Springer
the language of science

Dear Author/Editor,

Here are the proofs of your chapter as well as the metadata sheets.

**Metadata**

- Please carefully proof read the metadata, above all the names and address.
- In case there were no abstracts for this book submitted with the manuscript, the first 10-15 lines of the first paragraph were taken. In case you want to replace these default abstracts, please submit new abstracts with your proof corrections.

**Page proofs**

- Please check the proofs and mark your corrections either by
  - entering your corrections online
    or
  - opening the PDF file in Adobe Acrobat and inserting your corrections using the tool "Comment and Markup"
    or
  - printing the file and marking corrections on hardcopy. Please mark all corrections in dark pen in the text and in the margin at least ¼” (6 mm) from the edge.
- You can upload your annotated PDF file or your corrected printout on our Proofing Website. In case you are not able to scan the printout , send us the corrected pages via fax.
- Please note that any changes at this stage are limited to typographical errors and serious errors of fact.
- If the figures were converted to black and white, please check that the quality of such figures is sufficient and that all references to color in any text discussing the figures is changed accordingly. If the quality of some figures is judged to be insufficient, please send an improved grayscale figure.

# Metadata of the chapter that will be visualized online

| | | |
|---|---|---|
| Book Title | | Model-Based Cognitive Neuroscience |
| Chapter Title | | An Introduction to the Diffusion Model of Decision Making |
| Copyright | | Springer Science+Business Media, LLC 2015 |
| Corresponding Author | Prefix | |
| | Family name | **Smith** |
| | Particle | |
| | Given name | **Philip L.** |
| | Suffix | |
| | Division | Melbourne School of Psychological Sciences |
| | Organization | The University of Melbourne |
| | Address | Melbourne, VIC, Australia |
| Author | Prefix | |
| | Family name | **Ratcliff** |
| | Particle | |
| | Given name | **Roger** |
| | Suffix | |
| | Division | Department of Psychology |
| | Organization | The Ohio State University |
| | Address | Columbus, OH, USA |
| Abstract | | The diffusion model assumes that two-choice decisions are made by accumulating successive samples of noisy evidence to a response criterion. The model has a pair of criteria that represent the amounts of evidence needed to make each response. The time taken to reach criterion determines the decision time and the criterion that is reached first determines the response. The model predicts choice probabilities and the distributions of response times for correct responses and errors as a function of experimental conditions such as stimulus discriminability, speed-accuracy instructions, and manipulations of relative stimulus frequency, which affect response bias. This chapter describes the main features of the model, including mathematical methods for obtaining response time predictions, methods for fitting it to experimental data, including alternative fitting criteria, and ways to represent the fit to multiple experimental conditions graphically in a compact way. The chapter concludes with a discussion of recent work in psychology that links evidence accumulation to processes of perception, attention, and memory, and in neuroscience, to neural firing rates in the oculomotor control system in monkeys performing saccade-to-target decision tasks. |
| Keywords | | Diffusion process - Random walk - Decision-making - Response time - Choice probability |

# Chapter 3
# An Introduction to the Diffusion Model of Decision Making

**Philip L. Smith and Roger Ratcliff**

**Abstract** The diffusion model assumes that two-choice decisions are made by accumulating successive samples of noisy evidence to a response criterion. The model has a pair of criteria that represent the amounts of evidence needed to make each response. The time taken to reach criterion determines the decision time and the criterion that is reached first determines the response. The model predicts choice probabilities and the distributions of response times for correct responses and errors as a function of experimental conditions such as stimulus discriminability, speed-accuracy instructions, and manipulations of relative stimulus frequency, which affect response bias. This chapter describes the main features of the model, including mathematical methods for obtaining response time predictions, methods for fitting it to experimental data, including alternative fitting criteria, and ways to represent the fit to multiple experimental conditions graphically in a compact way. The chapter concludes with a discussion of recent work in psychology that links evidence accumulation to processes of perception, attention, and memory, and in neuroscience, to neural firing rates in the oculomotor control system in monkeys performing saccade-to-target decision tasks.

## 3.1 Historical Origins

The human ability to translate perception into action, which we share with nonhuman animals, relies on our ability to make rapid decisions about the contents of our environment. Any form of coordinated, goal-directed action requires that we be able to recognize things in the environment as belonging to particular cognitive categories or classes and to select the appropriate actions to perform in response. To a very significant extent, coordinated action depends on our ability to provide rapid answers to questions of the form: "What is it?" and "What should I do about it?" When viewed in this way, the ability to make rapid decisions—to distinguish

P. L. Smith (✉)
Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC, Australia

R. Ratcliff
Department of Psychology, The Ohio State University, Columbus, OH, USA



B. U. Forstmann, E.-J. Wagenmakers (eds.), *Model-Based Cognitive Neuroscience*,
DOI 10.1007/978-1-4939-2236-9_3

**Book ID: 300277_1_En ChapterID: 3 Dispatch Date: 04-12-2014 Proof No: 1**

predator from prey, or friend from foe—appears as one of the basic functions of the brain and central nervous system. The purpose of this chapter is to provide an introduction to the mathematical modeling of decisions of this kind.

Historically, the study of decision-making in psychology has been closely connected to the study of sensation and perception—an intellectual tradition with its origins in philosophy and extending back to the nineteenth century. Two strands of this tradition are relevant: psychophysics, defined as the study of the relationship between the physical magnitudes of stimuli and the sensations they produce, and the study of reaction time or response time (RT). Psychophysics, which had its origins in the work of Gustav Fechner in the Netherlands in 1860 on "just noticeable differences," led to the systematic study of decisions about stimuli that are difficult to detect or to discriminate. The study of RT was initiated by Franciscus Donders, also in the Netherlands, in 1868. Donders, inspired by the pioneering work of Hermann von Helmholtz on the speed of nerve conduction, sought to develop methods to measure the speed of mental processes. These two strands of inquiry were motivated by different theoretical concerns, but led to a common realization, namely, that decision-making is inherently variable. People do not always make the same response to repeated presentation of the same stimulus and the time they take to respond to it varies from one presentation to the next.

Trial-to-trial variation in performance is a feature of an important class of models for speeded, two-choice decision-making developed in psychology, known as *sequential-sampling* models. These models regard variation in decision outcomes and decision times as the empirical signature of a noisy evidence accumulation process. They assume that, to make a decision, the decision maker accumulates successive samples of noisy evidence over time, until sufficient evidence for a response is obtained. The samples represent the momentary evidence favoring particular decision alternatives at consecutive time points. The decision time is the time taken to accumulate a sufficient, or criterion, amount of evidence and the decision outcome depends on the alternative for which a criterion amount of evidence is first obtained. The idea that decision processes are noisy was first proposed on theoretical grounds, to explain the trial-to-trial variability in behavioral data, many decades before it was possible to use microelectrodes in awake, behaving animals to record this variability directly. The noise was assumed to reflect the moment-to-moment variability in the cognitive or neural processes that represent the stimulus [1–4].

In this chapter, we describe one such sequential-sampling model, the diffusion model of Ratcliff [5]. Diffusion models, along with random walk models, comprise one of the two main subclasses of sequential-sampling models in psychology; the other subclass comprises accumulator and counter models. For space reasons, we do not consider models of this latter class in this chapter. The interested reader is referred to references [2–4] and [6] for discussions. To distinguish Ratcliff's model from other models that also represent evidence accumulation as a diffusion process, we refer to it as the *standard diffusion model.* Historically, this model was the first model to represent evidence accumulation in two-choice decision making as a diffusion process and it remains, conceptually and mathematically, the benchmark against



Book ID: 300277_1_En ChapterID: 3 Dispatch Date: 04-12-2014 Proof No: 1

which other models can be compared. It is also the model that has been most extensively and successfully applied to empirical data. We restrict our consideration here to two-alternative decision tasks, which historically and theoretically have been the most important class of tasks in psychology.

