct RTs, with this difference decreasing for stimuli further from the bound (Ashby & Maddox, 1994). This latter prediction has generally been supported in categorization RT studies, although frequently errors are faster than corrects for stimuli where the categorization judgment is particularly easy (e.g., Maddox *et al.*, 1998). The relation between correct and incorrect categorization RTs parallels the relation between correct and incorrect RTs in more general two-choice RT tasks. In particular, in most two-choice RT studies, errors are slower than correct responses when the discrimination is difficult, but faster than corrects when the discrimination is easy (for reviews, see, e.g., Link & Heath, 1975; Luce, 1986). In addition, there is a long history of using the relative speed of correct versus incorrect responses to test sequential sampling models (e.g., Laming, 1968: Pike, Dalgleish, & Wright, 1977; Vickers, Caudrey, & Willson, 1971).

Maddox and Ashby (1996) tested the ability of several specific RT–distance GRT models to account for the RT distributions from a speeded classification study. Two specific versions were tested. In one, RT was assumed to decrease as an exponential function of distance-to-bound. and in the other the decrease was assumed to be a power function of distance-to-bound. Although both models provided reasonable accounts of the RT distributions, they also both showed systematic deviations, and so neither can be considered an adequate account of human categorization RT data. Of the two, the exponential model provided the better fits, but it systematically underpredicted the tails of the RT distributions. In other words, the data contained more long RTs than predicted by the exponential RT–distance model. Long RTs are most likely to occur to percepts near the decision bound. According to the exponential RT–distance model, categorization time T is given by $T = \alpha \exp(-\beta D)$, where α and β are nonnegative constants and D is distance-to-bound. Thus, when a percept falls exactly on the bound, $D = 0$, and the exponential RT–distance model predicts that $T = \alpha$. The largest RTs observed by Maddox and Ashby (1996) were considerably greater than the value of α that provided the best fits to the entire RT distributions.

Figure 5 shows mean categorization time as predicted by a stochastic GRT model and by the most similar exponential RT–distance model. Note that the stochastic GRT model predicts much longer RTs than the exponential RT–distance model for percepts close to the decision bound. As such, the stochastic GRT model

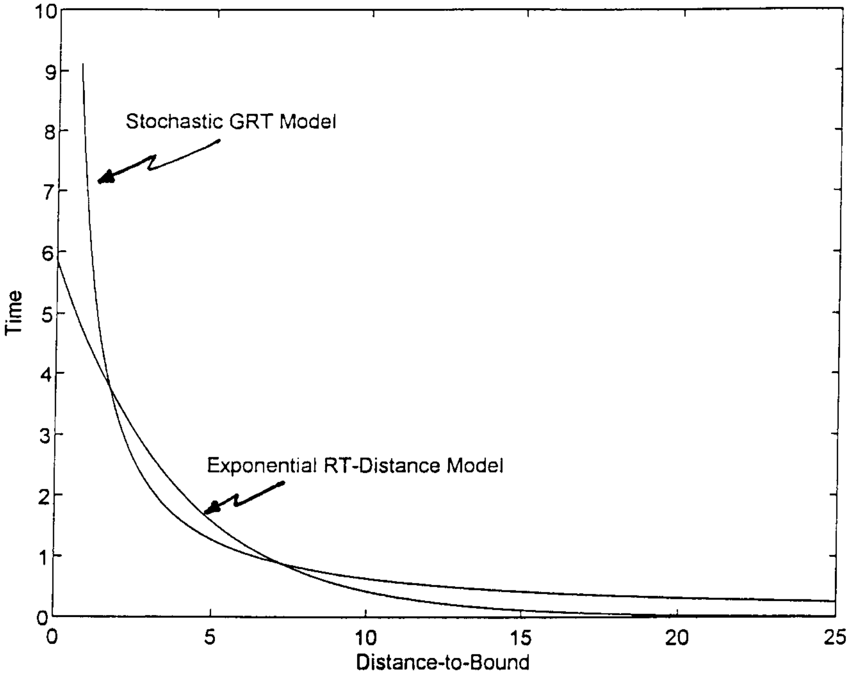

FIG. 5. Mean categorization time versus distance-to-bound, as predicted by the stochastic version of GRT and by the exponential RT–distance model.

should provide a better account of the tails of categorization RT density functions than the exponential model.

