

Modeling Simple Decisions and Applications Using a Diffusion Model

Roger Ratcliff and Philip Smith

Abstract

The diffusion model is one of the major sequential-sampling models for two-choice decision-making and choice response time in psychology. The model conceives of decision-making as a process in which noisy evidence is accumulated until one of two response criteria is reached and the associated response is made. The criteria represent the amount of evidence needed to make each decision and reflect the decision maker's response biases and speed-accuracy trade-off settings. In this chapter we examine the application of the diffusion model in a variety of different settings. We discuss the optimality of the model and review its applications to a number of cognitive tasks, including perception, memory, and language tasks. We also consider its applications to normal and special populations, to the cognitive foundations of individual differences, to value-based decisions, and its role in understanding the neural basis of decision-making.

Key Words: diffusion model, sequential-sampling, drift rate, choice, decision time, accuracy, confidence, perceptual decision, memory decision, lexical decision

Diffusion Models for Rapid Decisions

Over the last 30 or 40 years, there has been a steady development of models for simple decision-making that deal with both the accuracy of decisions and the time taken to make them. The models assume that decisions are made by accumulating noisy information to decision criteria, one criterion for each possible choice. The models successfully account for the probability that each choice is made and the response time (RT) distributions for correct responses and errors. The models are highly constrained by the behavior of these dependent variables. The most frequent applications of these models have been to tasks that require two-choice decisions that are made reasonably quickly, typically with mean RTs less than 1.0–2.0 s. This is fast enough that one can assume that the decisions come from a single decision process and not from multiple, sequential processes (anything much slower and the single-process assumption would be suspect).

The models have been applied successfully to many different tasks including perceptual, numerical, and memory tasks with a variety of subject populations, including older adults, children, dyslexics, and adults undergoing sleep deprivation, reduced blood sugar, or alcohol intoxication.

An important feature of human decision-making is that the processing system is very flexible because humans can switch tasks, stimulus dimensions, and output modalities very quickly, from one trial to the next. There are many different kinds of decisions that can be made about any stimulus. If the stimulus is a letter string, decisions can be made about whether it is word or a nonword, whether it was studied earlier, whether the color is red or green, whether it is upper or lower case, and so on. Responses can be made in different modalities and in different ways in those modalities (for example, manually, vocally, or via eye movements). The same decision mechanism might operate for all these

tasks or the mechanism might be task and modality specific.

For two-choice tasks, the assumption usually made is that all decision-related information, that is, all the information that comes from a stimulus or memory, is collapsed onto a single variable, called *drift rate*, that characterizes the discriminative or preference information in the stimulus. In some situations, subjects may be asked to make judgments based on more than one dimension that cannot be combined in this way. In such cases, the systems factorial methods of Townsend and colleagues (e.g., Townsend, 1972; see the review in Townsend & Wenger, 2004) may be able to be used to determine whether processing on the different dimensions is serial or parallel, or some hybrid of the two.

In this chapter, we focus on one model of the class of sequential sampling models of evidence accumulation, the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008; Smith, 2000). A comparison of the diffusion model with other sequential-sampling models, such as the Poisson counter model (Townsend & Ashby, 1983), the Vickers accumulator model (Smith & Vickers, 1988; Vickers, 1970), and the leaky competing accumulator model (Usher & McClelland, 2001) can be found in Ratcliff and Smith (2004). In the diffusion model, for a two-choice task, noisy evidence accumulates from a starting point (Figure 3.1), toward one of two decision criteria or boundaries and the quality of the information that enters the decision process determines the rate of accumulation.

Fitting the model to data provides independent estimates of drift rates, decision boundaries, and a parameter representing the duration of nondecision processes. The model's ability to separate these components is one of its key contributions and places major constraints on its ability to explain data. Stimulus difficulty affects drift rate but not the criteria, and to a good approximation, speed-accuracy shifts are represented in the criteria, not drift rate. If difficulty varies, changes in drift rate alone must accommodate all the changes in performance, namely accuracy and the changes in the spreads and locations of the correct and error RT distributions. Likewise, changes in the criteria affect all the aspects of performance. In these ways, the model is tightly constrained by data.

In a perceptual task, drift rate depends on the quality of the perceptual information from a stimulus; in a memory task, it depends on the quality of the match between a test item and memory. In

a brightness discrimination task, for example, if the accumulated evidence reaches the top boundary, a "bright" response is executed and a "dark" response would then correspond to the bottom boundary. Figure 3.1 shows an example, using a brightness discrimination task. Evidence accumulates from a stimulus to the "bright" boundary or to the "dark" boundary. The solid arrow shows the drift rate for a bright stimulus, the dashed arrow shows the drift rate for a less bright stimulus, and the dotted arrow shows the drift rate for a dark stimulus.

The three paths in Figure 3.1 show three different outcomes, all with the same drift rate. Noise in the accumulation process produces errors when the accumulated evidence reaches the incorrect boundary and it produces variable RTs that form a distribution of RTs that has the shape of empirically obtained distributions. In the figure, one path leads to a fast correct decision, one to a slow correct decision, and one to an error. Most responses are reasonably fast, but there are slower ones that spread out the right-hand tails of the distributions (as in the distribution at the top of Figure 3.1). As drift rate changes from a large value to near zero, the mean of the RT distribution for both correct and error responses increases because the tail of the RT distribution spreads out. Figure 3.2 shows simulated individual RTs from the model as a function of drift rate, which is assumed to vary from trial to trial. The shortest RTs change little with drift rate, and so a fast response says nothing about the difficulty of the trial. The probability of obtaining a slow response from a high drift rate is very small (e.g., Figure 3.2) and so conditions with the slowest responses come from lower drift rates (see Ratcliff, Philastides, & Sajda, 2009).

Figure 3.1 show the accumulation-of-evidence process. Besides this, there are processes that

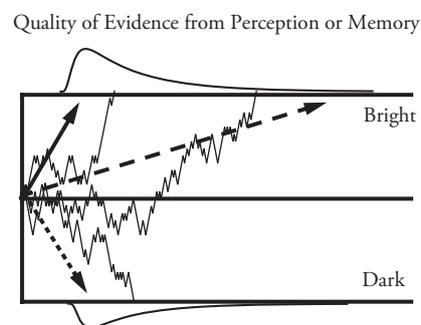


Fig. 3.1 The diffusion decision model with three simulated paths and three different drift rates.

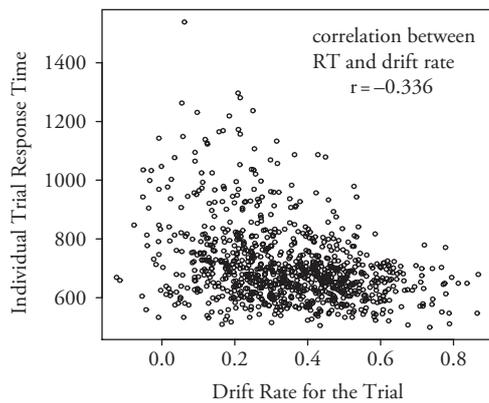


Fig. 3.2 Plots of individual RTs as a function of drift rate for the trial. The parameters of the diffusion model were, boundary separation, $a = 0.107$, starting point $z = 0.048$, duration of processes other than the decision process, $T_{er} = 0.48$ s, SD in drift rate across trial $\eta = 0.20$, range in starting points $s_z = 0.02$, range in nondecision time $s_t = 0.18$ s, drift rate $v = 0.3$.

encode stimuli, access memory, transform stimulus information into a decision-related variable that determines drift rate, and execute responses. These components of processing are combined into one “nondecision” component in the model, that has mean T_{er} . The total processing time for a decision is the sum of the time taken by the decision process and the time taken by the nondecision component.

The boundaries of the decision process can be manipulated by instructions (“respond as quickly as possible” or “respond as accurately as possible”), differential rewards for the two choices, and the relative frequencies with which the two stimuli are presented in the experiment. Changes in instructions, rewards, or biases affect both RTs and accuracy but in the model, to a good approximation, the effects on RTs and accuracy are due to shifts in boundary settings alone, not drift rates or nondecision time. (However, if subjects are pushed very hard to go fast, then nondecision time and drift rates can be lower (e.g., Starns, Ratcliff, & McKoon, 2012).)

Figure 3.3, left panel, shows boundaries moving in for speed relative to accuracy instructions and the right panel shows how subjects can be biased toward the top response versus the bottom response by moving decision criteria from the dashed line to the solid line settings. It is also possible (Figure 3.3 right panel) to adjust the zero point of drift rate (the drift rate criterion) to accommodate biases between the two responses (see Leite & Ratcliff, 2011; Ratcliff, 1985; Ratcliff & McKoon, 2008, Figure 3.3).

A problem with early random walk models, which were precursors to the diffusion model, was that they predicted equal correct and error RT distributions if the drift rates for two stimuli were equal in magnitude but opposite in sign (Laming, 1968; Stone, 1960; but see Link & Heath, 1975). This prediction is also made by the diffusion model in the absence of across-trial variability in model parameters. In fact, the patterns of the relative speed of correct versus error responses are as follows: with accuracy instructions and/or difficult tasks, errors are slower than correct responses, and with speed instructions and/or easy tasks, errors are faster than correct responses (Luce, 1986).

In the diffusion model, the observed patterns of correct versus error RTs fall out naturally because there is trial-to-trial variability in drift rate and starting point (e.g., Ratcliff, 1981). Figure 3.4 illustrates how this mixing works with just two drift rates or two starting points instead of their full distributions. In Figure 3.4 left panel, the v_1 drift rate produces high accuracy and fast responses, the v_2 one lower accuracy and slow responses. The mixture of these produces errors slower than correct responses because 5% of the 400 ms process averaged with 20% of the 600 ms process gives a weighted mean of 560 ms, which is slower than the weighted mean for correct responses (491 ms). In Figure 3.4, right panel, the distributions to the left are for processes that start near the correct boundary (the dotted arrow shows the distance the process has to go to make an error—the larger the distance, the slower the response) and the distributions to the right are for processes that start further away from the correct boundary. Processes that start near to the correct boundary have few errors and those errors are slow, whereas processes that start further away have more errors and the errors are fast, leading to errors faster than correct responses. In practice, drift rate is assumed to be normally distributed from trial to trial and the starting point is uniformly distributed, but these specific functional forms are not critical (Ratcliff, 2013).

Some researchers have argued that across-trial variability in the parameters is not needed (Palmer, Huk, & Shadlen, 2005; Usher & McClelland, 2001). However, it is unreasonable to assume that subjects can set their processing components to identical values on every equivalent trial of an experiment (i.e., ones with the same stimulus value). For drift rates, across-trial variability in drift rate is exactly analogous to variability in stimulus or memory strength in signal detection theory. Later

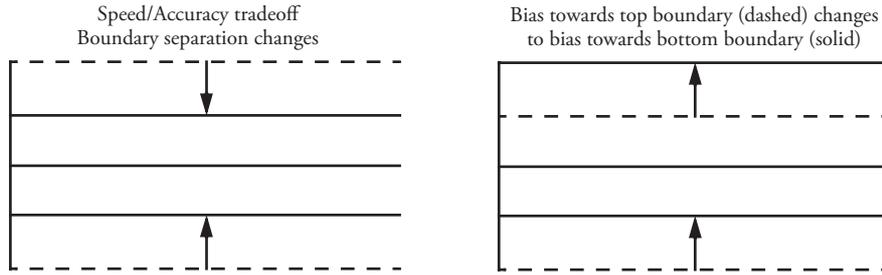


Fig. 3.3 In the left panel, boundary separation alone changes between speed and accuracy instructions. In the right panel, the starting point varies with bias.

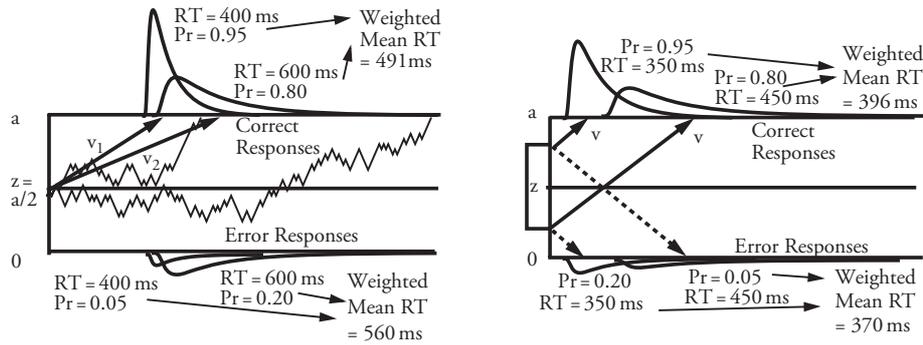


Fig. 3.4 Variability in drift rate and starting point and the effects on speed and accuracy. The left panel shows two process with drift rates v_1 and v_2 and the starting point halfway between the boundaries with correct and error RTs of 400 ms for v_1 and of 600 ms for v_2 . Averaging these two illustrates the effects of variability in drift rate across trials and in the illustration yields error responses slower than correct responses. The right panel shows processes with two starting points and drift rate v . Averaging processes with starting point $0.5a + 0.5$ (high accuracy and short RTs) and starting point $0.5a - 0.5$ (lower accuracy and short RTs) yields error responses faster than correct responses.

we describe an EEG study of perceptual decision-making that provides independent evidence for across-trial variability in drift rate and mention another that provides evidence for variability in starting point.