## 3.2 Diffusion Processes and Random Walks

Mathematically, diffusion processes are the continuous-time counterparts of random walks, which historically preceded them as models for decision-making. A random walk is defined as the running cumulative sum of a sequence of independent random variables, $Z_j$, $j = 1, 2, \ldots$. In models of decision-making, the values of these variables are interpreted as the evidence in a sequence of discrete observations of the stimulus. Typically, evidence is assumed to be sampled at a constant rate, which is determined by the minimum time needed to acquire a single sample of perceptual information, denoted $\Delta$. The random variables are assumed to take on positive and negative values, with positive values being evidence for one response, say $R_a$, and negative values evidence for the other response, $R_b$. For example, in a brightness discrimination task, $R_a$ might correspond to the response "bright" and $R_b$ correspond to the response "dim." The mean of the random variables is assumed to be positive or negative, depending on the stimulus presented. The cumulative sum of the random variables,

$$X_i = \sum_{j=1}^{i} Z_j,$$

is a random walk. If the $Z_j$ are real-valued, the domain of the walk is the positive integers and the range is the real numbers. To make a decision, the decision-maker sets a pair of evidence criteria, $a$ and $b$, with $b < 0 < a$ and accumulates evidence until the cumulative evidence total reaches or exceeds one of the criteria, that is, until $X_i \geq a$ or $X_i \leq b$. The time taken for this to occur is the *first passage time* through one of the criteria, defined formally as

$$T_a = \min\{i\Delta : \; X_i \geq a | X_k > b; \; k < i\}$$
$$T_b = \min\{i\Delta : \; X_i \leq b | X_k < a; \; k < i\}.$$

If the first criterion reached is $a$, the decision maker makes response $R_a$; if it is $b$, the decision maker makes response $R_b$. The decision time, $T_D$, is the time for this to occur

$$T_D = \min\{T_a, T_b\}.$$

If response $R_a$ is identified as the correct response for the stimulus presented, then the mean, or expected value, of $T_a$, denoted $E[T_a]$, is the mean decision time for



Book ID: 300277_1_En ChapterID: 3 Dispatch Date: 04-12-2014 Proof No: 1

correct responses; $E[T_b]$ is the mean decision time for errors, and the probability of a correct response, $P(C)$, is the *first passage probability* of the random walk through the criterion $a$,

$$P(C) = \text{Prob}\{T_a < T_b\}.$$

Although either $T_a$ or $T_b$ may be infinite on a given realization of the process, $T_D$ will be finite with probability one; that is, the process will terminate with one or other response in finite time [7]. This means that the probability of an error response, $P(E)$, will equal $1 - P(C)$.

Random walk models of decision-making have been proposed by a variety of authors. The earliest of them were influenced by Wald's sequential probability ratio test (SPRT) in statistics [8] and assumed that the random variables $Z_j$ were the log-likelihood ratios that the evidence at each step came from one as opposed to the other stimulus. The most highly-developed of the SPRT models was proposed by Laming [9]. The later *relative judgment theory* of Link and Heath [10] assumed that the decision process accumulates the values of the noisy evidence samples directly rather than their log-likelihood ratios. Evaluation of these models focused primarily on the relationship between mean RT and accuracy and the ordering of mean RTs for correct responses and errors as a function of experimental manipulations [2–4, 9, 10].

## 3.3 The Standard Diffusion Model

A diffusion process may be thought of as random walk in continuous time. Instead of accumulating evidence at discrete time points, evidence is accumulated continuously. Such a process can be obtained mathematically via a limiting process, in which the sampling interval is allowed to go to zero while constraining the average size of the evidence at each step to ensure the variability of the process in a given, fixed time interval remains constant [7, 11]. The study of diffusion processes was initiated by Albert Einstein, who proposed a diffusion model for the movement of a pollen particle undergoing random Brownian motion [11]. The rigorous study of such processes was initiated by Norbert Wiener [12]. For this reason, the simplest diffusion process is known variously as the Wiener process or the Brownian motion process.

In psychology, Ratcliff [5] proposed a diffusion model of evidence accumulation in two-choice decision-making—in part because it seemed more natural to assume that the brain accumulates information continuously rather than at discrete time points. Ratcliff also emphasized the importance of studying RT distributions as a way to evaluate models. Sequential-sampling models not only predict choice probabilities and mean RTs, they predict entire distributions of RTs for correct responses and errors. This provides for very rich contact between theory and experimental data, allowing for strong empirical tests.

The main elements of the standard diffusion model are shown in Fig. 3.1. We shall denote the accumulating evidence state in the model as $X_t$, where $t$ denotes time.



Book ID: 300277_1_En ChapterID: 3 Dispatch Date: 04-12-2014 Proof No: 1

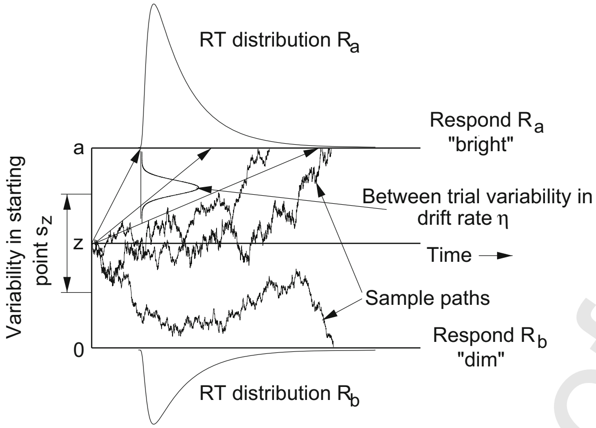


**Fig. 3.1** Diffusion model. The process starting at $z$ accumulates evidence between decision criteria at 0 and $a$. Moment-to-moment variability in the accumulation process means the process can terminate rapidly at the correct response criterion, slowly at the correct response criterion, or at the incorrect response criterion. There is between-trial variability in the drift rate, $\xi$, with standard deviation $\eta$, and between-trial variability in the starting point, $z$, with range $s_z$

Before describing the model, we should mention that there are two conventions used in psychology to characterize diffusion models. The convention used in the preceding section assumes the process starts at zero and that the criteria are located at $a$ and $b$, with $b < 0 < a$. The other is based on Feller's [13] analysis of the so-called gambler's ruin problem and assumes that the process starts at $z$ and that the criteria are located at 0 and $a$, with $0 < z < a$. As the latter convention was used by Ratcliff in his original presentation of the model [5] and in later work, this is the convention we shall adopt for the remainder of this chapter. The properties of the process are unaltered by translations of the starting point; such processes are called *spatially homogeneous*. For processes of this kind, a change in convention simply represents a relabeling of the $y$-axis that represents the accumulating evidence state. Other, more complex, diffusion processes, like the Ornstein-Uhlenbeck process [14–16], are not spatially homogeneous and their properties are altered by changes in the assumed placement of the starting point.

As shown in the figure, the process, starting at $z$, begins accumulating evidence at time $t = 0$. The rate at which evidence accumulates, termed the *drift* of the process and denoted $\xi$, depends on the stimulus that is presented and its discriminability. The identity of the stimulus determines the direction of drift and the discriminatory of the stimulus determines the magnitude. Our convention is that when stimulus $s_a$ is presented the drift is positive and the value of $X_t$ tends to increase with time, making it is more likely to terminate at the upper criterion and result in response $R_a$. When stimulus $s_b$ is presented the drift is negative and the value of $X_t$ tends to decrease with time, making it is more likely to terminate at the lower boundary with response $R_b$. In our example brightness discrimination task, bright stimuli lead to positive values of drift and dim stimuli lead to negative values of drift. Highly discriminable stimuli are associated with larger values of drift, which lead to more rapid information accumulation and faster responding. Because of noise in the process, the accumulating evidence is subject to moment-to-moment perturbations. The time course of evidence accumulation on three different experimental trials, all with the same drift rate, is shown in the figure. These noisy trajectories are termed the *sample paths* of the process. A unique sample path describes the time course of evidence accumulation on a given experimental trial. The sample paths in the figure show some of the different outcomes that are possible for stimuli with the same drift rate. The sample paths in the figure show: (a) a process terminating with a correct response made rapidly; (b) a process terminating with a correct response made slowly, and (c) a process terminating with an error response. In behavioral experiments, only the response and the RT are observables; the paths themselves are not. They are theoretical constructs used to explain the observed behavior.