Once a drift rate is specified, the stochastic version of GRT is similar to Ratcliff's (1978) univariate diffusion model (see also, Ratcliff, 1981, 1985; Ratcliff & Rouder, 1998; Ratcliff *et al.*, 1999). The only important difference is that Ratcliff and his colleagues have argued persuasively for trial-by-trial variability in (mean) drift rate and in the starting point of the diffusion process.[1] As a result, in current versions of his model, both of these parameters are assumed to be random variables. This generally adds two free parameters to the model—a starting point variance and the variance in mean drift.[2] Once these two parameters are added to the stochastic GRT model, it then makes identical RT predictions as the Ratcliff *et al.* (1999) model. This includes predictions about mean RT, RT variance, the RT density function, the RT hazard function, and the relation between mean RT on trials when the subject responds correctly or incorrectly. In effect, the stochastic version of GRT provides a front end for Ratcliff's diffusion model. The multivariate diffusion process provides a detailed model of the perceptual representation, and the decision bound (i.e., the cumulated discriminant process) provides an alternative mechanism for converting the percept into a drift rate.

[1] Laming (1968) was the first to postulate variability in the starting point, and he showed that the general effect of such variability was to increase the speed of errors relative to correct responses.

[2] This latter parameter is not to be confused with the variance of the diffusion process. The variance in mean drift specifies trial-by-trial variability in mean drift, whereas the variance of the diffusion process specifies within trial variability in drift.

Maddox *et al.* (1998) reported the results of a sophisticated statistical investigation of the RTs from several standard categorization experiments. When distance-to-bound was used as an (inverse) measure of task difficulty, the RTs from these experiments looked very similar to RTs from other two-choice RT tasks (e.g., memory scanning, subitizing, YES–NO detection). For example, plots of the RT hazard functions for different values of distance-to-bound were qualitatively identical to plots of memory scanning RT hazard functions for different memory set sizes (reported by Ashby. Tein, & Balakrishnan, 1993). In both cases, the hazard functions were ordered by task difficulty, they were nondecreasing on difficult trials (i.e., stimuli near the bound and large memory set sizes), and they were increasing-then-decreasing on easy trials (i.e., stimuli far from the bound and small memory set sizes). Ratcliff and his colleagues (Ratcliff & Rouder, 1998; Ratcliff *et al.*, 1999) have shown that the univariate diffusion model provides impressive fits to a wide variety of two-choice RT data, so it is plausible that the stochastic version of GRT will also provide good fits to categorization RT data.

As a preliminary test of this assumption, Ashby and Schwarz (1996) fit the stochastic GRT model to data from two experiments reported by Maddox *et al.* (1998). The stimuli in both experiments were circles of varying diameter containing a radial line of varying orientation. In one experiment. the bound separating the two categories was unidimensional (so the only relevant dimension was diameter), and in the other experiment, it was diagonal (with a slope of -1). For each subject, the model was fit simultaneously to the mean correct RTs and the proportion corrects. The stochastic GRT model provided good fits to the data from both experiments. Perhaps most impressive, however, was that the parameter values estimated during the model fitting procedure provided good predictions of the mean RTs from trials when the subject responded incorrectly.[3] The only consistent mispredictions occurred for mean error RT to stimuli that were far from the decision bound. In both experiments, these observed mean RTs were consistently faster than predicted by the model (although the model correctly predicted that these errors were faster than correct responses to the same stimuli). Ratcliff *et al.* (1999) pointed out that fast errors on easy trials are empirically common, and they showed that their diffusion model can account for fast errors by allowing trial-by-trial variability in the starting location of the diffusion process and in the mean drift rate. Thus, it appears likely that the ability of the stochastic GRT model to account for the fast errors on easy trials observed by Maddox *et al.* (1998) could be improved by allowing trial-by-trial variability in the starting location of the diffusion