It is important to understand that the diffusion model is highly falsifiable, not by mean RTs and accuracy values but by RT distributions. If empirical distributions are not right skewed, and do not shift and spread in exactly the right ways across experimental conditions, the model is falsified. Ratcliff (2002) generated sets of data with RT distributions that are plausible but never obtained in real experiments. For one set, the shapes and locations of the RT distributions were changed as a function of task difficulty, and for the other, the shapes and locations were changed as a function of speed versus accuracy instructions. For none of the resulting distributions was the model able to fit the data. In addition, the distributional predictions of the model are tested every time it is fit to empirical data.

EXPRESSIONS FOR ACCURACY AND RT DISTRIBUTIONS

For a two-boundary diffusion process with no across-trial variability in any of the parameters, the equation for accuracy, the proportion of responses terminating at the boundary at zero, is given by

$$P(v, a, z) = \frac{e^{-2va/s^2} - e^{-2vz/s^2}}{e^{-2va/s^2} - 1} \quad (1)$$

(or $1 - z/a$ if drift is zero), and the cumulative distribution of finishing times at the same boundary is given by

$$G(t, v, a, z) = P(v, a, z) - \frac{\pi s^2}{a^2} e^{-\frac{vz}{s^2}} \times \sum_{k=1}^{\infty} \frac{2k \sin\left(\frac{k\pi z}{a}\right) e^{-\frac{1}{2}\left(\frac{v^2}{s^2} + \frac{k^2\pi^2 s^2}{a^2}\right)t}}{\left(\frac{v^2}{s^2} + \frac{k^2\pi^2 s^2}{a^2}\right)} \quad (2)$$

where a is boundary separation (the top boundary is at a , the bottom boundary is at 0 and the

distribution of finishing times is the distribution at the bottom boundary), z is the starting point, v is drift rate, and s is the SD in the normal distribution of within-trial variability (square root of the diffusion coefficient).

These expressions can be derived as a solution of the partial differential equation for the first passage-time probability for the diffusion process (Feller, 1968). The results are described in detail in Ratcliff (1978) and Ratcliff and Smith (2004). Because Equation 2 contains an infinite sum, values of the RT density function need to be computed numerically. The series needs to be summed until it converges; this means that terms have to be added until subsequent terms become so small that they do not affect the total. This is complicated by the sine term, which can allow one value in the sum to be small, whereas the next one is not small. To deal with this practically, it is necessary to require that two or three successive terms are very small.

The predictions from the model are obtained by integrating the results from Equations 1 and 2 over the distributions of the model's across-trial variability parameters using numerical integration. In the standard model, drift rate is normally distributed across trials with SD η , the starting point is uniformly distributed with range s_z , and nondesign time is uniformly distributed with range s_t . The predicted values are "exact" numerical predictions in the sense that they can be made as accurate as necessary (e.g., 0.1 ms or better) by using more and more steps in the infinite sum and more and more steps in the numerical integrations (packages perform fitting are mentioned later).

Alternative computational methods for obtaining predictions for diffusion models have been described by Smith (2000) and Diederich and Busemeyer (2003). The approach described by Smith uses integral equation methods derived from renewal theory. It was originally developed in mathematical biology to model the firing rates of integrate-and-fire neurons (Buonocore, Giorno, Nobile, & Ricciardi, 1990). The method is more computationally intensive than the infinite series approach of Equation 2, but has the advantage that it can be applied to processes in which the drift rates or decision criteria change over time or in which the accumulated information decays during the course of a trial. Smith (1995) and Smith and Ratcliff (2009) have proposed models in which drift rates depend on the outputs of visual and memory processes that change during a trial.

They obtained predictions for these models using the integral equation method.

Diederich and Busemeyer (2003) proposed a matrix method for obtaining predictions for diffusion models. In their approach, a continuous-time, continuous-state diffusion process is approximated by a discrete-time, discrete-state birth-death process. The probability that the process takes a step up or down at each time point is characterized by a transition matrix whose entries express the rules by which the process evolves over time. By approximating the process in this way, the problem of obtaining RT distributions and response probabilities can be reduced to one of repeated matrix multiplication. This solution can be expensive computationally, but mean RTs and response probabilities can be obtained efficiently by solving the associated algebraic eigenvalue problem, avoiding the need for repeated matrix multiplication. The method can also be applied to more complex problems that cannot be solved using the method of Equation 2 and has the advantage that it is very robust computationally.

In some situations, it is important to generate predictions by simulation because simulated data can show the effects of all the sources of variability on a subject's RTs and accuracy. The number of simulated observations can be increased sufficiently that the data approach the predictions that would be determined exactly from the numerical method. The expression for the update of evidence, Δx , on every time step Δt during the decision process, is determined by the drift rate, v , plus a noise term (Gaussian random variable, ϵ_i with SD σ) to represent variability in processing:

$$\Delta x_i = v_i \Delta t + \sigma \eta_i \sqrt{\Delta t} \quad (3)$$

This equation provides the most straightforward method of simulating the diffusion process, but it is not the most efficient. Tuerlinckx, Maris, Ratcliff, & De Boeck (2001) examined four methods for simulating diffusion processes and found that a random walk approximation is better than using Equation 3. They also showed that a "rejection" method is even more efficient. However, if the process is nonstationary and complicated (e.g., with time varying drift rate, or boundaries that have some functional form) or there are several diffusion processes running to model multiple choice tasks, simulation is the simplest way to produce predictions, and the random walk approximation is likely the most efficient.

In fitting the diffusion model to data, accuracy and RT distributions for correct and error responses for all the conditions of the experiment must be simultaneously fit and the values of all of the components of processing estimated simultaneously. One commonly used fitting method uses quantiles of the RT distributions for correct and error responses for each condition (the 0.1, 0.3, 0.5, 0.7, and 0.9 quantile RTs). The model predicts the cumulative probability of a response at each RT quantile. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between adjacent quantiles. For a chi-square computation, these are the expected values, to be compared to the observed proportions of responses between the quantiles (i.e., the proportions between 0.1, 0.3, 0.5, 0.7, and 0.9, are each 0.2, and the proportions below 0.1 and above 0.9 are both 0.1) multiplied by the number of observations. Summing over $(\text{Observed}-\text{Expected})^2/\text{Expected}$ for correct and error responses for each condition gives a single chi-square value that is minimized with a general SIMPLEX minimization routine. The parameter values for the model are adjusted by SIMPLEX until the minimum chi-square value is obtained.

In any data set, there is the potential problem of outlier RTs, which could be fast (e.g., fast guesses) or slow (e.g., inattention). The quantile based method provides a good compromise that reduces the influence of outliers because the proportion of responses between the quantiles is used and extreme RTs within the bins have no influence on fitting. To additionally deal with outliers, a model of such processes is used in some model fitting approaches so that data is assumed to be a mixture of diffusion processes plus a small proportion of outliers. For details of the fitting methods for the standard diffusion model and modeling outliers, see Ratcliff and Tuerlinckx (2002)

New methods for fitting the diffusion model have been developed recently and, over the last 6 or 7 years, fitting packages have been made available by Vandekerckhove and Tuerlinckx (2007) and Voss and Voss (2007). Also, Bayesian methods have been developed (Vandekerckhove, Tuerlinckx, & Lee, 2011) and a Bayesian package by Wiecki, Sofer & Frank (2012) has been made available. These Bayesian methods also implement hierarchical modeling schemes, in which model parameters for individual subjects are assumed to be random samples from population distributions that are specified within the model. The means and

variances of the population distributions, which are estimated in fitting, determine a range of probable values of drift rates and decision boundaries for individual subjects. Because all subjects are fit simultaneously using these methods, the parameters are constrained by the group parameters especially with low numbers of observations. The application of these hierarchical methods are in their infancy and some applications with large numbers of subjects, both simulated and real, that show their benefit over and above the more traditional methods are needed.

To show how well the diffusion model fits data, we plot RT quantiles against the proportions for which the two responses are made. The top panel of Figure 3.5 shows a histogram for an RT distribution. The 0.1–0.9 quantile RTs and the 0.005 and 0.995 quantiles are shown on the *x*-axis. The rectangles represent equal areas of 0.2 probability mass between the 0.1–0.3, 0.3–0.5, etc. quantile RTs (and as can be seen, these represent the histogram reasonably well).

These quantiles can be used to construct a quantile-probability plot by plotting the 0.1–0.9 quantile RTs vertically, as in the second panel of Figure 3.5, against the response proportion of that condition on the *x*-axis. Usually, correct responses are on the right of 0.5 and errors to the left (if there is no bias toward one or the other of the responses). Example RT distributions constructed from the equal area rectangles are also shown in grey. When there is a bias in starting point or when the two response categories are not symmetric (as in lexical decision and memory experiments), two quantile probability are needed, one for each response category.

With quantile probability plots, changes in RT distribution locations and spread as a function of response proportion can be seen easily and compared with model fits. In the bottom panel of Figure 3.5, the 1–5 symbols are the data and the solid lines are the predictions from fits of the model to the data (with circles denoting the exact location of the predictions). As can be seen in this example, as response proportion changes from about 0.6 to near 1.0, the 0.1 quantile (leading edge) changes little, but the 0.9 quantile changes by as much as 400 ms. This is in line with the model predictions. Also, as can be seen, error responses are slower than correct responses mainly in the spread, not in the leading edge location. Thus, quantile-probability plots allow all the important aspects of the data to be read from a single plot.

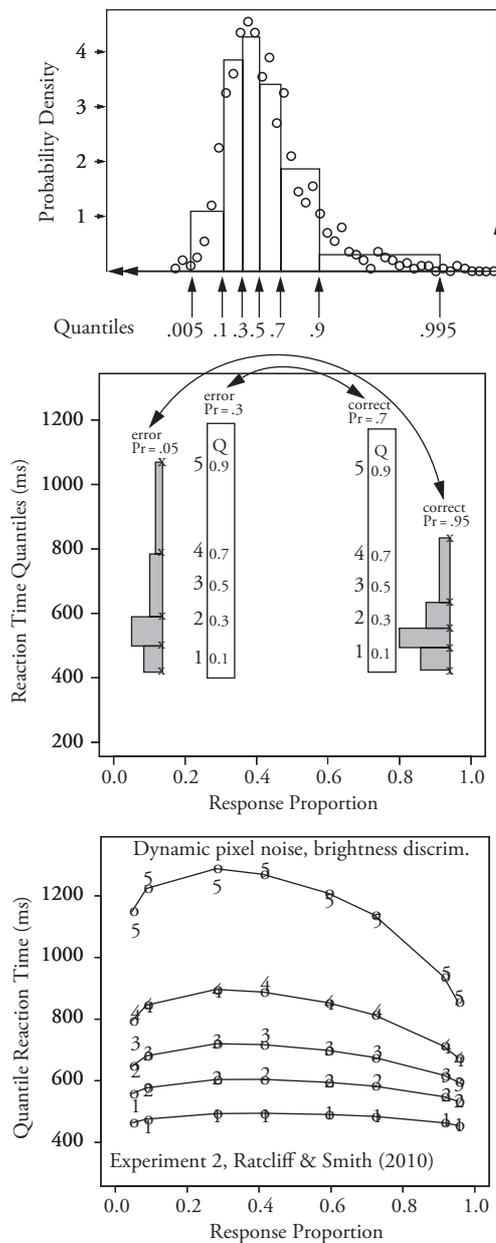


Fig. 3.5 The top panel shows a RT distribution overlaid with 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles, where the area outside the .1 quantile ranges from 0.005 to 0.1 and the area outside the .9 quantile ranges from 0.9 to 0.995. The areas between each pair of middle quantiles are 0.2 and the areas below 0.1 and above 0.9 are 0.095. The quantile rectangles capture the main features of the RT distribution and therefore a reasonable summary of overall distribution shape. The middle panel shows quantile RTs for the 0.1, 0.3, 0.5 (median), 0.7, and 0.9 quantiles (stacked vertically) plotted against response proportion for each of the six conditions. Correct responses are plotted to the right, and error responses to the left. The bottom panel shows a quantile probability function from Ratcliff and Smith (2010, Experiment 2) with the numbers representing data and the lines representing predictions.

Variants of the Standard Two-Choice Task

Up to this point, we have discussed how the diffusion model explains the results of experiments in which subjects respond with one of the two choices in their own time. The model has also been successfully applied to paradigms in which decision time is manipulated. Here we discuss three of these.