The noisiness, or variability, in the accumulating evidence is controlled by a second parameter, the *infinitesimal standard deviation*, denoted $s$. Its square, $s^2$, is termed the *diffusion coefficient*. The diffusion coefficient determines the variability in the sample paths of the process. Because the parameters of a diffusion model are only identified to the level of a ratio, all the parameters of the model can be multiplied by a constant without affecting any of the predictions. To make the parameters estimable, it is common practice to fix $s$ arbitrarily. The other parameters of the model are then expressed in units of infinitesimal standard deviation, or infinitesimal standard deviation per unit time.

## 3.4 Components of Processing

As shown in Fig. 3.1, the diffusion model predicts RT distributions for correct responses and errors. Moment-to-moment variability in the sample paths of the process, controlled by the diffusion coefficient, means that on some trials the process will finish rapidly and on others it will finish slowly. The predicted RT distributions have a characteristic unimodal, positively-skewed shape: More of the probability mass in the distribution is located below the mean than above it. As the drift of the process changes with changes in stimulus discriminability, the relative proportions of correct responses and errors change, and the means and standard deviations of the RT distributions also change. However, the shapes of the RT distributions change very little; to a good approximation, RT distributions for low discriminability stimuli are scaled copies of those for high discriminability stimuli [17].

One of the main strengths of the diffusion model is that the shapes of the RT distributions it predicts are precisely those found in empirical data. Many experimental tasks, including low-level perceptual tasks like signal detection and higher-level cognitive tasks like lexical decision and recognition memory, yield families of RT distributions like those predicted by the model [6]. In contrast, other models, particularly those of the accumulator/counter model class predict distribution shapes that become more symmetrical with reductions in discriminability [6]. Such distributions tend not to be found empirically, except in situations in which people are forced to respond to an external deadline.

One of the problems with early random walk models of decision-making—which they shared with the simplest form of the diffusion model—is they predicted that mean RTs for correct responses and errors would be equal [2]. Specifically, if $E[R_j|s_i]$, denotes the mean RT for response $R_j$ to stimulus $s_i$, with $i, j \in \{a, b\}$, then, if the drifts for the two stimuli are equal in magnitude and opposite in sign, as is natural to assume for many perceptual tasks, the models predicted that $E[R_a|s_a] = E[R_a|s_b]$ and $E[R_b|s_a] = E[R_b|s_b]$; that is, the mean time for a given response made correctly is the same as the mean time for that response made incorrectly. They also predicted, when the starting point is located equidistantly between the criteria, $z = a/2$, that $E[R_a|s_a] = E[R_b|s_a]$ and $E[R_a|s_b] = E[R_b|s_b]$; that is, the mean RT for correct responses to a given stimuli is the same as the mean error RT to that same stimulus. This prediction holds regardless of the relative magnitudes of the drifts. Indeed, a stronger prediction holds; the models predicted equality not only of mean RTs, but of the entire distributions of correct responses and errors. These predictions almost never hold empirically. Rather, the typical finding is that when discriminability is high and speed is stressed, error mean times are shorter than correct mean times. When discriminability is low and accuracy is stressed, error mean times are longer than correct mean times [2]. Some studies show a crossover pattern, in which errors are faster than correct responses in some conditions and slower in others [6].

A number of modifications to random walk models were proposed to deal with the problem of the ordering of mean RTs for correct responses and errors, including asymmetry (non-normality) of the distributions of evidence that drive the walk [1, 10], and biasing of an assumed log-likelihood computation on the stimulus information at each step [18], but none of them provided a completely satisfactory account of the full range of experimental findings. The diffusion model attributes inequality of the RTs for correct responses and errors to between-trial variability in the operating characteristics, or "components of processing," of the model. The diffusion model predicts equality of correct and error times only when the sole source of variability in the model is the moment-to-moment variation in the accumulation process. Given the complex interaction of perceptual and cognitive processes involved in decision-making, such an assumption is probably an oversimplification. A more realistic assumption is that there is trial-to-trial variability, both in the quality of information entering the decision process and in the decision-maker's setting of decision criteria or starting points. Trial-to-trial variability in the information entering the decision process would arise either from variability in the efficiency of the perceptual encoding of stimuli or from variation in the quality of the information provided by nominally equivalent stimuli. Trial-to-trial variability in decision criteria or starting points would arise as the result of the decision-maker attempting to optimize the speed and accuracy of responding [4]. Most RT tasks show sequential effects, in which the speed and accuracy of responding depends on the stimuli and/or the responses made on preceding trials, consistent with the idea that there is some kind of adaptive regulation of the settings of the decision process occurring across trials [2, 4].

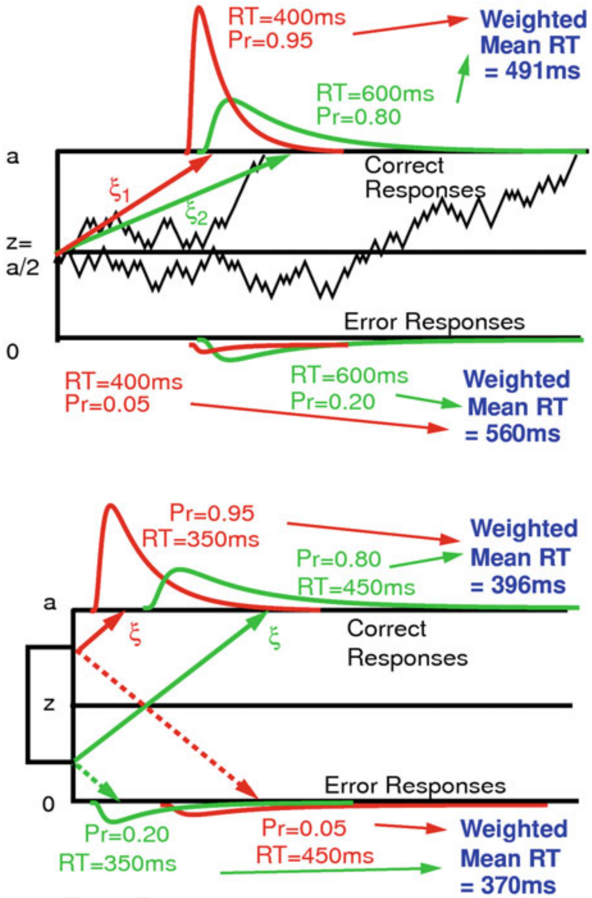


**Fig. 3.2** Effects of trial-to-trial variability in drift rates and starting points. The predicted RT distributions are probability mixtures across processes with different drift rates (*top*) or different starting points (*bottom*). Variability in drift rates leads to slow errors; variability in starting points leads to fast errors

The diffusion model assumes that there is trial-to-trial variation in both drift rates and starting points. Ratcliff [5] assumed that the drift rate on any trial, $\xi$, is drawn from a normal distribution with mean $\nu$ and standard deviation $\eta$. Subsequently Ratcliff, Van Zandt, and McKoon [19] assumed that there is also trial-to-trial variability in the starting point, $z$, which they modeled as a rectangular distribution with range $s_z$. They chose a rectangular distribution mainly on the grounds of convenience, because the predictions of the model are relatively insensitive to the distribution's form. The main requirement is that all of the probability mass of the distribution must lie between the decision criteria, which is satisfied by a rectangular distribution with $s_z$ suitably constrained. The distributions of drift and starting point are shown in Fig. 3.1.