[3] The stochastic GRT model had six free parameters in these applications: the infinitesimal perceptual variance (i.e., the infinitesimal variance–covariance matrices were all assumed to equal $\Sigma_{Di} = \sigma^2 I$, where I is the identity matrix), the slope and intercept of the decision bound, the distance to the two absorbing barriers (i.e., A and B), and the mean residual time. Parameter estimates were obtained via a minimum chi-squared procedure. In the unidimensional experiment, the mean correlation (across four subjects) between the observed and the predicted values was 0.933 for proportion correct and 0.931 for mean correct RT. In the diagonal experiments, these correlations were 0.818 and 0.933, respectively. In the unidimensional experiment, the correlation between observed and predicted mean error RT was 0.728, and in the diagonal experiment, it was 0.731 (in the case of mean error RT, these are true predicted values, since the error RTs were not used in the parameter estimation process).

process and in the mean drift rate. Although this initial application was promising, it is difficult to draw stronger conclusions. For example, it is possible that other categorization RT models might also have successfully fit these data (e.g. Nosofsky & Palmeri's, 1997, EBRW model). Certainly, much more work needs to be done to evaluate the empirical validity of the stochastic GRT model developed here.

RELATION TO OTHER VERSIONS OF GRT

The theory developed here generalizes the discrete-time model proposed by Ashby (1989). The perceptual representation in the two models is essentially the same, although the continuous-time multivariate diffusion process proposed here makes the current model more analytically tractable than the version developed by Ashby (1989). In addition, the decision process of the current model is more general and more natural than the decision process proposed by Ashby (1989).

Ashby and Maddox derived RT predictions from GRT by incorporating the so-called RT–distance hypothesis into the classical version of the theory (Ashby & Maddox, 1991, 1994; Maddox & Ashby, 1996; see also, Ashby *et al.*, 1994). As discussed above, the stochastic version of GRT developed in this article is closely related to the RT–distance GRT models because under appropriate distributional assumptions, the stochastic GRT model satisfies the RT–distance hypothesis (i.e.. see Theorem 2). On the other hand, the model developed here has a number of attractive advantages over the RT–distance GRT models. Most importantly, unlike the RT–distance models, the stochastic version of GRT is a process model, since it makes explicit assumptions about underlying psychological processes. Second, the stochastic version of GRT is more general than the RT–distance models, since the stochastic model is guaranteed to satisfy the RT–distance hypothesis only under restrictive distributional assumptions (i.e., $\Sigma_{Di} = \Sigma_D$, for all i).

Theorem 1 shows that for any particular static GRT model, there exists a dynamic GRT model (with $A = B$) that makes identical (or nearly identical) response accuracy predictions. Since the static GRT model has been successful at accounting for response accuracy data across a wide variety of identification (Ashby & Lee, 1990; Ashby & Perrin, 1988) and categorization (Ashby & Lee. 1990; Maddox & Ashby, 1993) experiments, it is clear that the stochastic version of GRT would be equally successful. Of course, a major difference between the static and the dynamic versions of GRT is that the dynamic version also makes RT predictions, whereas the static version does not.

CONCLUSIONS

In this article, I proposed a stochastic generalization of GRT that replaces the static point representation of the percept of classical GRT with a multivariate diffusion process. The decision process is assumed to cumulate distance-to-bound, and this variable drives a univariate diffusion process with two absorbing barriers. The stochastic version of GRT has a number of attractive advantages over the original static version. First, it makes predictions about response time as well as about response accuracy. Second, it provides a more realistic model of the perceptual representation than the static version of GRT. In particular, the multivariate diffusion process representation is closely associated with the notion of interacting, dynamic sensory channels. The model also provides a structure that makes it easy to replace this part of the model with alternative versions of the sensory channels that are even more biologically plausible (e.g., by replacing the multivariate diffusion process with a multivariate Ornstein–Uhlenbeck process). Third, the stochastic version of GRT can be viewed as enriching the perceptual representation of Ratcliff's (1978) diffusion model, which, arguably, has been the most successful of all two-choice response time models. Finally, the equivalence mapping between the stochastic and static versions of GRT indicate that traditional estimates of perceptual noise may often be corrupted by decisional influences.