RESPONSE SIGNAL AND DEADLINE TASKS

For response signal and deadline tasks, a signal is presented after the stimulus and a subject is required to respond as quickly as possible (in, say, 200–300 ms). For a deadline paradigm, the time between the stimulus and the signal is fixed across trials. For a response signal paradigm, the time varies from trial to trial (Reed, 1973; Schouten & Bekker, 1967; Wickelgren, 1977; Wickelgren, Corbett, & Doshier, 1980). With the deadline paradigm, subjects can adopt different strategies or criteria for different deadlines. This is not the case for the response signal paradigm in which processing can be assumed to be the same up to the signal.

To apply the diffusion model to response signal data, Ratcliff (1988, 2006) assumed that there are response criteria just as for the standard two-choice task, and at some signal lag, responses come from a mixture of processes, those that have terminated at one or the other of the boundaries and those that have not. This is in accord with subjects' intuitions that, at the long lags, the decision has already been made, the response has been chosen, and the subject is simply waiting for the signal. As the time between stimulus and signal decreases, a larger and larger proportion of processes will have failed to terminate. Differences among experimental conditions of different difficulties appear as differences in the proportions of accumulated information at the different lags. At the longest lags (2 or more seconds), all or almost all processes will have terminated. For nonterminated processes, there are two possibilities: that decisions are made on the basis of the partial information that has already been accumulated (Figure 3.6 top panel) or that they are simply guesses (Figure 3.6 middle panel). Ratcliff (2006) tested between these possibilities with a numerosity discrimination experiment (subjects decide whether the number of asterisks displayed on a PC monitor is greater than or less than 50). The same subjects participated in the response signal task and the standard task and examples of the response signal data and model fits are shown in Figure 3.7. When the model

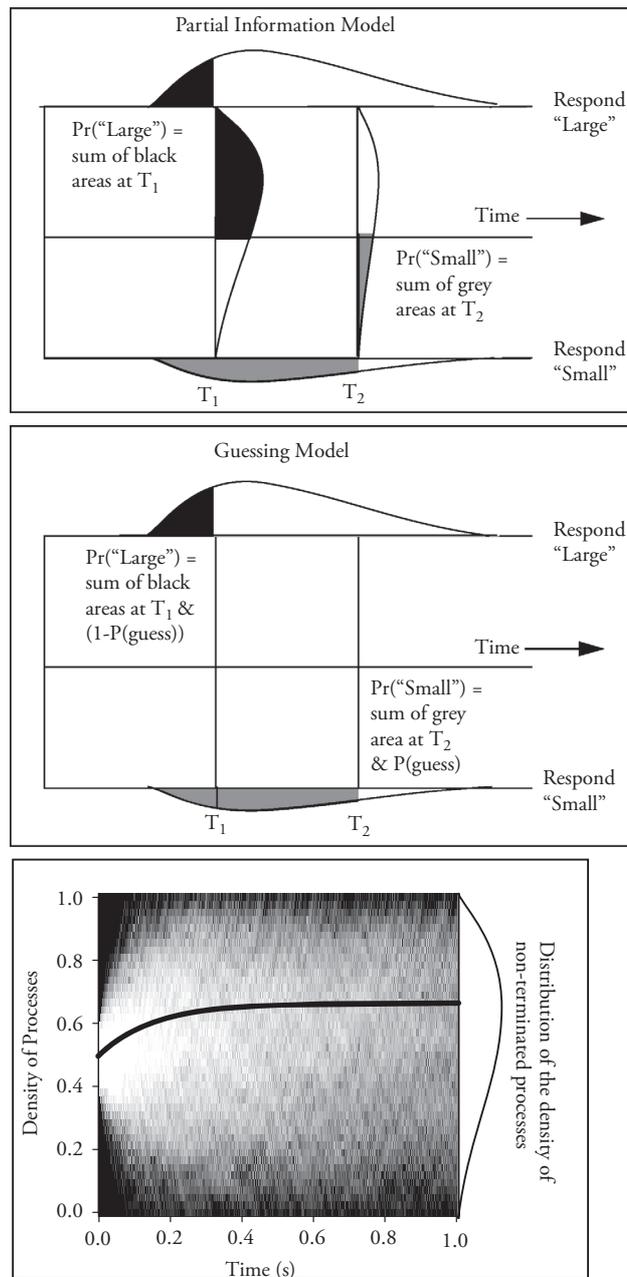


Fig. 3.6 The top two panels show two models for how the diffusion model accounts for response signal data. In the top panel, the proportion of “large” responses at time T_1 is the sum of processes that have terminated at the “large” boundary (the black area above the boundary) and nonterminated processes (the black area still within the diffusion process), i.e., partial information. The middle panel shows the same assumption as the top panel except that if a process has not terminated, a guess is used instead of partial information. The bottom panel shows a heat map of simulated paths for the diffusion model. White corresponds to high path density and black to a low path density. For the diffusion model, the distribution to the right corresponds to the asymptotic distribution of path positions after about 0.2 seconds (i.e., the vertically oriented distributions in the top panel).

was fit to the two sets of data simultaneously, it fit well and it fit equally well for the two possibilities for nonterminated processes. In other words, “guessing and partial information models could not be discriminated.

MEYER, IRWIN, OSMAN, & KOUNIOS, (1988) PARTIAL INFORMATION PARADIGM

This paradigm used a variant of the response signal task in which, on each trial, subjects responded either in the regular way unless a signal

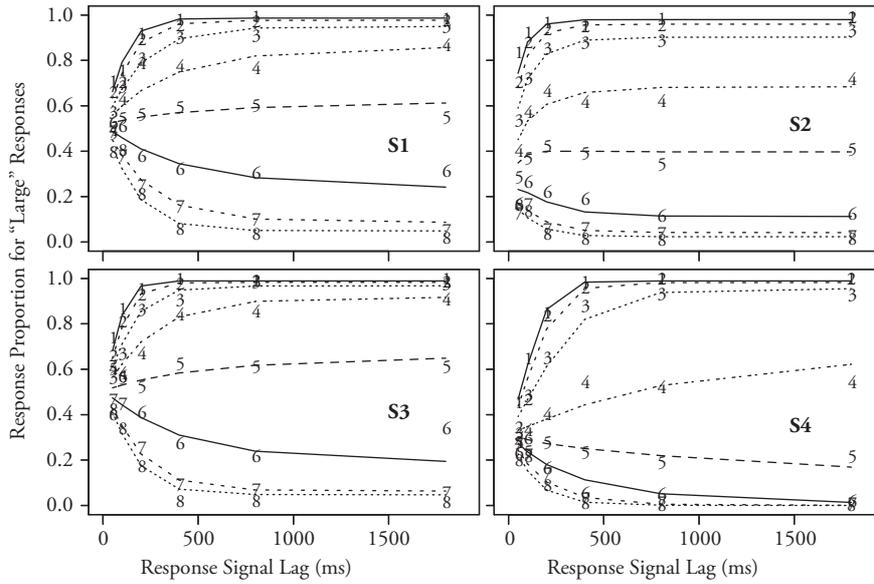


Fig. 3.7 Plots of response proportion as a function of response signal lag from a numerosity discrimination experiment (Ratcliff, 2006) for four subjects. The task required subjects to judge whether the number of dots in a 10x10 array was greater than 50 or less or equal to 50. The digits 1–8 (in reverse order) and the eight lines represent eight groupings of numbers of dots (e.g., 13–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80, and 81–87 dots).

to respond occurred, in which case they were to respond immediately. Thus, any trial could be a signal trial or a regular trial. Meyer et al. developed a method based on a race model that decomposes accuracy on the signal trials (at each signal lag) into a component from fast finishing regular trials and a component based on partial information. Results showed that partial information, in some tasks (see also Kounios, Osman, & Meyer, 1987) grew quickly and leveled off at about one-third the accuracy level of regular processes.

Ratcliff (1988) examined the predictions of the diffusion model with the assumption that decisions on signal trials were a mixture of processes that terminated at a boundary and processes based on position in the decision process, that is, partial information. Therefore, if a process was above the starting point (i.e., the black area in the vertical distribution in the top panel of Figure 3.6), the decision corresponded to the choice at the upper boundary.

Figure 3.6 bottom panel shows a heat map of the evolution of simulated diffusion processes. The map shows the density of processes as they begin at the starting point and spread out to the boundaries. The hotter the color (whiter), the more processes in that region. As time goes by, the color becomes

cooler because there are fewer and fewer processes that have not terminated. As in the top panel, the evolution of paths moves the mean position (the thick black line) from the starting point at 0.5 to a point a little above 0.6 by about 0.2 s. This produces an almost stationary distribution (the distribution to the right of the heat map), which gradually collapses over time (the two vertical distributions in the top panel of Figure 3.6).

For the case in which partial information is used in the decision, the expression for the distribution of the positions x of decision processes at time t is given by:

$$p(x, t) = e^{\frac{v(x-z)}{s^2}} \sum_{n=1}^{\infty} \frac{2}{a} \sin\left(\frac{n\pi z}{a}\right) \times \sin\left(\frac{n\pi x}{a}\right) e^{-\frac{1}{2}\left(\frac{v^2}{s^2} + \frac{n^2\pi^2 s^2}{a^2}\right)t} \quad (4)$$

where s^2 is the diffusion coefficient, z is the starting point, a is the separation between the boundaries, and v is the drift rate. For model fitting, the expression in Equation 4 must be integrated over the normal distribution of drift rates and the uniform distribution of starting points to include variability in drift rate and starting point across trials. This can be accomplished with numerical integration using Gaussian quadrature. The series in

Equation 4 must be summed until it converges; this means that terms have to be added until subsequent terms become so small that they do not affect the total (i.e., the series has converged to within some criterion, e.g., 10^{-5}). Then, to obtain the probability of choosing each response alternative, the proportion of processes between 0 and $a/2$ (for the negative alternative) and between $a/2$ and a (for the positive alternative) is calculated by integrating the expression for the density over position.

TIME-VARYING PROCESSING

Ratcliff (1980) examined two cases in which drift rate changes across the time course of processing. For one, drift rate changes discretely at one fixed time. Because there is an explicit expression for the distribution of evidence at that time, this distribution can be used as a starting distribution for a second diffusion process. If the time at which evidence changes is not a fixed time but has a distribution over time, this can be integrated over. This allows both response signal and regular RT tasks to be modeled. For another case, boundaries are removed completely and drift rate and the drift coefficient varied continuously over time. Only the first case has been used in modeling response signal data (as in Ratcliff, 1988, 2006).

GO/NOGO TASK

In the go/no go task, subjects are told to respond for one of the two choices but to make no response for the other choice. Withholding responses for one of the choices is similar to the response signal task in which responses must be held until the signal. Gomez, Ratcliff, and Perea (2007) proposed that there are two response boundaries for the go/no go task just as for the standard task, but subjects made a response only when accumulated evidence reaches the “go” boundary. Gomez et al. successfully fit the model simultaneously to data from the standard task and data from the go/no go task. They also tested a variant for which there was only one boundary, the “go” boundary, but this variant could not fit the data well.

Application of the diffusion model simultaneously to the standard task and response signal task or to the standard task and go/no go tasks places powerful constraints on the model and, when it is successful, it offers new insights into the cognitive processes involved in these tasks. It also provides theoretical convergence between the three tasks, with two boundaries for all three tasks and withheld responses for the latter two.

The first conclusion is that applying models to multiple tasks simultaneously produces strong constraints on models that (if they successfully account for data) lead to new understanding of how the tasks are performed. In the context of the sequential sampling models discussed in this article, this approach yielded a new view of response signal performance: responses increase in accuracy over time mainly because the proportion of terminated processes increases and the increase in accuracy does not come entirely from the increasing availability of partial information. Moreover, versions of the models that provide quite good fits to the data from the standard RT and response signal tasks individually could not account for both sets of data simultaneously with parameters that were consistent across tasks.

Optimality

In animal studies, performance has been described in terms of how close it comes to maximizing reward rate. This is part of a larger theme in neuroscience, which reprises the classical signal detection and sequential-sampling literatures, in which reward rate is used as a criterion for understanding whether neural computations approach optimality. For animals, how close performance is to optimal in terms of reward rate is a reasonable question to ask because animals are deprived of water or food and their overwhelming desire is to obtain them. Also they are trained for many sessions and so there is ample opportunity to optimize reward. However, when this kind of optimality is translated to human studies, the a priori reasonableness comes into question. This is because humans do not aim to get the most correct per unit time. Instead, they aim to get the most correct in the available time. If a student takes a 2-hour exam and obtains 60% correct in 1 hour, but another student gets 80% correct in 2 hours, the first has more correct per unit time, but the second would be more likely to pass the course.

Bogacz, Brown, Moehlis, Holmes, & Cohen et al. (2006) performed extensive analyses of optimality and set the stage for analyses of data. They showed that optimality as defined by reward rate can be adjusted by changing boundary settings. If the boundaries are too far apart, subjects are accurate, but slow and so there are few correct per unit of time. If boundaries are too narrow, RT is short but accuracy is low and there are few correct responses per unit of time. Thus, there is a boundary setting that maximizes the number

correct per unit of time and it is possible to test whether subjects set criteria near to this value.