Trial-to-trial variation in drift rates allows the model to predict slow errors; trial-to-trial variation in starting point allows it to predict fast errors. The combination of the two allows it to predict crossover interactions, in which there are fast errors for high discriminability stimuli and slow errors for low discriminability stimuli. Figure 3.2a shows how trial-to-trial variability in drift results in slow errors. The assumption that

Book ID: 300277_1_En ChapterID: 3 Dispatch Date: 04-12-2014 Proof No: 1

drift rates vary across trials means that the predicted RT distributions are probability mixtures, made up of trials with different values of drift. When the drift is small (i.e., near zero), error rates will be high and RTs will be long. When the drift is large, error rates will be low and RTs will be short. Because errors are more likely on trials on which the drift is small, a disproportionate number of the trials in the error distribution will be trials with small drifts and long RTs. Conversely, because errors are less likely on trials on which drift is large, a disproportionate number of the trials in the correct response distribution will be trials with large drifts and short RTs. In either instance, the predicted mean RT will be the weighted mean of the RTs on trials with small drift and large drifts.

Figure 3.2a illustrates how slow errors arise in a simplified case in which there are just two drifts, $\xi_1$ and $\xi_2$, with $\xi_1 > \xi_2$. When the drift is $\xi_1$, the mean RT is 400 ms and the probability of a correct response, $P(C)$, is 0.95. When the drift is $\xi_2$, the mean RT is 600 and $P(C) = 0.80$. The predicted mean RTs are the weighted means of large drift and small drift trials. The predicted mean RT for correct responses is $(0.95 \times 400 + 0.80 \times 600)/1.75 = 491$ ms. The predicted mean for error responses $(0.05 \times 400 + 0.20 \times 600)/0.25 = 560$ ms. Rather than just two drifts, the diffusion model assumes that the predicted means for correct responses and errors are weighted means across an entire normal distribution of drift. However, the effect is the same: predicted mean RTs errors are longer than those for correct responses.

Figure 3.2b illustrates how fast errors arise as the result of variation in starting point. Again, we have shown a simplified case, in which there are just two starting points, one of which is closer to the lower, error, response criterion and the other of which is closer to the upper, correct, response criterion. In this example, a single value, of drift, $\xi$, has been assumed for all trials. The model predicts fast errors because the mean time for the process to reach criterion depends on the distance it has to travel and because it is more likely to terminate at a particular criterion if the criterion is near the starting point rather than far from it. When the starting point is close to the lower criterion, errors are faster and also more probable. When the starting point is close to the upper criterion, errors are slower, because the process has to travel further to reach the error criterion, and are less probable. Once again, the predicted distributions of correct responses and errors are probability mixtures across trials with different values of starting point.

In the example shown in Fig. 3.2b, when the process starts near the upper criterion, the mean RT for correct responses is 350 ms and $P(C) = 0.95$. When it starts near the lower criterion, the mean RT for correct responses is 450 ms and $P(C) = 0.80$. The predicted mean RTs for correct responses and errors are again the weighted means across starting points. In this example, the mean RT for correct responses is $(0.95 \times 350 + 0.80 \times 450)/1.75 = 396$ ms; the mean RT for errors is $(0.20 \times 350 + 0.05 \times 450)/0.25 = 370$ ms. Again, the model assumes that the predicted mean times are weighted means across the entire distribution of starting points, but the effect is the same: predicted mean times for errors are faster than those correct responses. When equipped with both variability in drift and starting point, the model can predict both the fast errors and the slow errors that are found experimentally [6].

The final component of processing in the model is the non-decision time, denoted $T_{\text{er}}$. Like many other models in psychology, diffusion model assumes that RT can be additively decomposed into the decision time, $T_D$, and the time for other processes, $T_{\text{er}}$:

$$RT = T_D + T_{\text{er}}.$$

The subscript in the notation means "encoding and responding." In many applications of the model, it suffices to treat $T_{\text{er}}$ as a constant. In practice, this is equivalent to assuming that it is an independent random variable whose variance is negligible compared to that of $T_D$. In other applications, particularly those in which discriminability is high and speed is emphasized and RT distributions have small variances, the data are better described by assuming that $T_{\text{er}}$ is rectangularly distributed with range $s_t$. As with the distribution of starting point, the rectangular distribution is used mainly as a convenience, because when the variance of $T_{\text{er}}$ is small compared to that of $T_D$, the shape of the distribution will be determined almost completely by the shape of the distribution of decision times. The advantage of assuming some variability in $T_{\text{er}}$ in these settings is that it allows the model to better capture the leading edge of the empirical RT distributions, which characterizes the fastest 5–10 % of responses, and which tends to be slightly more variable than the model predicts.

## 3.5 Bias and Speed-Accuracy Tradeoff Effects

Bias effects and speed-accuracy tradeoff effects are ubiquitous in experimental psychology. Bias effects typically arise when the two stimulus alternatives occur with unequal frequency or have unequal rewards attached to them. Speed-accuracy tradeoff effects arise as the result of explicit instructions emphasizing speed or accuracy or as the result of an implicit set on the part of the decision-maker. Such effects can be troublesome in studies that measure only accuracy or only RT, because of the asymmetrical way in which these variables can be traded off. Small changes in accuracy can be traded off against large changes in RT, which can sometimes make it difficult to interpret a single variable in isolation [2].

One of the attractive features of sequential-sampling models like the diffusion model is that they provide a natural account of how speed-accuracy tradeoffs arise. As shown in Fig. 3.3, the models assume that criteria are under the decision-maker's control. Moving the criteria further from the starting point (i.e., increasing $a$ while keeping $z = a/2$) increases the distance the process must travel to reach a criterion and also reduces the probability that it will terminate at the wrong criterion because of the cumulative effects of noise. The effect of increasing criteria will thus be slower and more accurate responding. This is the speed-accuracy tradeoff.

The diffusion model with variation in drift and starting point can account for the interactions with experimental instructions emphasizing speed or accuracy that are found experimentally. When accuracy is emphasized and criteria are set far from the starting point, variations in drift have a greater effect on performance than do



Book ID: 300277_1_En ChapterID: 3 Dispatch Date: 04-12-2014 Proof No: 1

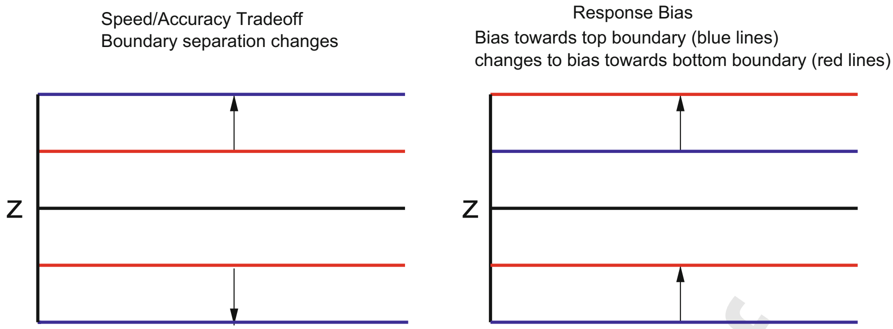


**Fig. 3.3** Speed-accuracy tradeoff and response bias. Reducing decision criteria leads to faster and less accurate responding. Shifting the starting point biases the process towards the response associated with the nearer criterion

variations in starting point, and so slow errors are found. When speed is emphasized and criteria are near the starting point, variations in starting point have a greater effect on performance than do variations in drift and fast errors are found.

Like other sequential-sampling models, the diffusion model accounts for bias effects by assuming unequal criteria, represented by a shift in the starting point towards the upper or lower criterion, as shown in Fig. 3.3. Shifting the starting point towards a particular response criterion increases the probability of that response and reduces the average time taken to make it. The probability of making the other response is reduced and the average time to make it is correspondingly increased. The effect of changing the prior probabilities of the two responses, by manipulating the relative stimulus frequencies, is well described by a change in the starting point (unequal decision criteria). In contrast, unequal reward rates not only lead to a bias in decision criteria, they also lead to a bias in the way stimulus information is classified [20]. This can be captured in the idea of a *drift criterion*, which is a criterion on the stimulus information, like the criterion in signal detection theory. The effect of changing the drift criterion is to make the drift rates for the two stimuli unequal. Both kinds of bias effects appear to operate in tasks with unequal reward rates.