APPENDIX

Proof of Theorem 1. From Eq. (1), we see that if $B = A$, then

$$P(A \mid i) = \frac{\exp\left(\frac{4\mu_i A}{\sigma_i^2}\right)\left[1 - \exp\left(-\frac{2\mu_i A}{\sigma_i^2}\right)\right]}{\exp\left(\frac{4\mu_i A}{\sigma_i^2}\right) - 1}$$

$$= \frac{\exp\left(\frac{2\mu_i A}{\sigma_i^2}\right)\left[\exp\left(\frac{2\mu_i A}{\sigma_i^2}\right) - 1\right]}{\left[\exp\left(\frac{2\mu_i A}{\sigma_i^2}\right) + 1\right]\left[\exp\left(\frac{2\mu_i A}{\sigma_i^2}\right) - 1\right]}$$

$$= \frac{1}{1 + \exp\left(-\frac{2\mu_i A}{\sigma_i^2}\right)}.$$

In the stochastic GRT model, $\mu_i = \mathbf{b}'_D \boldsymbol{\mu}_{Di} + c_D$ and $\sigma_i^2 = \mathbf{b}'_D \Sigma_{Di} \mathbf{b}_D + \sigma_D^2$. Therefore,

$$P(A \mid i) = \frac{1}{1 + \exp\left[-\frac{2A(\mathbf{b}'_D \boldsymbol{\mu}_{Di} + c_D)}{(\mathbf{b}'_D \Sigma_{Di} \mathbf{b}_D + \sigma_D^2)}\right]}.$$

In the static model, if the noise is logistic, then

$$P(A \mid i) = \frac{1}{1 + \exp\left[-\frac{\pi}{\sqrt{3}} \frac{\mathbf{b}'_S \boldsymbol{\mu}_{Si} + c_S}{\sqrt{\mathbf{b}'_S \Sigma_{Si} \mathbf{b}_S + \sigma_S^2}}\right]}.$$

Therefore, the two models predict identical response probabilities if and only if

$$\frac{\pi}{\sqrt{3}}\frac{\mathbf{b}'_S\boldsymbol{\mu}_{Si}+c_S}{\sqrt{\mathbf{b}'_S\Sigma_{Si}\mathbf{b}_S+\sigma_S^2}}=\frac{2A(\mathbf{b}'_D\boldsymbol{\mu}_{Di}+c_D)}{\mathbf{b}'_D\Sigma_{Di}\mathbf{b}_D+\sigma_D^2}.$$

Next, define A^* as the standardized distance to either absorbing barrier in the dynamic model—that is,

$$A^*=\frac{A}{\sqrt{\mathbf{b}'_D\Sigma_{Di}\mathbf{b}_D+\sigma_D^2}}.$$

In terms of A^*, equality holds if and only if

$$\frac{\pi}{\sqrt{3}}\frac{\mathbf{b}'_S\boldsymbol{\mu}_{Si}+c_S}{\sqrt{\mathbf{b}'_S\Sigma_{Si}\mathbf{b}_S+\sigma_S^2}}=\frac{2A^*(\mathbf{b}'_D\boldsymbol{\mu}_{Di}+c_D)}{\sqrt{\mathbf{b}'_D\Sigma_{Di}\mathbf{b}_D+\sigma_D^2}}.$$

The theorem follows immediately from this result.

Proof of Theorem 2. Suppose the upper absorbing barrier is associated with category A and the lower barrier with category B. With a linear decision bound, the drift at time t is $y_i(t)=\mathbf{b}'_D\mathbf{x}_i(t)+c_D+\varepsilon_D(t)$, and the mean drift is $(\mathbf{b}'_D\boldsymbol{\mu}_{Di}+c_D)\,t$. Therefore, the set of all stimuli that elicit a mean drift of Ct at time t satisfies

$$C=\mathbf{b}'_D\boldsymbol{\mu}_{Di}+c_D$$

or equivalently

$$\mathbf{b}'_D\boldsymbol{\mu}_{Di}+(c_D-C)=0.$$

Now, the decision bound is defined as the set $\{\mathbf{x}(t)\,|\,\mathbf{b}'_D\mathbf{x}(t)+c_D=0\}$. Therefore, the infinitesimal mean percepts of all stimuli that elicit the same mean drift of Ct at time t fall on a line that is parallel to the decision bound. As such, the set of all such mean percepts are equally distant from the decision bound.