Starns and Ratcliff (2012) tested undergraduate subjects on a simple numerosity discrimination task in which different groups of subjects were tested at different levels of difficulty. They were tested in blocks of trials that had a fixed total duration for which they were instructed to get as many correct in the time allowed and in blocks of trials in which the number of trials was the same no matter how fast they went. Reward-rate optimality predicts that when difficulty increases, subjects should speed up and sacrifice accuracy per unit time. Results showed subjects did the opposite, slowing down with increases in difficulty. This is the result we might expect from years of academic training to spend more time on difficult problems.

Starns and Ratcliff (2010) analyzed several published data sets with young and older adults and found that young adults with accuracy feedback sometimes approached reward-rate optimality. But older adults rarely moved more than a few percent away from asymptotic accuracy. Young adults in the context of psychology experiments (or perhaps practice with video games, some of which promote speed) will sometimes be able to optimize performance in terms of number correct per unit of time. In general, however, concerns about accuracy that have been trained for years appear to dominate.

Domains of Application

One criterion for how well a model performs is whether it simply reiterates what is already known from traditional analyses. Here we describe a number of applications, some of which provide new insights into processing, individual differences and differences among subject groups are obtained. But in other cases, the obvious results are obtained. Even when obvious results are obtained, the model integrates the three dependent variables, namely, accuracy and correct and error RT distributions, into a common theoretical framework that provides explanations of data that many hypothesis-testing approaches do not. Hypothesis-testing approaches usually select only accuracy or only mean RT as the dependent variable. In some cases, the two variables tell the same empirical story, but in other cases, they are inconsistent. The model based approach helps to resolve such inconsistencies.

Perceptual Tasks

Recently diffusion models have been applied to psychophysical discrimination tasks in which

stimuli are presented very briefly, often at low levels of contrast, sometimes with backward masks to limit iconic persistence. The focus has been to understand the perceptual processes involved in the computation of drift rates. Psychophysical paradigms have historically been used mainly with threshold or accuracy measures but recent studies have collected accuracy and RT data.

Ratcliff and Rouder (2000) and Smith, Ratcliff, and Wolfgang (2004) found that the diffusion model provided a good account of accuracy and distributions of RT from tasks with brief backward-masked stimuli. They compared the model with a constant drift rate from starting point to boundaries to the model with varying drift rate. Drift rates might be thought to decrease over time if they either tracked stimulus information or were governed by a decaying perceptual trace. However, there was no evidence in either study of increased skewness in the RT distributions or very slow error RTs at low levels of stimulus discriminability, as would have been expected if the decision process had been driven by a decaying perceptual trace. Instead, it appears that the information that drives the decision is relatively durable.

The standard application of the model assumes that, at some point in time after stimulus encoding, the decision process turns on, and evidence is accumulated toward a decision. This time is assumed to be the same across conditions and drift rate is assumed to be at a constant values from the point the process turns on. The assumption of a constant drift rate could be relaxed: Ratcliff (2002) generated predicted accuracy and RT quantiles for several conditions under the assumption that drift rate ramped up from zero to a constant level over 50 ms. He fit the standard model to these predicted values and found that the model fit well with nondecision time increased by 25 ms and with starting point, and nondecision time variability increased. Thus, a ramped onset of drift rate over a small time range will be indistinguishable from an abrupt onset.

Smith and Ratcliff (2009) developed a model, the integrated system model, that is a continuous-flow model comprised of perceptual, memory, and decision processes operating in cascade. The perceptual encoding processes are linear filters (Watson, 1986) and the transient outputs of the filters are encoded in a durable form in visual short-term memory (VSTM), which is under the control of spatial attention. The strength of the VSTM trace determines the drift rate for the diffusion

process and the moment-to-moment variations in trace strength act as a source of noise in the decision process. Because the VSTM trace in the model increases over time (i.e., drift rate is time varying), predictions for the model are obtained using the integral equation methods described previously (Smith, 2000). The model has successfully accounted for accuracy and RT distributions in tasks with brief backward-masked stimuli.

The main area of application of the integrated system model has been to tasks in which spatial attention is manipulated by spatial cues. In many cuing tasks, in which a single well-localized stimulus is presented in an otherwise empty display, attention shortens RT but increases accuracy only when stimuli are masked (Smith, Ratcliff, & Wolfgang, 2004; Smith, Ellis, Sewell, & Wolfgang, 2010). The model assumes that attention increases the efficiency with which perceptual information is transferred to VSTM and that masks interrupt the process of VSTM trace formation before it is complete. These two processes interact to produce a cuing effect in accuracy only when stimuli are masked but an unconditional effect in RT. The model has successfully accounted for the distributions of RT and accuracy in attention tasks in which the timing of stimulus localization is manipulated via onset transients and localizing markers (Sewell & Smith, 2012). These studies have helped illuminate the way in which performance is determined by perceptual, memory, attention, and decision processes acting in concert.

Diederich and Busemeyer (2006) also considered the effects of attention on decision-making in a diffusion-process framework, studying decisions about multi-attribute stimuli for which it is plausible that people shift their attention sequentially from one attribute of a stimulus to the next. They assumed that some attributes would provide more information than others and modeled this successfully as a sequence of step changes in drift rate during the course of a trial.

Recognition Memory

One of the early applications of the diffusion model was to recognition memory. In global memory models, a test item is matched against all memory in parallel, and the output is a single value of strength or familiarity (Gillund & Shiffrin, 1984; Hintzman, 1986; Murdock, 1982, and later, Dennis & Humphreys, 2001; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). From this point of view, the diffusion model provides

a meeting point between the decision process and memory, specifically, the drift rate for a test item represents the degree of match between a test item and memory.

In signal detection approaches to recognition memory, there has been considerable interest in the relative standard deviations (SDs) in strength between old and new test items, typically measured by confidence judgement paradigms. The common finding is that z -ROC functions (i.e., z -score transformed receiver operating characteristics) are approximately linear with a slope less than 1 (e.g., Ratcliff, Sheu, & Gronlund, 1992). There have been two interpretations of this finding. One is a single-process model that assumes the SD of memory strength is normally distributed, but the SD for old items is larger than that for new items. The other is a dual-process model in which the familiarity of old and new items comes from normal distributions with equal SDs but there is an additional recollection process (e.g., Yonelinas, 1997).

In fits of the diffusion model to recognition memory data, it has been usually assumed that the SD in drift rate across trials is the same for studied and new items. Starns and Ratcliff (2014) performed an analysis of existing data sets that allowed the across-trial variability in drift rate to be different for studied and new items. They found that the across-trial variability in drift rate was larger (in about 66% of the cases for individual subjects) for studied items than for new items. It also turned out that the interpretations of the other model parameters did not change when variability was allowed to differ. The advantage of this analysis is that the relative variability of studied and new items was able to be determined from two-choice data and did not require confidence judgments.

Lexical Decision

Much like recognition memory, a test item for lexical decision is matched against memory. The output is a value of how “wordlike” the item is. For sequential sampling models, proposals about how lexical items are accessed in memory must provide output values that, when mapped through a sequential sampling model, produce RTs and accuracy that fit data (Ratcliff, Gomez, & McKoon, 2004). (Note that there are other models that have integrated RT and accuracy with lexical processes, in particular, Norris, 2006).

Often, lexical decision response time (RT) has been interpreted as a direct measure of the speed

with which a word can be accessed in the lexicon. For example, some researchers have argued that the well-known effect of word frequency—shorter RTs for higher frequency words—demonstrates the greater accessibility of high frequency words (e.g., their order in a serial search, Forster, 1976; the resting levels of activation in units representing the words in a parallel processing system, Morton, 1969). However, other researchers have argued, as we do here, against a direct mapping from RT to accessibility. For example, Balota and Chumbley (1984) suggested that the effect of word frequency might be a by-product of the nature of the task itself, and not a manifestation of accessibility. In the research presented here, the diffusion model makes explicit how such a by-product might come about.

Semantic and Recognition Priming Effects

For semantic priming, the task is usually a lexical decision. A target word is immediately preceded in a test list either by a word related to it (e.g., cat dog) or some other word (e.g. table dog). For recognition priming, the task is old/new recognition and a target word is immediately preceded by a word that was studied near to it in the list of items to be remembered or far from it. In the diffusion model, the simplest assumption about priming effects is that they result from higher drift rates for primed than unprimed items.

It has been hypothesized that the difference in drift rates between primed and unprimed items arises from the familiarity of compound cues to memory (McKoon & Ratcliff, 1992; McNamara, 1992, 1994; Ratcliff & McKoon, 1988, 1994). The compound cue for an item is a multiplicative combination of the familiarity of the target word and the familiarity of the prime (see examples in Ratcliff & McKoon, 1988). If the prime and target words are related in memory, the combination produces a higher value of the joint familiarity than if they were not related. For primed items, the prime and target share associates in memory, the joint familiarity would be higher than if the prime and target do not share associates. This model was capable of explaining a number of phenomena in research on priming including the range of priming, the decay in priming, the onset of priming, and so on.

McKoon and Ratcliff (2012) compared priming in word recognition to associative recognition. Subjects studied pairs of words and then performed either a single-word recognition task or

an associative recognition task (see also Ratcliff, Thapar, & McKoon, 2011). For the associative recognition task, subjects decided whether two words of a test pair had or had not appeared in the same pair at study. In the single-word task, some test words were immediately preceded in the test list by the other word of their studied pair (primed) and some by a word from a different pair (unprimed). Data from the two tasks were fit with the diffusion model and the results showed parallel behavior: the drift rates for associative recognition and those for priming were parallel across ages and IQ, indicating that they are based, at least to some degree, on the same information in memory.

Value-Based Judgments

Busemeyer and Townsend (1993) developed a diffusion model called decision field theory to explain choices and decision times for decisions under uncertainty, and later Roe, Busemeyer, Townsend (2001) extended it to multi-alternative and multi-attribute situations. According to the theory, at each moment in time, options are compared in terms of advantages/ disadvantages with respect to an attribute, these evaluations are accumulated across time until a threshold is reached, and the first option to cross the threshold determines the choice that is made. The theory accounts for a number of findings that seem paradoxical from the perspective of rational choice theory. Usher and McClelland (2004) proposed another diffusion model to account for a similar range of findings.

Milosavljevic, Malmaud, Huth, Koch, & Rangel (2010) examined several variants of diffusion models for value-based judgments. They found that the standard model with across-trial variability in model parameters provided a good account of data from their paradigm. More recently, Krajbich and Rangel (2011) have used a model similar in character to decision field theory. They examined value-based judgments for food items and had subjects choose which of two alternatives they preferred. They monitored eye fixations and in modeling, they assumed evidence was accumulated at a higher rate for the alternative fixated. Their model accounted for RTs and accuracy and for the influence of which of the two choices was fixated and for how long.

Philiastides and Ratcliff (2013) examined value-based judgments of consumer choices with brand names presented on some trials as well as the items for which the choices were made. When the quality of the brand name was in conflict with the perceived quality of the item, the probability of choosing the

item was lower than when they were consistent. Application of the diffusion model showed that the effect of the brand was to alter drift rate but none of the other parameters of the model. This means that the value of the stimulus and brand name were processed as a whole.

Currently, there is a growing interest in the application of diffusion models to decision-making in marketing and economics, including neuroeconomics. Wide application of diffusion models in this domain are in their infancy, but the potential for theoretical advancement is great, as is demonstrated by these examples.

Aging

The application of the diffusion model to studies of aging has been especially successful, producing a different view of the effects of aging on cognition than has been usual in aging research. The general finding in the literature has been that older adults are slower than young adults (but not necessarily less accurate) on most tasks, and this has been interpreted as a decline with age in all or almost all cognitive processes. However, application of the diffusion model showed that this is not correct (Ratcliff, Thapar, & McKoon, 2003, 2004, 2006, 2007; Ratcliff, Thapar, Gomez, & McKoon, 2004). For example, Ratcliff, Thapar, and McKoon (2010) tested old and young adults on numerosity discrimination, lexical decision, and recognition memory. What they found is that older adults had slower nondecision times and set their boundaries wider, but their drift rates were not lower than those of young adults. In contrast, in some tasks (associative recognition and letter discrimination), large declines in drift rate with age have been found (Ratcliff et al., 2011; Thapar et al., 2003).

Individual Differences

The diffusion model has been used to examine individual differences. To do so requires that the SDs in model parameters from estimation variability are smaller than the SDs between subjects. In the aging studies described earlier, with about 45 minutes of data collection, individual differences in drift rates, boundary settings, and nondecision time were three to five times larger than the SDs of the model parameters. (See Ratcliff & Tuerlinckx, 2002, for tables of SDs in model parameters).

Schmiedek, Oberauer, Wilhelm, Suß, & Wittmann (2007) analyzed data from eight choice-RT tasks (including verbal, numerical, and

spatial tasks) from Oberauer, Suß, Wilhelm, and Wittmann (2003). They found that drift rates in the diffusion model mapped onto working memory, speed of processing, and reasoning ability measures (each of these was measured by aggregated performance on several tasks).