## 3.6 Mathematical Methods For Diffusion Models

Diffusion processes can be defined mathematically either via partial differential equations or by stochastic differential equations. If $f(\tau, y; t, x)$ is the transition density of the process $X_t$, that is, $f(\tau, y; t, x)\,dx$ is the probability that a process starting at time $\tau$ in state $y$ will be found at time $t$ in a small interval $(x, x + dx)$, then the accumulation process $X_t$, with drift $\xi$ and diffusion coefficient $s^2$, satisfies the partial differential equation

$$-\frac{\partial f}{\partial \tau} = \frac{1}{2} s^2 \frac{\partial^2 f}{\partial y^2} + \xi \frac{\partial f}{\partial y}.$$

This equation is known in the probability literature as Kolmogorov's backward equation, so called because its variables are the starting time $\tau$ and the initial state $y$. The process also satisfies a related equation known as Kolmogorov's forward equation, which is an equation in $t$ and $x$ [7, 11]. The backward equation is used to derive RT distributions; the forward equation is useful for studying evidence accumulation by a process unconstrained by criteria [5].

Alternatively, the process can be defined as satisfying the stochastic differential equation [11]:

$$dX_t = \xi dt + s\, dW_t.$$

The latter equation is useful because it provides a more direct physical intuition about the properties of the accumulation process. Here $dX_t$ is interpreted as the small, random change in the accumulated evidence occurring in a small time interval of duration $dt$. The equation says that the change in evidence is the sum of a deterministic and a random part. The deterministic part is proportional to the drift rate, $\xi$; the random part is proportional to the infinitesimal standard deviation, $s$. The term on the right, $dW_t$, is the differential of a Brownian motion or Wiener process, $W_t$. It can be thought of as the random change in the accumulation process during the interval $dt$ when it is subject to the effects of many small, independent random perturbations, described mathematically as a *white noise* process. White noise is a mathematical abstraction, which cannot be realized physically, but it provides a useful approximation to characterize the properties of physical systems that are perturbed by broad-spectrum, Gaussian noise. Stochastic differential equations are usually written in the differential form given here, rather than in the more familiar form involving derivatives, because of the extreme irregularity of the sample paths of diffusion processes, which means that quantities of the form $dX_t/dt$ are not well defined mathematically.

Solution of the backward equation leads to an infinite series expression for the predicted RT distributions and an associated expression for accuracy[5, 7, 11]. The stochastic differential equation approach leads to a class of integral equation methods that were developed in mathematical biology to study the properties of integrate-and-fire neurons. The interested reader is referred to references [6, 16, 21] for details. For a two-boundary process with drift $\xi$, boundary separation $a$, starting point $z$, and infinitesimal standard deviation $s$, with no variability in any of its parameters, the probability of responding at the upper barrier, $P(\xi, a, z)$, is

$$P(\xi, a, z) = \frac{\exp\left(-2\xi a/s^2\right) - \exp\left(-2\xi z/s^2\right)}{\exp\left(-2\xi a/s^2\right) - 1}.$$

The cumulative distribution of first passage times at the upper boundary, $a$, is

$$G(t, \xi, a, z) = P(\xi, a, z) - \frac{\pi s^2}{a^2} e^{-\xi z/s^2} \sum_{k=1}^{\infty} \frac{2k \sin\left(\frac{k\pi z}{a}\right) \exp\left\{-\frac{1}{2}\left(\frac{\xi^2}{s^2} + \frac{k^2\pi^2 s^2}{a^2}\right) t\right\}}{\left(\frac{\xi^2}{s^2} + \frac{k^2\pi^2 s^2}{a^2}\right)}.$$

The probability of a response and the cumulative distribution of first passage times at the lower boundary are obtained by replacing $\xi$ with $-\xi$ and $z$ with $a - z$ in the preceding expressions. More details can be found in reference [5].

In addition to the partial differential equation and integral equation methods, predictions for diffusion models can also be obtained using finite-state Markov chain methods or by Monte Carlo simulation [22]. The Markov chain approach, developed by Diederich and Busemeyer [23], approximates a continuous-time, continuous-state, diffusion process by a discrete-time, discrete-state, birth-death process. A transition matrix is defined that specifies the probability of an increment or a decrement to the process, conditional on its current state. The entries in the transition matrix express the relationship between the drift and diffusion coefficients of the diffusion process and the transition probabilities of the approximating Markov chain [24]. The transition matrix includes two special entries that represent criterion states, which are set equal to 1.0, expressing the fact that once the process has transitioned into a criterion state, it does not leave it. An initial state vector is defined, which represents the distribution of probability mass at the beginning of the trial, including the effects of any starting point variation. First passage times and probabilities can then be obtained by repeatedly multiplying the state vector by the transition matrix. These alternative methods are useful for more complex models for which an infinite-series solution may not be available. There are now software packages available for fitting the standard diffusion model that avoid the need to implement the model from first principles [25–27].

## 3.7 The Representation of Empirical Data

The diffusion model predicts accuracy and distributions of RT for correct responses and errors as a function of the experimental variables. In many experimental settings, the discriminability of the stimuli is manipulated as a within-block variable, while instructions, payoffs, or prior probabilities are manipulated as between-block variables. The model assumes that manipulations of discriminability affect drift rates, while manipulations of other variables affect criteria or starting points. Although criteria and starting points can vary from trial to trial, they are assumed to be independent of drift rates, and to have the same average value for all stimuli in a block. This assumption provides an important constraint in model testing.

To show the effects of discriminability variations on accuracy and RT distributions, the data and the predictions of the model are represented in the form of a *quantile-probability plot*, as shown in Fig. 3.4. To construct such a plot, each of the RT distributions is summarized by an equal-area histogram. Each RT distribution is represented by a set of rectangles, each representing 20 % of the probability mass in the distribution, except for the two rectangles at the extremes of the distribution, which together represent the 20 % of mass in the upper and lower tails. The time-axis bounds of the rectangles are distribution quantiles, that is, those values of time that cut off specified proportions of the mass in the distribution. Formally, the $p$th

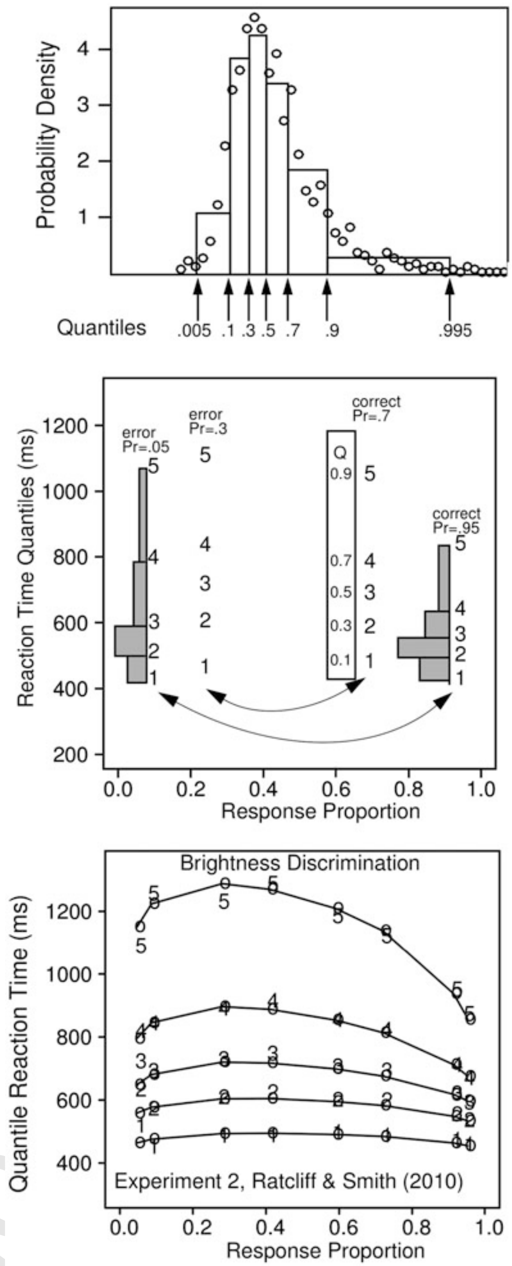


**Fig. 3.4** Representing data in a quantile probability plot. *Top panel*: An empirical RT distribution is summarized using an equal-area histogram with bins bounded by the distribution quantiles. *Middle panel*: The quantiles of the RT distributions for correct responses and errors are plotted vertically against the probability of a correct response on the *right* and the probability of an error response on the *left*. *Bottom panel*: Example of an empirical quantile probability plot from a brightness discrimination experiment