Next, let $C_1>C_2>0$, and let $\boldsymbol{\mu}_1\in\{\boldsymbol{\mu}\,|\,\mathbf{b}'_D\boldsymbol{\mu}+c_D=C_1\}$ and $\boldsymbol{\mu}_2\in\{\boldsymbol{\mu}\,|\,\mathbf{b}'_D\boldsymbol{\mu}+c_D=C_2\}$. Thus, $\boldsymbol{\mu}_1$ elicits a larger mean drift than $\boldsymbol{\mu}_2$. Now $\boldsymbol{\mu}_1$ falls on a line with intercept c_D-C_1, and $\boldsymbol{\mu}_2$ falls on a line with intercept c_D-C_2. Since $C_1>C_2$, $\boldsymbol{\mu}_1$ is farther from the decision bound than $\boldsymbol{\mu}_2$. A similar argument shows that, of the percepts eliciting a negative drift, the faster drift (toward the barrier at -B) is elicited by the stimulus with mean infinitesimal percept farther from the decision bound.

The drift variance at time t is $(\mathbf{b}'_D\Sigma_{Di}\mathbf{b}_D+\sigma_D{}^2)\,t$. Thus, if $\Sigma_{Di}=\Sigma_D$, for all i, then obviously, all stimuli will elicit the same drift variance.

REFERENCES

Ashby, F. G. (1983). A biased random walk model for two choice reaction times. *Journal of Mathematical Psychology*, **27**, 277–297.

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: Elsevier.

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., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neurospychological 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. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, **120**, 150–172.

Ashby, F. G., & Lee, W. W. (1992). On the relationship among identification, similarity and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General*, **121**, 385–393.

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: Elsevier.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 598–612.

Ashby, F. G., & Maddox, W. T. (1991). A response time theory of perceptual independence. In J. P. Doignon & J. C. Falmagne, (Eds.), *Mathematical psychology: Current developments*, pp. 389–413. New York: Springer-Verlag.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 50–71.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372–400.

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., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, **95**, 124–150.

Ashby, F. G., & Schwarz, W. (1996). Stochastic general recognition theory. Paper presented at the International Congress of Psychology, Montreal, Canada.

Ashby, F. G., Tein, J. Y., & Balakrishnan, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, **37**, 526–555.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154–179.

Balakrishnan, J. D. (1998a). Measures and interpretations of vigilance performance: Evidence against the detection criterion. *Human Factors*, **40**, 601–623.

Balakrishnan, J. D. (1998b). Some more sensitive measures of sensitivity and response bias. *Psychological Methods*, **3**, 68–90.

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. London: Methuen.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Laming, D. (1968). *Information theory of choice reaction times*. New York: Academic Press.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77–105.

Luce, R. D. (1986). *Response times*. New York: Oxford.

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.

Murdock, B. B. (1985). An analysis of the strength-latency hypothesis. *Memory & Cognition*, **13**, 571–521.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266–300.

Pachella, R. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing*: *Tutorials in performance and cognition*. Hillsdale, NJ: Erlbaum.

Perrin, N. A. (1992). Uniting identification, similarity and preference: General recognition theory. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition*, pp. 123–146. Hillsdale, NJ: Erlbaum.

Perrin, N. A., & Ashby, F. G. (1991). A test of perceptual independence with dissimilarity data. *Applied Psychological Measurement*, **15**, 79–93.

Pike, A. R. (1973). Response latency models for signal detection. *Psychological Review*, **80**, 53–68.

Pike, R., Dalgleish, L., & Wright, J. (1977). A multiple-observations model for response latency and the latencies of correct and incorrect responses in recognition memory. *Memory & Cognition*, **5**, 580–589.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59–108.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, **88**, 552–572.

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative response. *Psychological Review*, **92**, 212–225.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347–356.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261–300.

Stone, M. (1960). Model for choice-reaction time. *Psychometrika*, **25**, 251–260.

Tanner, W. P. (1956). Theory of recognition. *The Journal of the Acoustical Society of America*, **28**, 882–888.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, **61**, 401–409.

Thomas, E. A. C. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology*, **8**, 303–332.

Thomas, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology*, **40**, 318–341.

Townsed, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge Univ. Press.

Vickers, D., Caudrey, D., & Willson, R. (1971). Discriminating between the frequency of occurrence of two alternative events. *Acta Psychologica*, **35**, 151–172.

Received: October 1, 1997