In aging studies by Ratcliff et al. (2010, 2011), IQs ranged from about 80 to about 140. Applying the model showed that drift rate varied with IQ (by as much as 2:1 for high versus lower IQ subjects) but boundary separation and nondecision time did not. This is the opposite of the pattern for aging. This dissociation provides strong support for the model because it extracts regularity from the three dependent variables (accuracy and correct and error RT distributions).

Individual differences across tasks in model parameters provide strong evidence for common abilities across tasks. In the Ratcliff et al. (2010) study, in the lexical decision, item recognition, and associative recognition tasks, there are strong correlations across subjects in drift rate, and these correlated with IQ as measured by WAIS vocabulary and matrix reasoning. Also, boundary separation correlates across tasks as did nondecision time. These results show that the diffusion model extracts components of processing that show systematic individual differences across tasks.

Consistent boundary setting across tasks are of special interest because boundary settings are optional, because they can be easily changed by instruction (e.g., go fast or be accurate). In most real-life situations, we rarely encounter more than single decisions on a particular stimulus class (except perhaps at Las Vegas or in psychology experiments). This means that there is little chance of adjusting decision criteria in real life because there is little extended experience with a task in which the decision maker can extract statistics from a long sequence of trials in which the structure of the trials do not change. The diffusion model assumes that a decision maker uses this decision mechanism across many tasks, and so we would expect to see correlations in boundary separation across tasks. This is a result that has been obtained whenever the comparison has been made.

Child Development

A natural extension from the aging studies is to test children on similar tasks to those performed with older adults to trace the course of development within the model framework. Ratcliff, Love,

Thompson, and Opfer (2012) tested several groups of children on a numerosity discrimination task and a lexical decision task. The results showed that relative to college age subjects, children's drift rates were lower, boundary separation was larger, and nondecision time was longer. These differences were larger for younger relative to older children. In other laboratories, drift rates have been found to be lower for ADHD and dyslexic children relative to normal controls (ADHD, Mulder et al., 2010; dyslexia, Zeguers et al., 2011). These studies show that the diffusion model can be applied to data collected from children, a domain in which there has been relatively little research with decision models.

Clinical Applications

In research on psychopathology and clinical populations, two-choice tasks are commonly used to investigate processing differences between patients and healthy controls. For highly anxious individuals, it is well-established that they show enhanced processing with threat-provoking materials, but this is found reliably only when there are two or more stimuli competing for processing resources, not one. However, when White, Ratcliff, Vasey, & McKoon (2010) applied the diffusion model to the RT and accuracy data from two-choice lexical decision task with single words that included threatening and control words, they found a consistent processing advantage for threatening words in high-anxious individuals, whereas traditional comparisons showed no significant differences. Because the diffusion model makes use of both RT and accuracy data, it has more power to detect differences among subject populations than simply RT or accuracy alone.

Studies of depression have had somewhat different patterns of results. Depressive symptoms are more closely linked with abnormal emotional processing with a negative emotional bias in clinical depression, even-handedness (i.e., no emotional bias) in dysphoria, and a positive emotional bias in nondepressed individuals. However, item recognition and lexical decision tasks often fail to produce significant results. White, Ratcliff, Vasey, & McKoon (2009) used the diffusion model to examine emotional processing in dysphoric (i.e., moderately high levels of depressive symptoms) and nondysphoric college students to examine differences in memory and lexical processing of positive and negative emotional words (which were presented among many neutral filler words). They found positive emotional bias in nondysphoric

subjects and even-handedness in dysphoric subjects in drift rates. As before, this pattern was not apparent with comparisons of reaction times or accuracy, consistent with previous null findings.

One limitation of these studies and similar ones is that there may be relatively few materials with the right kinds of properties or structures (also in language processing experiments for example). The emotional word pools for the experiments only contained 30 words each. This left relatively few observations (especially for errors) to use in fitting the diffusion model, which would result in unreliable parameter estimates. To remedy this, the model was fit to all conditions simultaneously, including the neutral filler conditions which had hundreds of observations. The only parameter that was allowed to vary between the conditions was drift rate. Estimates for the other parameters (e.g., nondecision time and boundary separation) largely determined by the filler conditions because the fitting method essentially weighted estimation of the parameters common to all conditions by the number of observations for each condition. Thus, the filler conditions largely determined all model parameters except the drift rates for the critical conditions, resulting in an increase in power. The results showed a bias for positive emotional words in the nondysphoric participants, but not in the dysphoric participants (White et al., 2009).

This difference in emotional bias was not significant when the diffusion model was fit only to the emotional conditions with few observations, nor was it significant in comparisons of mean RT or accuracy.

Another study examined the effects of aphasia in a lexical decision task. The impairments produce the exaggerated lexical decision reaction times typical of neurolinguistic patients. In diffusion model analyses, decision and nondecision processes were compromised, but the quality of the information upon which the decisions were based did not differ much from that of unimpaired subjects (Ratcliff, Perea, Colangelo, & Buchanan, 2004).

Manipulations of Homeostatic State

Ratcliff and Van Dongen (2009) looked at effects of sleep deprivation with a numerosity discrimination task, van Ravenzwaaij, Dutilh, and Wagenmakers (2012) looked at the effects of alcohol consumption with a lexical decision task, and Geddes et al. (2010) looked at the effects of reduced blood sugar with a numerosity

discrimination task. Applying the model to all of these studies, the main effect was a reduced drift rate but with either small or no effect on boundary separation and nondecision time.

These results show that the diffusion model is useful in providing interpretations of group differences among different subject populations. Furthermore, as noted earlier, the model can be used to examine individual differences (even with only 45 minutes of data collection for a task). This means that this modeling approach, when paired with the right tasks, may have a useful role to play in neuropsychological assessment.

Situations in Which the Standard Model Fails

There are several cases in which the standard diffusion model fails to account for experimental data. These fall into two classes: one involves dynamic noise and categorical stimuli and the other involves conflict paradigms. For both, the main way the model fails is that there are cases for which the onset of the RT distribution (i.e., the leading edges) for one condition is delayed relative to the onset for other conditions.

Ratcliff and Smith (2010) and Smith, Ratcliff, & Sewell (2014) tested letter discrimination, horizontal versus vertical bars discrimination, and Gabor patch orientation discrimination with stimuli degraded with either static noise or with dynamic noise. Noise was implemented by reversing the contrast polarity of some proportion of the pixels (randomly selected) for each of the letter, random bars, and Gabor patch stimuli. For dynamic noise, a different random sample of pixels was chosen on every frame of the display, whereas static noise used a single image with one random sample reversed. Dynamic noise and, to a lesser extent static noise, produced large shifts in the leading edges of the RT distribution. The shapes of the RT distributions were consistent with the model but increasing noise increased estimates of the nondecision time parameter T_{er} . This finding is inconsistent with the hypothesis that noise increases RTs simply by reducing the rate at which evidence accumulates in the decision process. Instead, it implies that noise delays the onset of the diffusion process.

Smith, Ratcliff, and Sewell (2014) showed that shifts in onsets can be explained by Smith and Ratcliff's (2009) integrated system model, with the assumption that noise slows the process of forming a stable perceptual representation of the stimulus. In

the integrated system model, drift rate and diffusion noise grow in proportion to one another to an asymptote. Unlike the standard model, in which the onset of evidence accumulation is abrupt, the onset of evidence accumulation in the integrated system model is gradual, controlled by the growth of diffusion noise. Smith Ratcliff & Sewell, 2014 showed that this model could explain the shifts in the onsets of RT distributions found by Ratcliff and Smith (2010).

Smith, et al. (2014) also considered a second, release-from-inhibition model, which was motivated, in part, by physiological principles. They modeled release from inhibition using an Ornstein-Uhlenbeck (OU) diffusion process with a time-varying decay coefficient. In the OU process, information accumulation is opposed by a decay term that pulls the process back toward its starting point. The larger the decay, the harder it is for the process to accumulate enough information to reach a criterion and trigger a response. In the standard OU process, decay is proportional to the distance of the process from its starting point, but does not vary with time. Smith et al. (2012) assumed that decay was time-locked to the stimulus. At the start of the trial, before a perceptual representation of the stimulus is formed, the decay term is large and the process remains near its starting point with high probability. As stimulus information becomes available, the decay term progressively decreases, allowing information to accumulate in the same way as it does in the standard model. This model was also able to account for data like those reported by Ratcliff and Smith (2010). Because the inhibition process behaves somewhat like the standard model with variable starting point, the release-from-inhibition model was able to account for the fast errors found at high stimulus discriminability in dynamic noise tasks without the assumption of starting point variability.

Ratcliff and Frank (2012) also found shifts in the leading edges of RT distributions in a reinforcement learning conflict experiment for which the stimuli were three pairs of letters (the same three throughout the experiment). On each trial, one of the pairs of letters were presented in random order and the subject had to choose and respond to one of the letters. One of the letters of the pair was reinforced more often than the other (in this case, reinforcement was simply a "correct" or "incorrect" message). After a training phase, on a small proportion of the trials, letters from different pairs were presented together. When the two letters

were the highly reinforced members of the pairs, they were chosen nearly equally often and there was no slowing of the RT distribution. But when the letters that were reinforced with low probability were presented together, there was a delay in the leading edge of the RT distribution, an average delay of over 100 ms. This was explained in two ways, one in terms of the basal ganglia model of Frank (2006), and one in terms of the diffusion model. For the diffusion model to explain the data, a delay in the onset of the decision process could be used to produce good fits to the data. But this was, to some degree, a redescription of the empirical result. The basal ganglia model explained these conflict trials by an increase in threshold in the neural circuitry. This was linked to the diffusion model by showing that a transient increase in boundary separation was also capable of explaining the result (the delay in onset of the RT distribution). It turned out that an increase in boundary separation with an exponential decay mimics a delayed onset.

White, Ratcliff and Starns (2011) also found leading edge shifts in a flanker task. In their experiment, a target angle bracket was presented that pointed in the direction of the correct response. On conflict trials, the target bracket was embedded in a string pointing the other way. Again, RT distributions could not be explained with only a difference in drift rates, but a model with drift rate changing over the time course of the decision, starting by being dominated by the flankers and then gradually focusing on the central symbol, was successful.

All of these paradigms suggest that, in these conflict situations, drift rate is not stationary over time. It is necessary to go beyond the basic decision model and begin to integrate it with models of perceptual and cognitive processing.

Competing Two-Choice Models

The diffusion model described to this point is one of a class of sequential sampling models that share many features. They all have given the same interpretations of effects of independent variables, which are the same across the models (e.g., Donkin, Brown, Heathcote, & Wagenmakers, 2011; Ratcliff, Thapar, Smith, & McKoon, 2005). This means, for example, that the effects of aging on model components are the same, whichever model is used.

The leaky competing accumulator (LCA) model (Usher & McClelland, 2001) was developed as

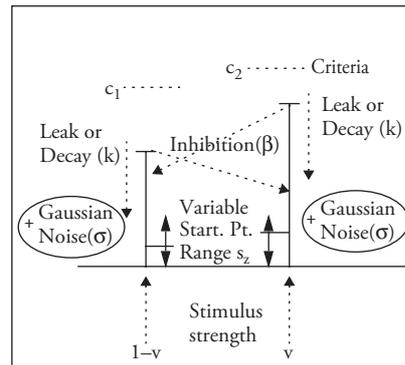


Fig. 3.8 An illustration of the leaky competing accumulator. The model includes an inhibition term ($-\beta x_j$) in which the increment to evidence in accumulator i is reduced as a function of activity in the other accumulator (x_j) and a decay term in which the increment to evidence is reduced as a function of activity in the accumulator ($-\beta x_i$). The decision criteria for the two accumulators are c_1 and c_2 , the accumulation rates are v_1 and v_2 ($v_1 + v_2 = 1$), and there is variability in the starting points that is uniformly distributed across trials with range s_T . Variability in processing within a trial is normally distributed with standard deviation σ .

an alternative to the diffusion model. Part of the motivation was to implement neurobiological principles that the authors believed should be incorporated into RT models, especially mutual inhibition mechanisms and decay of information across time.

In the LCA model, like the diffusion model, information is accumulated continuously over time. There are two accumulators, one for each response, as shown in Figure 3.8, and a response is made when the amount of information in one of the counters reaches its criterion amount. The rate of accumulation, the equivalent of drift rate in the diffusion model, is a combination of three components. The first is the input from the stimulus (v), with a different value for each experimental condition. If the input to one of the accumulators is v , the input to the other is $1-v$ so that the sum of the two rates is 1. The second component is decay in the amount of accumulated information, k , with size of decay growing as the amount of information in the accumulator grows, and the third is inhibition from the other accumulator, β , with the amount of inhibition growing as the amount of information in the other accumulator grows. If the amount of inhibition is large, the model exhibits features similar to the diffusion model because an increase in accumulated information for one of the response choices produces a decrease for the other choice.

