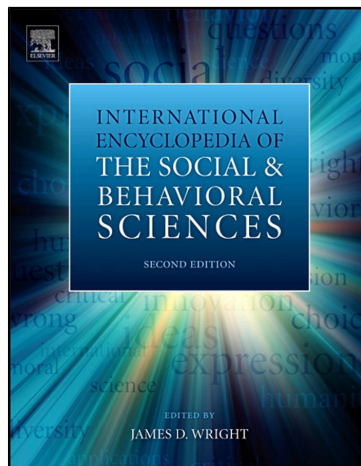


**Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.**

This article was originally published in the *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

From Smith, P.L., Ratcliff, R., 2015. Diffusion and Random Walk Processes. In: James D. Wright (editor-in-chief), *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, Vol 6. Oxford: Elsevier. pp. 395–401.

ISBN: 9780080970868

Copyright © 2015 Elsevier Ltd. unless otherwise stated. All rights reserved.

Elsevier

Diffusion and Random Walk Processes

Philip L. Smith, The University of Melbourne, VIC, Australia
 Roger Ratcliff, The Ohio State University, Columbus OH, USA

© 2015 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by R. Ratcliff, volume 20, pp. 3668–3673, © 2001, Elsevier Ltd.

Abstract

Diffusion and random walk processes comprise one of the main classes of sequential-sampling models for decision making and choice response time in psychology. The models assume that decisions are made by accumulating samples of noisy evidence to a response criterion, either in continuous or in discrete time. The models successfully predict the main features of performance in speeded decision tasks: accuracy, the ordering of mean response times for correct responses and errors, the shapes of response time distributions, and the effect of instructions. Neural correlates of the evidence accumulation process have been identified in studies of the brain regions involved in decision making.

Historical Foundations

Diffusion and random walk models form one of two major classes of sequential-sampling models for speeded decision making and choice response time in psychology. The other class comprises accumulator and counter models. A taxonomy of these two classes of models, first compared by Audley and Pike (1965), is shown in Figure 1. Models of both classes conceive of decision making as a statistical process, in which successive samples of noisy evidence or stimulus information are accumulated to a response criterion. The evidence is assumed to be noisy, either because the stimulus is

a sequence of variable or noisy events, or because of noisy coding in neural systems. The criterion represents the amount of evidence needed to make a decision. In random walk models, evidence for one response is evidence against the other. The accumulating evidence is represented as a single fluctuating sum, which increases or decreases as evidence favoring one or other response is sampled. In counter and accumulator models, evidence for the two responses is accumulated in parallel, in separate totals. The decision process in these models has the form of a race between accumulating evidence totals, with the winner of the race determining the response.