quantile, $Q_p$, is defined to be the value of time such that the proportion of RTs in the distribution that are less than or equal to $Q_p$ is equal to $p$. The distribution in the figure has been summarized using five quantiles: the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles. The 0.1 and 0.9 quantiles represent the upper and lower tails of the distribution, that is, the fastest and slowest responses, respectively. The 0.5 quantile is the median and represents the distribution's central tendency. As shown in the figure, the set of five quantiles provides a good summary of the location, variability, and shape of the distribution.

To construct a quantile probability plot, the quantile RTs for correct responses and errors are plotted on the $y$-axis against the choice probabilities (i.e., accuracy) on the $x$-axis for each stimulus condition, as shown in the middle panel of the figure. Specifically, if, $Q_{i,p}(C)$ and $Q_{i,p}(E)$ are, respectively, the quantiles of the RT distributions for correct responses and errors in condition $i$ of the experiment, and $P_i(C)$ and $P_i(E)$ are the probabilities of a correct response and an error in that condition, then the values of $Q_{i,p}(C)$ are plotted vertically against $P_i(C)$ for $p = 0.1, 0.3, 0.5, 0.7, 0.9$, and the values of $Q_{i,p}(E)$ are similarly plotted against $P_i(E)$. All of the distribution pairs and choice probabilities from each condition are plotted in a similar way.

The bottom panel of the figure shows data from a brightness discrimination experiment from Ratcliff and Smith [28] in which four different levels of stimulus discriminability were used. Because of the way the plot is constructed, the two outermost distributions in the plot represent performance for the most discriminable stimuli and the two innermost distributions represent performance for the least discriminable stimuli. The value of the quantile-probability plot is that it shows how performance varies parametrically as stimulus discriminability is altered, and how different parts of the RT distributions for correct responses and errors are affected differently. As shown in the figure, most of the change in the RT distribution with changing discriminability occurs in the upper tail of the distribution (e.g., the 0.7 and 0.9 quantiles); there is very little change in the leading edge (the 0.1 quantile). This pattern is found in many perceptual tasks and also in more cognitive tasks like recognition memory. The quantile-probability plot also shows that errors were slower than correct responses in all conditions. This appears as a left-right asymmetry in the plot; if the distributions for correct responses and errors were the same, the plot would be mirror-image symmetrical around its vertical midline. The predicted degree of asymmetry is a function of the standard deviation of the distribution of drift rates, $\eta$ and, when there are fast errors, of the range of starting points, $s_z$. The slow-error pattern of data in Fig. 3.4 is typical of difficult discrimination tasks in which accuracy is emphasized.

The pattern of data is Fig. 3.4 is rich and highly-constrained and represents a challenge for any model. The success of the diffusion model is that it has shown repeatedly that it can account for data of this kind. Its ability to do so is not a just a matter of model flexibility. It is not the case that the model is able to account for any pattern of data whatsoever [29]. Rather, as noted previously, the model predicts families of RT distributions that have a specific and quite restricted form. Distributions of this particular form are the ones most often found in experimental data.

## 3.8 Fitting the Model to Experimental Data

Fitting the model to experimental data requires estimation of its parameters by iterative, nonlinear minimization. A variety of minimization algorithms have been used in the literature, but the Nelder-Mead SIMPLEX algorithm has been popular because of its robustness [30]. Parameters are estimated to minimize a fit statistic, or loss function, that characterizes the discrepancy between the model and the data. A variety of fit statistics have been used in applications, but chi-square-type statistics, either the Pearson chi-square ($\chi^2$) or the likelihood-ratio chi-square ($G^2$), are common. For an experiment with $m$ stimulus conditions, these are defined as

$$\chi^2 = \sum_{i=1}^{m} n_i \sum_{j=1}^{12} \frac{(p_{ij} - \pi_{ij})^2}{\pi_{ij}}$$

and

$$G^2 = 2\sum_{i=1}^{m} n_i \sum_{j=1}^{12} p_{ij} \ln\left(\frac{p_{ij}}{\pi_{ij}}\right),$$

respectively. In these equations, the outer summation over $i$ indexes the $m$ conditions in the experiment and the inner summation over $j$ indexes the 12 bins defined by the quantiles of the RT distributions for correct responses and errors. (The use of five quantiles per distribution gives six bins per distribution, or 12 bins per correct and error distribution pair.) The quantities $p_{ij}$ and $\pi_{ij}$ are the observed and predicted proportions of probability mass in each bin, respectively, and $n_i$ is the number of stimuli in the $i$th experimental condition. For bins defined by the quantile bounds, the values of $p_{ij}$ will equal 0.2 or 0.1, depending on whether or not the bin is associated with a tail quantile, and the values of $\pi_{ij}$ are the differences in the probability mass in the cumulative finishing time distributions, evaluated at adjacent quantiles, $G(Q_{i,p}, \nu, a, z) - G(Q_{i,p-1}, \nu, a, z)$. Here we have written the cumulative distribution as a function of the mean drift, $\nu$, rather than the trial-dependent drift, $\xi$, to emphasize that the cumulative distributions are probability mixtures across a normal distribution of drift values. Because the fit statistics keep track of the distribution of probability mass across the distributions of correct responses and errors, minimizing them fits both RT and accuracy simultaneously.

Fitting the model typically requires estimation of around 8–10 parameters. For an experiment with a single experimental condition and four different stimulus discriminabilities like the one shown in Fig. 3.4, a total of 10 parameters must be estimated to fit the full model. There are four values of the mean drift, $\nu_i$, $i = 1, \ldots, 4$, a boundary separation parameter, $a$, a starting point, $z$, a non-decision time, $T_{\mathrm{er}}$, and variability parameters for the drift, starting point, and non-decision time, $\eta$, $s_z$, and $s_t$, respectively. As noted previously, to make the model estimable, the infinitesimal standard deviation is typically fixed to an arbitrary value (Ratcliff uses $s = 0.1$ in his work, but $s = 1.0$ has also been used). In experiments in which there is no evidence of response bias, the data can be pooled across the two responses to create one distribution of correct responses and one distribution of errors per stimulus condition. Under these conditions, a symmetrical decision process can be assumed ($z = a/2$) and the number of free parameters reduced by one. Also, as discussed previously, in many applications the non-decision time variability parameter can be set to zero without worsening the fit.

Although the model has a reasonably large number of free parameters, it affords a high degree of data reduction, defined as the number of degrees of freedom in the data divided by the number of free parameters in the model. There are $11m$ degrees of freedom in a data set with $m$ conditions and six bins per distribution (one degree of freedom is lost for each correct-error distribution pair, because the expected and observed masses are constrained to be equal in each pair, giving $12 - 1 = 11$ degrees of freedom per pair). For the experiment in Fig. 3.4, there are 44 degrees of freedom in the data and the model had nine free parameters, which represents a data reduction ratio of almost 5:1. For larger data sets, data reduction ratios of better than 10:1 are common. This represents a high degree of parsimony and explanatory power.