Just as in the diffusion model, the accumulation of information is assumed to be variable over the course of a trial, with a normal distribution with standard deviation σ . Because of the decay and inhibition in the accumulation rates, the tails of RT distributions are longer than they would be if produced without these factors (cf., Smith & Vickers, 1988; Vickers, 1970, 1979; Vickers, Caudrey, Willson, 1971), which leads to good matches with the skewed shape of empirical distributions.

The expression for the change in the amount of accumulated information at time t in counter i , is:

$$\Delta x_i = \left(v_i - kx_i - \sum_{j \neq i} \beta x_j \right) \Delta t + \sigma \eta_i \sqrt{\Delta t} \quad i = 1, 2 \quad (5)$$

The amount of accumulated information is not allowed to take on values below zero, so if it is computed to be below zero, it is reset to zero. This is theoretically equivalent to constraining the diffusion process with a reflecting boundary at zero.

The LCA model without cross-trial variability in any of its components predicts errors slower than correct responses. To produce errors faster than correct responses and the crossover pattern such that errors are faster than correct responses for easy conditions and slower for difficult conditions, Usher and McClelland assumed variability in the accumulators' starting points, just as is assumed in the diffusion model and by Laming (1968).

In the diffusion model, moving a boundary position is equivalent to moving the starting point. Moving the starting point an amount y toward one boundary is the same as moving that boundary an amount y toward the starting point and the other boundary an amount y away from the starting point. In the LCA model, changing the starting point is not equivalent to changing a boundary position because decay is a function of the distance of the accumulated amount of evidence from zero. Increasing the starting point by an amount y increases decay by an amount proportional to y , but with the starting point at zero, reducing the boundary by y has no effect on decay. Usher and McClelland (2001) implemented variability in starting point by assuming rectangular distributions of the starting points with minimums at zero.

No explicit solution is known for the pair of coupled equations in Eq. 5, when they are constrained by decision criteria and the requirement that the

accumulated information remain positive. Thus, as in Usher and McClelland (2001), predictions from the model are obtained by simulation. There have been several analyses of this model. Bogacz et al. (2006) showed that the model could be reduced to a single diffusion process if leak and inhibition were balanced and examined notions of optimality (but see van Ravenzwaaij, van der Maas, & Wagenmakers, 2012).

The Linear Ballistic Accumulator (LBA, Brown and Heathcote, 2008) is similar to the LCA in that it uses two accumulators, but it has no within-trial variability, no decay, and no inhibition. The model assumes that the rate of evidence accumulation and the starting point for accumulation both vary randomly from trial to trial, but that the process of evidence accumulation itself is noise free. In essence, the model assumes that there is noise in the central nervous system on long, between-trial, time scales, but none on short, moment-to-moment, time scales that govern evidence accumulation within a trial. This assumption appears incompatible with the single-cell recording literature that has linked processes of evidence accumulation with neural firing rates in the oculomotor control system, because such neural spike trains are typically noisy. To reconcile these kinds of data with noiseless evidence accumulation requires an argument to the effect that individual neurons are noisy but the neural ensemble as a whole is effectively noise free. However, it is not clear that firing rates in weakly coupled networks of neurons exhibit the kinds of central limit theorem type properties that this argument requires (Zohary, Shadlen, & Newsome, 1994), and so the status central limit argument is unclear.

Multichoice Decision-Making and Confidence Judgments

Recently, interest in the neuroscience domain in multichoice decision-making tasks has developed for visual search (Basso & Wurtz, 1998; Purcell et al., 2010) and motion discrimination (Niwa & Ditterich, 2008; Ditterich, 2010). In psychology, there have been investigations using generalizations of standard two-choice tasks (Leite & Ratcliff, 2010) and in absolute identification (Brown, Marley, Donkin, & Heathcote, 2008). In addition, confidence judgments in decision-making and memory tasks are multichoice decisions, and diffusion models are being applied in these domains (Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013; Van Zandt, 2002).

It is clear that there is no simple way to extend the two-choice model to tasks with three or more choices. But models with racing accumulators can be extended. Some models with racing accumulators become standard diffusion models when the number of choices is reduced to two.

Ratcliff and Starns (2013) proposed a model for confidence judgments in recognition memory tasks that uses a multiple-choice diffusion decision process with separate accumulators of evidence for each confidence choice. The accumulator that first reaches its decision boundary determines which choice is made. Ratcliff and Starns compared five algorithms for accumulating evidence and found that one of them produced choice proportions and full RT distributions for each choice that closely matched empirical data. With this algorithm, an increase in the evidence in one accumulator is accompanied by a decrease in the others with the total amount of evidence in the system being constant.

Application of the model to the data from an earlier experiment (Ratcliff, McKoon, & Tindall, 1994) uncovered a relationship between the shapes of z -ROC functions and the behavior of RT distributions. For low-proportion choices, the RT distributions were shifted by as much as several hundred milliseconds relative to high proportion choices. This behavior and the shapes of z -ROC functions were both explained in the model by the behavior of the decision boundaries.

For generality, Ratcliff and Starns (2013) also applied the decision model to a three-choice motion discrimination task in which one of the alternatives was the correct choice on only a low proportion of trials. As for the confidence judgment data, the RT distribution for the low probability alternative was shifted relative to the higher probability alternatives. The diffusion model with constant evidence accounted for the shift in the RT distribution better than a competing class of models.

Research on multichoice decision making, including confidence judgments, is a growing industry but the constraints provided by RT distributions and response proportions for the different choices makes the modeling quite challenging.

One-Choice Decisions

Relatively little work has been done recently on one-choice decisions. In these, there is only one key to press when a stimulus is detected. Ratcliff and Van Dongen (2011) tested a model that used a single diffusion process to represent the process of accumulating evidence. The main application was

to the psychomotor vigilance task (PVT) for which a millisecond timer is displayed on a computer screen and it starts counting up at intervals between 2 and 12 s after the subject's last response. The subject's task is to hit a key as quickly as possible to stop the timer. When the key is pressed, the counter is stopped, and the RT in milliseconds is displayed for 1 s.

In single-choice decision-making tasks, the data are a distribution of RTs for hitting the response key. The one-choice diffusion model assumes the evidence begins accumulating on presentation of a stimulus until a decision criterion is hit, upon which, a response is initiated (Figure 3.9 illustrates the model). In the model, drift rate is assumed to vary from trial to trial. This relates it to the standard two-choice model, which makes this assumption to fit the relative speeds of correct and error responses. In application of the one-choice model to sleep deprivation data, across-trial variability in drift rate was needed to produce the long tails observed in the RT distributions.

Ratcliff and Van Dongen (2011) fit the model to RT distributions and their hazard functions from experiments with the PVT with over 2000 observations per RT distribution per subject. With only changes in drift rate, they found that the model accounted for changes in the shape of RT distributions. In particular, changes in drift rate accounted for the change in hazard function shape moving from a high tail under no sleep deprivation to a low tail with sleep deprivation. They also fit data for which the PVT was tested every 2 hours for 36 hours of sleep deprivation and found that drift rate was closely related to an independent measure of alertness, which provides an external validation of the model.

Neuroscience

One of the major advances in understanding decision making is in neuroscience applications using single cell recording in monkeys (and rats), human neuroscience including fMRI, EEG, and MEG. All these domains have had interactions between diffusion model theory and neuroscience measures. Hanes and Schall (1996) made the first connection between theory and single cell recording data, and this was taken up in work by Shadlen and colleagues (e.g., Gold and Shadlen, 2001).

MONKEY NEUROPHYSIOLOGY

In both psychology and neuroscience, theories of decision processes have been developed that

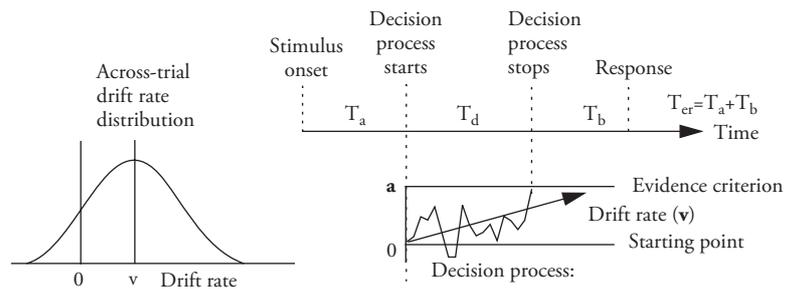


Fig. 3.9 An illustration of the one-choice diffusion model. Evidence is accumulated at a drift rate v with SD across trials η , until a decision criterion at a is reached after time T_d . Additional processing times include stimulus encoding time T_a and response output time T_b ; these sum to nondecision time T_{cr} , which has uniform variability across trials with range s_r .

assume that evidence is gradually accumulated over time (Boucher, Palmeri, Logan, & Schall, 2007; Churchland, Kiani, & Shadlen, 2008; Ditterich, 2006; Gold & Shadlen, 2001, 2007; Grinband, Hirsch, & Ferrera, 2006; Hanes & Schall, 1996; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Platt & Glimcher, 1999; Purcell et al., 2010; Ratcliff, Cherian, & Segraves, 2003; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007; Roitman & Shadlen, 2002; Shadlen & Newsome, 2001). In these studies, cells in the lateral intraparietal cortex (LIP), frontal eye field (FEF), and the superior colliculus (SC) exhibit behavior that corresponds to a gradual buildup in activity that matches the buildup in evidence in making simple perceptual decisions (see also Munoz & Wurtz, 1995; Basso & Wurtz, 1998). The neural populations that exhibit buildup behavior in LIP, FEF, and SC prior to a decision have been studied extensively. There is debate about where exactly the accumulation takes place, but it is clear that (at least) these three structures are part of a circuit that is involved in implementing the decision. These studies so far support the notion that there is a flow of information from LIP to FEF and then to SC prior to a decision.

In modeling the neurobiology of the decision process, there are a number of models applied to a range of different tasks. They all have the common theme that they assume evidence is accumulated to a decision criterion, or boundary, and that accumulated evidence corresponds to activity in populations of neurons corresponding to the decision alternatives. The models considered here have been explicitly proposed as models of oculomotor decision making in monkeys or argued to describe the evidence accumulation process in humans or monkeys. The models fall into several

classes (Ratcliff & Smith, 2004; Smith & Ratcliff, 2004), including those that assume accumulation of a single evidence quantity taking on positive and negative values (Gold & Shadlen, 2000, 2001; Ratcliff, 1978; Ratcliff et al., 2003; Ratcliff, Van Zandt, & McKoon, 1999; Smith, 2000) and those that assume that evidence is accumulated in separate accumulators corresponding to separate decisions (Churchland, et al., 2008; Ditterich, 2006; Mazurek et al., 2003; Ratcliff et al., 2007; Usher & McClelland, 2001). In this latter class of models, accumulation can be independent in separate accumulators, or it can be interactive so that as evidence grows in one accumulator, it inhibits evidence accumulation in the other accumulator. The single accumulator model can be seen as implementing perfect inhibition because a positive increment toward one boundary is an increment away from the other boundary.

The models with separate accumulators have the advantage because the two accumulators can be used to represent growth of activity in the populations of neurons corresponding to the two decisions. In the single diffusion process models, if the single process represented the aggregate activity in the two populations, then the growth of activity in the two populations would have to be perfectly negatively correlated. This is plausible if the resting activity level is relatively high in the neural populations (e.g., Roitman & Shadlen, 2002), but it is less plausible in populations in which the resting level is low (Hanes & Schall, 1996; Ratcliff et al., 2007). However, the two classes of models largely mimic each other at a behavioral level (Ratcliff, 2006; Ratcliff & Smith, 2004) and although the choice of models with racing diffusion processes seems to be superior in application in oculomotor

responses in monkeys, this does not rule out the viability of the single accumulator model for human behavioral and neural data (Philiastides, Ratcliff, & Sajda, 2006; Ratcliff et al., 2009).

Ratcliff et al. (2007; see also Ratcliff, Hasegawa, et al., 2011) applied a dual diffusion model to a brightness discrimination task. In the dual diffusion model, evidence for the two responses is accumulated by a pair of racing diffusion processes. In Ratcliff et al.'s model, there was competition at input (drift rates summed to a constant) but no inhibition (i.e., Figure 3.8 without the inhibition). Two rhesus monkeys were required to make a saccade to one of two peripheral choice targets based on the brightness of a central stimulus. Neurons in the deep layers of the SC exhibited a robust presaccadic activity when the stimulus specified a saccade toward a target within the neuron's response field, and the magnitude of this activity was unaffected by level of difficulty. Activity following brightness stimuli specifying saccades to targets outside the response field was affected by task difficulty, increasing as the task became more difficult, and this modulation correlated with performance accuracy. The model fit the full complexity of the behavioral data, accuracy and RT distributions for correct and error responses, over a range of levels of difficulty. Using the parameters from the fits to behavioral data, simulated paths of the process were generated and these provided numerical predictions for the behavior of the firing rates in SC neurons that matched most but not all the effects in the data.

Simulated paths from the model were compared to neuron activity. The assumption linking the paths to the neuron data is that firing rate is linearly related to position in the accumulation process; the nearer the boundary the decision process is, the higher the firing rate. The firing rate data show delayed availability of discriminative information for fast, intermediate, and slow decisions when activity is aligned on the stimulus and very small differences in discriminative information when activity is aligned on the saccade. The model produces exactly these patterns of results. The accumulation process is highly variable, allowing the process both to make errors, as is the case for the behavioral performance, and also to account for the firing rate results. Figure 3.10 shows sample results for the firing rate functions (black lines) and predicted firing rates (red lines).

There have also been significant modeling efforts to relate models based on spiking neurons to

diffusion models (e.g., Deco, Rolls, Albantakis, & Ramo, 2013; Roxin & Ledberg, 2008; Wong and Wang, 2006). Smith (2010) made an explicit connection between diffusion processes at a macro behavioral level and shot noise processes at a slightly abstract neural level.

Smith (2010) sought to show how diffusive information accumulation at a behavioral level could arise by aggregating neural firing rate processes. He modeled the representation of stimulus information at the neural level as the difference between excitatory and inhibitory Poisson shot noise processes. The shot noise process describes the cumulative effects of a number of time-varying disturbances or perturbations, each of which is initiated by a point event, which arrive according to a Poisson process. These discrete pulses are assumed to have exponential decay, and so, in time, some of these decaying traces add, and this is the shot noise process (e.g., Figure 3.1, Smith, 2010). In his model, the disturbances represent the flux in postsynaptic potentials in a cell population in response to a sequence of action potentials. Smith showed that the time integral of such Poisson shot-noise pairs follows an integrated Ornstein-Uhlenbeck process, whose long-time scale statistics are very similar to those assumed in the standard diffusion model. His analysis showed how diffusive information at a behavioral level could arise from Poisson-like representations at the neural level.

Subsequently Smith and McKenzie (2011) investigated a simple model of how long time scale information accumulation could be realized at a neural level. Wang (2002) previously argued that models of decision-making require information integration on a time scale that is an order of magnitude greater than any integration process found at a neural level. He argued that the most plausible substrate for such long-time scale integration is persistent activity in reverberation networks. Smith and McKenzie considered a very simple model of a recurrent loop in which spikes cycle around the loop with exponentially distributed cycle times and new spikes are added superposition. The activity in the loop could, therefore, be modeled as a superposition of Poisson processes.

They showed that a model based on such recurrent loops could realize the kind of long-time scale integration process described by Wang and that it, too, exhibited a form of diffusive information accumulation that closely matches what is found behaviorally. In particular, the resulting model successfully predicted the RT distributions and

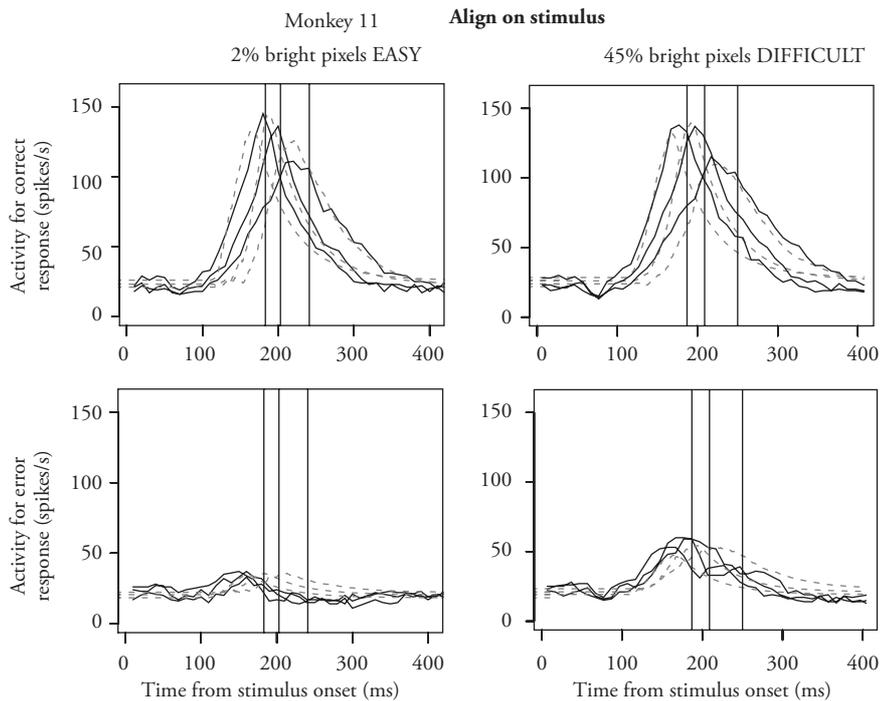


Fig. 3.10 Neural firing rates averaged over cells for firing rates aligned on the stimulus for the two monkeys from Ratcliff, Hasegawa et al. (2007). The firing rates are divided into thirds as a function of the behavioral response (fastest third, middle third, and slowest third). The left hand column shows easy conditions, bright responses to 98% white pixels and dark responses to 98% black pixels and the right hand column shows difficult conditions, bright responses to 55% white pixels and dark responses to 55% black pixels. The first row shows firing rates for cells in the receptive field of the target corresponding to the correct response and the correct response is made (target cell). The second row shows firing rates for cells in the receptive field of the target corresponding to the incorrect response for the stimulus when a correct response is made (competitor cell). The solid lines are the data and the dashed lines are model predictions.

choice probabilities from a signal detection experiment reported by Ratcliff and Smith (2004).

Human Neuroscience

Diffusion models are currently being combined with fMRI and EEG techniques to look for stimulus-independent areas that implement decision-making (e.g., vmPFC, Heekeren, Marrett, Bandettini, & Ungerleider, 2004) and to map diffusion model parameters onto EEG signals (Philiastides et al., 2006).

EEG SUPPORT FOR ACROSS-TRIAL VARIABILITY IN DRIFT RATE

Philiastides, Ratcliff, and Sajda (2006) used a face/car discrimination task with briefly presented degraded pictures. They recorded EEGs from multiple electrodes during the task and then weighted and combined the electrical signals to obtain a single number or regressor that best discriminated between faces and cars. This was repeated over 60

ms windows from stimulus onset on up. The single-trial regressor was significant at two times, around 180 ms and around 380 ms. Ratcliff, Philiastides, and Sajda (2009) reasoned that, if the regressor was an index of difficulty, then in each condition of the experiment, responses could be sorted into those that the electrical signal said were more facelike and those that were more carlike. When responses were sorted and the diffusion model fit to the two halves of each condition, the drift rates for the two halves differed substantially but only for the later component at 380 ms.

The diffusion model provides an estimate of nondetection time, which represents the duration of encoding and stimulus transformation processes prior to the decision time (as well as response output processes). This estimate shows that the decision process begins no earlier than 400 ms after stimulus onset, and so the late EEG signal component indexes difficulty on a trial-to-trial basis prior to the onset of the decision process. Therefore, these two features of the late component

provide evidence that drift rate varies from trial to trial.

EEG SUPPORT FOR ACROSS-TRIAL VARIABILITY IN STARTING POINT

Bode, Sewell, Lilburn, Forte, Smith and Stahl (2012) reported EEG evidence consistent with trial-to-trial biasing of the starting point of the diffusion process. They recorded EEG activity in a task requiring discrimination between briefly presented images of chairs or pianos that were presented in varying levels of noise and then backward masked. They applied a support vector machine pattern classifier to the EEG signals at successive time points and showed that decisions could be decoded (i.e., predicted) from the EEG several hundred milliseconds before the behavioral response. When the stimulus display contained only noise and no discriminative information, the decision outcome could still be predicted from the EEG, but only from the activity prior to stimulus presentation and not from any later time points. Bode et al. found that the RT distributions and accuracy in their task were well described by a diffusion model in which the starting point for evidence accumulation was biased toward the upper or lower boundary, depending on the participant's previous choice history. They proposed that the information in the prestimulus EEG was a neural correlate of the process of setting the starting point, which occurs prior to the start of evidence accumulation. When the display contained no stimulus information and the drift of the diffusion process was zero, the primary determinant of the decision outcome would be the participant's bias state: Processes starting near the upper boundary would be more likely to terminate at that boundary, and similarly for the lower boundary.

STRUCTURAL MRI

Studies that have examined structural connections between brain areas that are implicated in the control of decision making have found correlations between tract strength and decision-making variables. Forstmann et al. (2010) found a relationship between cortico-striatal connection strength and the ability of subjects to change their speed-accuracy tradeoff settings. Mulder, Boekel, Ratcliff, & Forstmann (2014) found correlations between subjects' ability to bias their responses in response to reward and vmPFC-STN connection

strength. These studies are the beginning of a new approach to brain structure and processing.

fMRI

A major problem with attempts to relate results from fMRI measurements to the growth of activity in decision-related brain areas is the sluggishness of the BOLD response. Despite this, there are many studies that use diffusion models in analyses of fMRI data. Mulder, Van Maanen, & Forstmann (2014) have reviewed a number of studies of perceptual decision making using fMRI methods and found evidence for regions associated with different components of diffusion models. Although there was some convergence, maps of the peak-coordinates of the activity for model components showed quite a large scatter across areas. This research would require a chapter by itself but the notion that some brain areas accumulate noisy evidence from other areas is certainly a mainstream belief in neuroscience and diffusion models are one theoretical framework that relates the neural to the behavioral level.

Conclusions

The use of diffusion models in representing simple decision-making in a variety of domains is an area of research that is seeing significant advances. The view that evidence is accumulated over time to decision criteria seems a settled view. The competing models seem to produce about the same conclusions about processing within experimental paradigms, and so broad interpretations do not depend on the specific model being used. In psychological applications, the basic theory and experimental applications are well established and somewhat mature. But application to individual differences (including neuropsychological testing) and different subject and patient populations are in their infancy. Also, neuroscience applications in both experimental and theoretical research are blossoming, with a variety of experimental methods being used as well as a variety of variants on the basic models developed in psychology.

Author Note

Preparation of this chapter was supported by grants NIA R01-AG041176, AFOSR grant FA9550-11-1-0130, IES grant R305A120189, and by ARC Discovery Grant DP140102970.

Glossary

Accumulator Model: A model in which positive increments are continuous random variable and the time at which the increments are made are discrete in time. The accumulators race to separate decision criteria.

Confidence Judgments: Tasks in which responses are made on a discrete scale using different response keys.

Decision Boundaries: These represent the amount of evidence needed to make a decision.

Decision criteria: The amount of evidence for one or other alternative to make a decision. In diffusion models, the criteria are represented as boundaries on the evidence space.

Diffusion Model: A model that assumes continuously available evidence in continuous time. Evidence accumulates in one signed sum and the process terminates when one of two decision criteria are reached.

Diffusion Process: A process in which continuously variable noisy evidence is accumulated in continuous time.

Drift rate: The average rate at which a diffusion process accumulates evidence.

Go/Nogo Tasks: Tasks in which subjects respond to one stimulus type but hold their response until a time out for the other response.

Leaky Competing Accumulator Model: A model in which evidence is continuously available in continuous time. Evidence is accumulated in separate accumulators (i.e., separate diffusion processes) and there is both decay in an accumulator and inhibition from other accumulators.

Nondecision Time: Duration of processes other than the decision process. These include encoding time, response output time, memory access time in memory tasks, and the time to transform the stimulus representation to a decision-based representation for perceptual tasks.

Optimality: Often defined in terms of “reward rate” or the number correct per unit time in simple decision making experiments by analogy with animal experiments.

Ornstein-Uhlenbeck diffusion process: This describes a noisy evidence accumulation process with leakage or decay; the standard (Wiener or Brownian motion) diffusion process describes a process in which there is no leakage.

Poisson Counter Model: A model in which increments are discrete equal-sized units, but the time at which they arrive as the accumulators are Poisson distributed (exponential delays between counts).

Poisson shot noise process: A process in which each point event in a Poisson process generates a continuous, time-varying disturbance or perturbation. The shot noise process is the cumulative sum of the perturbations. The shot noise process has been used as a model for a variety of phenomena, including the flow of electrons in vacuum tubes, the cumulative effects of earth tremors, and the flux in the postsynaptic potential in cell bodies in a neural population.

PVT: The psychomotor vigilance test in which a counter starts counting up and the subject simply hits a key to stop it counting.

Random walk model: A discrete-time counterpart of the diffusion process. A diffusion process accumulates evidence in continuous time, whereas a random walk accumulates evidence at discrete time points.

Response Signal and Deadline Tasks: Tasks in which the subject is required to respond at an experimenter-determined time. The dependent variable is usually accuracy and the task measures how it grows over time in the decision process.

Response Time Distributions: The distribution of times at which the decision process terminates (i.e., a histogram of times for data).

Single Cell Recording in Animals: Recordings from single neurons often in awake behaving animals.