It is possible to fit the diffusion model by maximum likelihood instead of by minimum chi-square. Maximum likelihood defines a fit statistic (a likelihood function) on the set of raw RTs rather than on the probability mass in the set of bins, and maximizes this (i.e., minimizes its negative). Despite the theoretical appeal of maximum likelihood, its disadvantage is that it is vulnerable to the effects of contaminants or outliers in a distribution. Almost all data sets have a small proportion of contaminant responses in them, whether from finger errors or from lapses in vigilance or attention, or other causes. RTs from such trials are not representative of the process of theoretical interest. Because maximum likelihood requires that all RTs be assigned a non-zero likelihood, outliers of this kind can disrupt fitting and estimation, whereas minimum chi-square is much less susceptible to such effects [31].

Many applications of the diffusion model have fitted it to group data, obtained by quantile-averaging the RT distributions across participants. A group data set is created by averaging the corresponding quantiles, $Q_{i,p}$, for each distribution of correct responses and errors in each experimental condition across participants. The choice probabilities in each condition are also averaged across participants. The advantage of group data is that it is less noisy and variable than individual data. A potential concern when working with group data is that quantile averaging may distort the shapes of the individual distributions, but in practice, the model appears to be robust to averaging artifacts. Studies comparing fits of the model to group and individual data have found that both methods lead to similar conclusions. In particular, the averages of the parameters estimated by fitting the model to individual data agree fairly well with the parameters estimated by fitting the model to quantile-averaged group data [32, 33]. Although the effects of averaging have not been formally characterized, the robustness of the model to averaging may be a result of the relative invariance of its families of distribution shapes, discussed previously.

## 3.9 The Psychophysical Basis of Drift

The diffusion model has been extremely successful in characterizing performance in a wide variety of speeded perceptual and cognitive tasks, but it does so by assuming that all of the information in the stimulus can be represented by a single value of drift, which is a free parameter of the model, and that the time course of the stimulus encoding processes that determine the drift can be subsumed within the non-decision time, $T_{\text{er}}$, which is also a free parameter. Recent work has sought to characterize the perceptual, memory, and attentional processes involved in the computation of drift and how the time course of these processes affects the time course of decision making [34].

Developments in this area have been motivated by recent applications of the diffusion model to psychophysical discrimination tasks, in which stimuli are presented very briefly, often at very low levels of contrast and followed by backward masks to limit stimulus persistence. Surprisingly, performance in these tasks is well described by the standard diffusion model, in which the drift rate is constant for the duration of an experimental trial [35, 36]. The RT distributions found in these tasks resemble those obtained from tasks with response-terminated stimuli, like those in Fig. 3.4, and show no evidence of increasing skewness at low stimulus discriminability, as would be expected if the decision process were driven by a decaying perceptual trace. The most natural interpretation of this finding is that the drift rate in the decision process depends on a durable representation of the stimulus stored in visual short-term memory (VSTM), which preserves the information it contains for the duration of an experimental trial.

This idea was incorporated in the *integrated system model* of Smith and Ratcliff [34], which combines submodels of perceptual encoding, attention, VSTM, and decision-making in a continuous-flow architecture. It assumes that transient stimulus information encoded by early visual filters is transferred to VSTM under the control of spatial attention and the rate at which evidence is accumulated by the decision process depends on the time-varying strength of the VSTM trace. Because the VSTM trace is time-varying, the decision process in the model is *time-inhomogeneous*. Predictions for time-inhomogeneous diffusion processes cannot be obtained using the infinite-series method, but can be obtained using either the integral equation method [16] or the Markov chain approximation [23]. The integrated system model has provided a good account of performance in tasks in which attention is manipulated by spatial cues and discriminability is limited by varying stimulus contrast or backward masks. It has also provided a theoretical link between stimulus contrast and drift rates, and an account of the shifts in RT distributions that occur when stimuli are embedded in dynamic noise, which is one of the situations in which the standard model fails [28, 37]. The main contribution of the model to our understanding of simple decision tasks is to show how performance in these tasks depends on the time course of processes of perception, memory, attention, and decision-making acting in concert.



Book ID: 300277_1_En ChapterID: 3 Dispatch Date: 04-12-2014 Proof No: 1

## 3.10 Conclusion

Recently, there has been a burgeoning of interest in the diffusion model and related models in psychology and in neuroscience. In psychology, this has come from the realization that the model can provide an account of the effects of stimulus information, response bias, and response caution (speed-accuracy tradeoff) on performance in simple decision tasks, and a way to characterize these components of processing quantitatively in populations and in individuals. In neuroscience, it has come from studies recording from single cells in structures of the oculomotor systems of awake behaving monkeys performing saccade-to-target decision tasks. Neural firing rates in these structures are well-characterized by assuming that they provide an online read-out of the process of accumulating evidence to a response criterion [38]. This interpretation has been supported by the finding that the parameters of a diffusion model estimated from monkeys' RT distributions and choice probabilities can predict firing rates in the interval prior to the overt response [39, 40]. These results linking behavioral and neural levels of analysis have been accompanied by theoretical analyses showing how diffusive evidence accumulation at the behavioral level can arise by aggregating the information carried in individual neurons across the cells in a population [41, 42].

There has also been recent interest in investigating alternative models that exhibit diffusive, or diffusion-like, model properties. Some of these investigations have been motivated by a quest for increased neural realism, and the resulting models have included features like racing evidence totals, decay, and mutual inhibition [43]. Although arguments have been made for the importance of such features in a model, and although these models have had some successes, none has yet been applied as systematically and as successfully to as wide a range of experimental tasks as has the standard diffusion model.

## 3.11 Suggestions for Further Reading

Anyone wishing to properly understand the RT literature should begin with Luce's (1986) classic monograph, *Response Times* [2]. Although the field has developed rapidly in the years since it was published, it remains unsurpassed in the depth and breadth of its analysis. Ratcliff's (1978) *Psychological Review* article [5] is the fundamental reference for the diffusion model, while Ratcliff and Smith's (2004) *Psychological Review* article [6] provides a detailed empirical comparison of the diffusion model and other sequential-sampling models. Smith and Ratcliff's (2004) *Trends in Neuroscience* article [38] discusses the emerging link between psychological models of decision-making and neuroscience.

## Exercises

Simulate a random walk with normally-distributed increments in Matlab, R, or some other software package. Use your simulation to obtain predicted RT distributions and choice probabilities for a range of different accumulation rates (means of the random variables, $Z_i$). Use a small time step of, say, 0.001 s to ensure you obtain a good approximation to a diffusion process and simulate 5000 trials or more for each condition. In most experiments to which the diffusion model is applied, decisions are usually made in around a second or less, so try to pick parameters for your simulation that generate RT distributions on the range 0–1.5 s.

1. The drift rate, $\xi$, and the infinitesimal standard deviation, $s$, of a diffusion process describe the change occurring in a unit time interval (e.g., during one second). If $\xi_{\mathrm{rw}}$ and $s_{\mathrm{rw}}$ denote, respectively, the mean and standard deviation of the distribution of increments, $Z_i$, to the random walk, what values must they be set to in order to obtain a drift rate of $\xi = 0.2$ and an infinitesimal standard deviation of $s = 0.1$ in the diffusion process? (Hint: The increments to a random walk are independent and the means and variances of sums of independent random variables are both additive).
2. Verify that your simulation yields unimodal, positively-skewed RT distributions like those in Fig. 3.1. What is the relationship between the distribution of correct responses and the distribution of errors? What does this imply about the relationship between the mean RTs for correct responses and errors?
3. Obtain RT distributions for a range of different drift rates. Drift rates of $\xi = \{0.4, 0.3, 0.2, 0.1\}$ with a boundary separation $a = 0.1$ are likely to be good choices with $s = 0.1$. Calculate the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles of the distributions of RT for each drift rate. Construct a Q-Q (quantile-quantile) plot by plotting the quantiles of the RT distributions for each of the four drift conditions on the $y$-axis against the quantiles of the largest drift rate (e.g., $\xi = 0.4$) condition on the $x$-axis. What does a plot of this kind tell you about the families of RT distributions predicted by a model?
4. Compare the Q-Q plot from your simulation to the empirical Q-Q plots reported by Ratcliff and Smith [28] in their Fig. 20. What do you conclude about the relationship?
5. Read Wagenmakers and Brown [17]. How does the relationship they identify between the mean and variance of empirical RT distributions follow from the properties of the model revealed in the Q-Q plot?