References

- Balota, D. A. & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? the role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357.
- Basso, M. A. & Wurtz, R. H. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *Journal of Neuroscience*, *18*, 7519–7534.
- Bode, S., Sewell, D. K., Lilburn, S., Forte, J. D., Smith, P. L. & Stahl, J. (2012). Predicting perceptual decisions from early brain activity. *Journal of Neuroscience*, *32*, 12488–12498.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review*, *113*, 700–765.
- Boucher, L., Palmeri, T., Logan, G., & Schall, J. (2007). Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review*, *114*, 376–397.
- Brown, S. D., & Heathcote, A. J. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.
- Brown, S. D., Marley, A. A. J., Donkin, C. & Heathcote, A. J. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, *115*, 396–425.
- Buonocore, A., Giorno, V., Nobile, A. G., & Ricciardi, L. (1990). On the two-boundary first-crossing- time problem for diffusion processes. *Journal of Applied Probability*, *27*, 102–114.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.
- Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.
- Deco, G., Rolls, E. T., Albantakis, L., & Romo, R. (2013). Brain mechanisms for perceptual and reward-related decision-making. *Progress in Neurobiology*, *103*, 194–213.
- Dennis, S. & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–477.

- Diederich, A., & Busemeyer, J.R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*, 304–322.
- Diederich, A., & Busemeyer, J. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics*, *68*, 194–207.
- Ditterich, J. (2006). Computational approaches to visual decision making. In D. J. Chadwick, M. Diamond, & J. Goode (Eds.), *Percept, decision, action: Bridging the gaps* (p.114). Chichester, UK: Wiley.
- Ditterich, J. (2010). A comparison between mechanisms of multi-alternative perceptual decision making: Ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory. *Frontiers in Neuroscience*, *4*, 184.
- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models for response time, same conclusions about psychological mechanisms? *Psychonomic Bulletin & Review*, *55*, 140–151.
- Feller, W. (1968). *An introduction to probability theory and its applications*. New York, NY: Wiley.
- Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam, Netherlands: North-Holland.
- Forstmann, B. U., Anwander, A., Schafer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., Bogacz, R., & Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences*, *107*, 15916–15920.
- Frank, M.J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, *19*, 1120–1136.
- Gillund, G., & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.
- Geddes, J., Ratcliff, R., Allerhand, M., Childers, R., Wright, R. J., Frier, B. M., & Deary, I. J. (2010). Modeling the effects of hypoglycemia on a two-choice task in adult humans. *Neuropsychology*, *24*, 652–660.
- Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*, 390–394.
- Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Science*, *5*, 10–16.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.
- Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*, 347–369.
- Grinband, J., Hirsch, J., & Ferrera, V.P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron*, *49*, 757–763.
- Hanes, D. P., and Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*, 427–430.
- Heekeren, H. R., Marrett, S., Bandettini, P. A., Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*, 859–62.
- Hintzman, D. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.
- Krajbich, I., & Rangel, A. (2011). A multi-alternative drift diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*, 13852–13857.
- Kounios, J., Osman, A. M., & Meyer, D. E. (1987). Structure and process semantic memory: New evidence based on speed-accuracy decomposition. *Journal of Experimental Psychology: General*, *116*, 3–25.
- Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York: Wiley.
- Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-choice decisions. *Attention, Perception and Psychophysics*, *72*, 246–273.
- Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making*, *6*, 651–687.
- 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, NY: Oxford University Press.
- Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision-making. *Cerebral Cortex*, *13*, 1257–1269.
- McClelland, J. L. & Chappell, M. (1998). Familiarity breeds differentiation: A Bayesian approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 724–760.
- McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1155–1172.
- McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language*, *66*, 416–437.
- McNamara, T. P. (1992) Priming and constraints it places on theories of memory and retrieval. *Psychological Review*, *99*, 650–662.
- McNamara, T. P. (1994). Priming and theories of memory: A reply to Ratcliff and McKoon. *Psychological Review*, *101*, 185–187.
- Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition: mental processes inferred from a speed-accuracy decomposition technique. *Psychological Review*, *95*, 183–237.
- Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The Drift Diffusion Model can account for the accuracy and reaction times of value-based choice under high and low time pressure. *Judgment and Decision Making*, *5*, 437–449.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
- Mulder, M. J., Boekel, W., Ratcliff, R., & Forstmann, B. U. (in press). Cortico-subthalamic connection predicts individual

- differences in value-driven choice bias. *Brain Structure & Function*, 219, 1239–1249.
- Mulder, M. J., Bos, D., Weusten, J. M. H., van Belle, J., van Dijk, S. C., Simen, P., van Engeland, H., & Durson, S. (2010). Basic impairments in regulating the speed-accuracy tradeoff predict symptoms of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 68, 1114–1119.
- Mulder, M., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences-A model-based review. *Neuroscience*, 277, 872–884.
- Munoz, D. P., & Wurtz, R. H. (1995). Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *Journal of Neurophysiology*, 73, 2313–2333.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626.
- Niwa, M., & Ditterich, J. (2008). Perceptual decisions between multiple directions of visual motion. *Journal of Neuroscience*, 28, 4435–4445.
- Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113, 327–357.
- Oberauer, K., Suß, H.-M., Wilhelm, O., Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31, 167–193.
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5, 376–404.
- Philiastides, M., & Ratcliff, R. (2013). Influence of branding on preference-based decision making. *Psychological Science*, 24, 1208–1215.
- Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience*, 26, 8965–8975.
- Platt, M., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400, 233–238.
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117, 864–901.
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally-constrained modeling of perceptual decision making. *Psychological Review*, 117, 1113–1143.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R. (1980). A note on modelling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, 21, 178–184.
- Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review*, 92, 212–225.
- Ratcliff, R. (1988). A note on the mimicking of additive reaction time models. *Journal of Mathematical Psychology*, 32, 192–204.
- Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a two choice brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin and Review*, 9, 278–291.
- Ratcliff, R. (2006). Modeling Response Signal and Response Time Data. *Cognitive Psychology*, 53, 195–237.
- Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*, 120, 281–292.
- 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. *Journal of Neurophysiology*, 90, 1392–1407.
- Ratcliff, R., & Frank, M. (2012). Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. *Neural Computation*, 24, 1186–1229.
- Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical-decision task. *Psychological Review*, 111, 159–182.
- Ratcliff, R., Hasegawa, Y. T., Hasegawa, Y. P., Childers, R., Smith, P. L., & Segraves, M. A. (2011). Inhibition in superior colliculus neurons in a brightness discrimination task? *Neural Computation*, 23, 1790–1820.
- Ratcliff, R., Hasegawa, Y. T., Hasegawa, Y. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, 97, 1756–1774.
- Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*, 83, 367–381.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95, 385–408.
- Ratcliff, R., & McKoon, G. (1994). Retrieving information from memory: Spreading activation theories versus compound cue theories. *Psychological Review*, 101, 177–184.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922.
- Ratcliff, R., McKoon, G., & Tindall, M. H. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 763–785.
- Ratcliff, R., Perea, M., Colangelo, A., & Buchanan, L. (2004). A diffusion model account of normal and impaired readers. *Brain & Cognition*, 55, 374–382.
- Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences*, 106, 6539–6544.
- Ratcliff, R., & Rouder, J.N. (2000). A diffusion model account of masking in letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 127–140.
- Ratcliff, R., Sheu, C-F, & Gronlund, S.D. (1992). Testing global memory models using ROC curves. *Psychological Review*, 99, 518–535.
- Ratcliff, R. & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111, 333–367.
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: the temporal relation

- between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, *139*, 70–94.
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*, 59–83.
- Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychological Review*, *120*, 697–719.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, *19*, 278–289.
- Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception and Psychophysics*, *65*, 523–535.
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*, 408–424. 1
- Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging*, *21*, 353–371.
- Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging*, *22*, 56–66.
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*, 127–157.
- Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*, 46–487.
- Ratcliff, R., Thapar, A., Smith, P. L., & McKoon, G. (2005). Aging and response times: A comparison of sequential sampling models. In J. Duncan, P. McLeod, & L. Phillips (Eds.), *Speed, Control, and Age*, Oxford, England: Oxford University Press.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, *9*, 438–481.
- Ratcliff, R., & Van Dongen, H. P. A. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin and Review*, *16*, 742–751.
- Ratcliff, R., & Van Dongen, H.P.A. (2011). A diffusion model for one-choice reaction time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*, 11285–11290.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*, 261–300.
- Reed, A.V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, *181*, 574–576.
- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision-making. *Psychological Review*, *108*, 370–392.
- Roitman, J. D. & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*, 9475–9489.
- Roxin, A., & Ledberg, A. (2008). Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Computational Biology*, *4*, e1000046.
- Schmiedek, F., Oberauer, K., Wilhelm, O., Suß, H-M., & Wittmann, W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429.
- Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143–153.
- Sewell, D. K., & Smith, P. L. (2012). Attentional control in visual signal detection: Effects of abrupt-onset and non-onset stimuli. *Journal Of Experimental Psychology: Human Perception And Performance*, *38*, 1043–1068.
- Shadlen, M. N. & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1935.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, *4*, 145–166.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*, 567–593.
- Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.
- Smith, P. L. (2010). From Poisson shot noise to the integrated Ornstein-Uhlenbeck process: Neurally-principled models of diffusive evidence accumulation in decision-making and response time. *Journal of Mathematical Psychology*, *54*, 266–283.
- Smith, P. L., Ellis, R., Sewell, D. K., & Wolfgang, B. J. (2010). Cued detection with compound integration-interruption masks reveals multiple attentional mechanisms. *Journal of Vision*, *10*, 1–28.
- Smith, P. L., & McKenzie, C. (2011). Diffusive information accumulation by minimal recurrent neural models of decision making. *Neural Computation*, *23*, 2000–2031.
- Smith, P. L., & Ratcliff, R. (2004). The psychology and neurobiology of simple decisions, *Trends in Neuroscience*, *27*, 161–168.
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*, 283–317.
- Smith, P. L., Ratcliff, R., & Sewell, D. K. (2014). Modeling perceptual discrimination in dynamic noise: Time-changed diffusion and release from inhibition. *Journal of Mathematical Psychology*, *59*, 95–113.
- Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. *Vision Research*, *44*, 1297–1320.
- Smith, P.L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*, 135–168.
- Sperling, G. & Doshier, B. A. (1986). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of perception and performance*. (Vol. 1, pp. 1–65). New York, NY: Wiley.

- Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging, 25*, 377–390.
- Starns, J. J., & Ratcliff, R. (2012). Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks. *Psychonomic Bulletin and Review, 19*, 139–145.
- Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language, 70*, 36–52.
- Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variability and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology, 64*, 1–34.
- Stone, M. (1960). Models for choice reaction time. *Psychometrika, 25*, 251–260.
- Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18*, 415–429.
- Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology, 25*, 168–197.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.
- Townsend, J. T. & Wenger, M.J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*, 1003–1035.
- Tuerlinckx, F., Maris, E., Ratcliff, R., & De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behavior, Research, Instruments, and Computers, 33*, 443–456.
- Usher, M. & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*, 550–592.
- Vandekerckhove, J. & Tuerlinckx, F. (2007) Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review, 14*, 1011–1026.
- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16*, 44–62.
- van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2012). A diffusion model decomposition of the effects of alcohol on perceptual decision making. *Psychopharmacology, 219*, 1017–1025.
- van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E.-J. (2012). Optimal decision making in neural inhibition models. *Psychological Review, 119*, 201–215.
- Van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Vol. Ed.) & H. Pashler (Series Ed.), *Stevens' Handbook of Experimental Psychology* (3rd ed.), Volume 4: *Methodology in Experimental Psychology* (pp. 461–516). New York, NY: Wiley.
- Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13*, 37–58.
- Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic .
- Vickers, D., Caudrey, D., & Willson, R. J. (1971). Discriminating between the frequency of occurrence of two alternative events. *Acta Psychologica, 35*, 151–172.
- Voss, A. & Voss, J. (2007) Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39*, 767–775.
- Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron, 36*, 955–968.
- Watson, A. B. (1986). Temporal sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp 6–1 to 6–43). New York, NY: Wiley.
- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology, 63*, 210–238.
- White, C., Ratcliff, R., Vasey, M. & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion model analysis. *Cognition and Emotion, 23*, 181–205.
- White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*, 39–52.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67–85.
- Wickelgren, W. A., Corbett, A. T., & Doshier, B. A. (1980). Priming and retrieval from short-term memory: A speed accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior, 19*, 387–404.
- Wiecki, T. V., Sofer, I. & Frank, M. J. (2012). Hierarchical Bayesian estimation of the drift diffusion model: quantitative comparison with maximum likelihood. Program No. 494.13/ CCC30. 2012 Neuroscience Meeting Planner. New Orleans, LA: Society for Neuroscience.
- Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism for time integration in perceptual decisions. *Journal of Neuroscience, 26*, 1314–1328.
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*, 747–763.
- Zeguers, M. H. T., Snellings, P., Tijms, J., Weeda, W. D., Tamboer, P., Bexkens, A. & Huizenga, H.M. (2011). Specifying theories of developmental dyslexia: A diffusion model analysis of word recognition. *Developmental Science, 14*, 1340–1354.
- Zohary, E., Shadlen, M., & Newsome, W. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature, 370*, 140–143.