## Solutions (These Go in a Separate Book of Answers)

1. You need to set $\xi_{\mathrm{rw}} = \xi h$ and $s_{\mathrm{rw}} = s\sqrt{h}$, where $h$ is the time step of the random walk. The number of increments, $n$, to the random walk in one second is $n = 1/h$, so a sum of $n$ independent random variables, each with mean $\xi_{\mathrm{rw}}$ and standard

deviation $s_{rw}$, will have a mean of $n\xi_{rw} = \xi$ and a variance of $ns_{rw}^2 = s^2$ and a standard deviation of $s$.

2. Your simulation should have yielded *joint distributions* of RT. The probability mass in each of the joint distributions is equal to the probability of making the associated response. You should find that, within the limits of the accuracy of your simulation, that the joint distributions of correct responses and errors for a given drift rate should be scaled copies of each other. *Conditional distributions* are obtained by dividing the joint distributions of RT for correct responses and errors by their associated response probabilities. making the probability mass in each distribution equal to 1.0. The conditional distributions should, within the limits of your simulation, be identical to one another. If the distributions of correct responses and errors are the same, the means (and variances) of correct responses and errors will be equal.
3. The Q-Q plot shows how the means, standard deviations, and shapes of the RT distributions vary as the drift of the process is systematically varied. The diffusion process (and the approximating random walk) generate Q-Q plots that are highly linear. This means that, to a good approximation, the predicted RT distribution in one condition can be obtained from the distribution in another condition by rescaling the time axis.
4. The empirical Q-Q plots reported by Ratcliff and Smith (2010) are highly linear, in agreement with the simulation.
5. Wagenmakers and Brown (2007) investigated the relationship between the mean and standard deviation of RT across a range of discriminability conditions in a number of different experiments. In each experiment, they found that the means and standard deviations of the RT distributions, considered as functions of the stimulus condition, were linearly related to one another. A linear relationship between the mean and standard deviation follows from the linearity of the predicted Q-Q plot, because the RT distribution in one condition, and hence also the mean and standard deviation, can obtained from that in another condition by multiplying the time scale by a constant.



## References


1. Link, SW (1992) The wave theory of difference and similarity. Erlbaum, Englewood Cliffs
2. Luce RD (1986) Response times. Oxford University Press, New York
3. Townsend JT, Ashby FG (1983) Stochastic modeling of elementary psychological processes. Cambridge University Press, Cambridge
4. Vickers D (1979) Decision processes in visual perception. Academic, London
5. Ratcliff R (1978) A theory of memory retrieval. Psychol Rev 85:59–108
6. Ratcliff R, Smith PL (2004) A comparison of sequential-sampling models for two choice reaction time. Psychol Rev 111:333–367
7. Cox DR, Miller HD (1965) The theory of stochastic processes. Chapman & Hall, London.
8. Wald A (1947) Sequential analysis. Wiley, New York
9. Laming DRJ (1968) Information theory of choice reaction time. Wiley, New York
10. Link SW, Heath RA (1975) A sequential theory of psychological discrimination. Psychometrika 40:77–105

11. Gardiner CW (2004) Handbook of stochastic methods, 3rd edn. Springer, Berlin
12. Wiener N (1923) Differential space. J Math Phys 2:131–174
13. Feller W (1967) An introduction to probability theory and its applications, 3rd edn. Wiley, New York
14. Busemeyer J, Townsend JT (1992) Fundamental derivations from decision field theory. Math Soc Sci 23:255–282
15. Busemeyer J, Townsend JT (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol Rev 100:432–459
16. Smith PL (2000) Stochastic dynamic models of response time and accuracy: a foundational primer. J Math Psychol 44:408–463
17. Wagenmakers E-J, Brown S (2007) On the linear relationship between the mean and standard deviation of a response time distribution. Psychol Rev 114:830–841
18. Ashby FG (1983) A biased random walk model for two choice reaction time. J Math Psychol 27:277–297
19. Ratcliff R, Van Zandt T, McKoon G (1999) Connectionist and diffusion models of reaction time. Psychol Rev 106:261–300
20. Leite FP, Ratcliff R (2011) What cognitive processes drive response biases? A diffusion model analysis. Judgm Decis Mak 6:651–687
21. Buonocore A, Giorno V, Nobile AG, Ricciardi L (1990) On the two-boundary first-crossing-time problem for diffusion processes. J Appl Probab 27:102–114
22. Tuerlinckx F, Maris E, Ratcliff R, De Boeck P (2001) A comparison of four methods for simulating the diffusion process. Behav Res Methods Instrum Comput 33:443–456
23. Diederich A, Busemeyer JR (2003) Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. J Math Psychol 47:304–322
24. Bhattacharya RB, Waymire EC (1990) Stochastic processes with applications. Wiley, New York
25. Vandekerckhove J, Tuerlinckx F (2008) Diffusion model analysis with MATLAB: a DMAT primer. Behav Res Methods 40:61–72
26. Vandekerckhove J, Tuerlinckx F, Lee MD (2011) Hierarchical diffusion models for two-choice response times. Psychol Res 16:44–62
27. Voss A, Voss J (2008) A fast numerical algorithm for the estimation of diffusion model parameters. J Math Psychol 52:1–9
28. Ratcliff R, Smith PL (2010) Perceptual discrimination in static and dynamic noise: the temporal relationship between perceptual encoding and decision making. J Exp Psychol Gen 139:70–94
29. Ratcliff R (2002) A diffusion model account of response time and accuracy in a brightness discrimination task: fitting real data and failing to fit fake but plausible data. Psychon Bull Rev 9:278–291
30. Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
31. Ratcliff R, Tuerlinckx F (2002) Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. Psychon Bull Rev 9:438–481
32. Ratcliff R, Thapar A, McKoon G (2003) A diffusion model analysis of the effects of aging on brightness discrimination. Percept Psychophys 65:523–535
33. Ratcliff R, Thapar A, McKoon G (2004) A diffusion model analysis of the effects of aging on recognition memory. J Mem Lang 50:408–424
34. Smith PL, Ratcliff R (2009) An integrated theory of attention and decision making in visual signal detection. Psychol Rev 116:283–317
35. Ratcliff R, Rouder J (2000) A diffusion model account of masking in two-choice letter identification. J Exp Psychol Hum Percept Perform 26:127–140
36. Smith PL, Ratcliff R, Wolfgang BJ (2004) Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. Vis Res 44:1297–1320
37. Smith PL, Ratcliff R, Sewell DK (in press) Modeling perceptual discrimination in dynamic noise: time-changed diffusion and release from inhibition. J Math Psychol

38. Smith PL, Ratcliff R (2004) Psychology and neurobiology of simple decisions. Trends Neurosci 27:161–168
39. Ratcliff R, Cherian A, Segraves M (2003) A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. J Neurophysiol 90:1392–1407
40. Ratcliff R, Hasegawa Y, Hasegawa R, Smith PL, Segraves M (2007) A dual diffusion model for single cell recording data from the superior colliculus in a brightness discrimination task. J Neurophysiol 97:1756–1797
41. Smith PL (2010) From poisson shot noise to the integrated Ornstein-Uhlenbeck process: Neurally-principled models of diffusive evidence accumulation in decision-making and response time. J Math Psychol 54:266–283
42. Smith PL, McKenzie CRL (2011) Diffusive information accumulation by minimal recurrent neural models of decision-making. Neural Comput 23:2000–2031
43. Usher M, McClelland JL (2001) The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 108:550–592