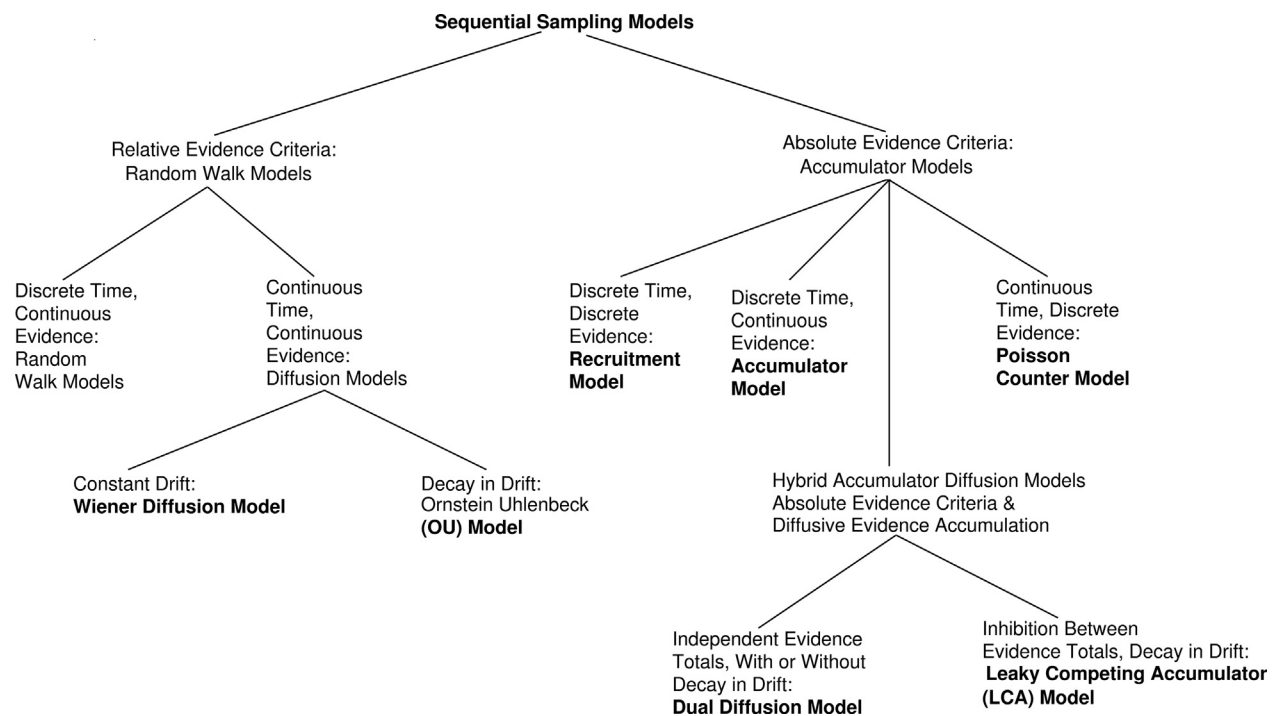


Figure 1 Sequential-sampling models. The branch on the left shows random walk and diffusion models. The hybrid accumulator–diffusion models shown under the branch on the right combine the attributes of accumulator models and diffusion models.

Random walk models are the discrete-time counterpart of diffusion process models and historically preceded them. Random walk models for decision making were introduced into psychology by Stone (1960) and Edwards (1965) and were developed subsequently by Laming (1968). The aim of the models was to provide a unified account of the processes underlying psychology's two ubiquitous dependent variables: accuracy and response time (RT). The early random walk models were influenced by Wald's (1947) sequential probability ratio test (SPRT) in statistics, which was based on the accumulation of log-likelihood ratios. SPRT models assume that the decision process acts on a sequence of noisy observations, each element of which provides evidence for one of two alternative hypotheses, H_a and H_b , which represent states of the world. In a choice response time setting, the hypotheses might be that either stimulus s_a was presented or stimulus s_b was presented. The task is to decide which of these hypotheses is true and to respond as quickly as possible. The evidence is represented as a sequence of random variables, Z_j , $j = 1, 2, \dots$, which is sampled at a constant rate. The decision process computes the log-likelihood ratio of the evidence for the two hypotheses obtained at step j and accumulates these as a running sum. If $p_a(Z_j)$ and $p_b(Z_j)$ denote the probability density functions for Z_j under H_a and H_b , respectively, then the decision process, X_i , at time i is the sum,

$$X_i = \sum_{j=1}^i \log \left[\frac{p_a(Z_j)}{p_b(Z_j)} \right]. \quad [1]$$

If the Z_j are Gaussian random variables, then the log-likelihood ratio is also Gaussian, and the running sum is a Gaussian random walk.

To make a decision, evidence is sampled until the accumulating total reaches or exceeds one of two absorbing boundaries or evidence criteria, a and b , with $b < 0 < a$. If the first criterion reached is a , the person makes response R_a (stimulus s_a was presented); if the first criterion reached is b , the person makes the alternative response R_b (stimulus s_b was presented). The decision time, T , is the minimum time to reach a criterion. Formally, if $T_a = \min\{i : X_i \geq a\}$ and $T_b = \min\{i : X_i \leq b\}$, then $T = \min(T_a, T_b)$. The time of the first boundary crossing, T , is guaranteed to be finite, which means that the process will terminate in finite time with one or other response. Wald's SPRT is optimum in the sense that it is the process that achieves a given criterion level of accuracy in the minimum expected time. It was this property that first recommended it as a model for decision making in settings in which there is a cost associated with sampling evidence, either in time or other units.

A later random walk model, the *relative judgment theory* of Link and Heath (1975), omitted the computation of likelihood ratios and assumed the walk accumulates the magnitudes of the noisy evidence sequence, Z_j , directly,

$$X_i = \sum_{j=1}^i Z_j. \quad [2]$$

This change was to allow the model to account for the orderings of mean RTs for correct responses and errors that are found experimentally. Some recent treatments of random walks and diffusion processes in psychology and neuroscience have

again begun to emphasize the connection with Wald's SPRT and optimality (Bogacz et al., 2006; Gold and Shadlen, 2002).

Diffusion Process Models

Mathematically, diffusion processes are the continuous-time counterparts of random walks. Whereas a random walk, X_i , accumulates evidence at discrete-time steps, $i = 1, 2, \dots$, a diffusion process, X_t , accumulates evidence in continuous time ($t > 0$). The study of diffusion processes was begun in physics in 1905 by Albert Einstein, who derived a partial differential equation for the probability distribution of a pollen particle undergoing Brownian motion. He showed that this equation had the same form as the classical heat equation of physics (Gardiner, 2004). In 1923, Wiener began the mathematical study of the path-space properties of diffusion processes, by proving the existence of a continuous stochastic process whose transition distribution satisfied the partial differential equation derived by Einstein. The resulting process is variously known as the *Brownian motion process* or *Wiener diffusion process*.

Ratcliff (1978) proposed a diffusion process model of two-choice decision making in psychology, arguing that information processing in the brain is better viewed as a continuous-time rather than a discrete-time process. In his model, which is shown in Figure 2, the accumulating evidence is described by the diffusion process studied by Einstein and Wiener. The model is similar to the earlier random walk models in that it conceptualizes decision making as diffusion on a line between absorbing barriers, which represent decision criteria. Ratcliff's model follows the gambler's ruin framing of

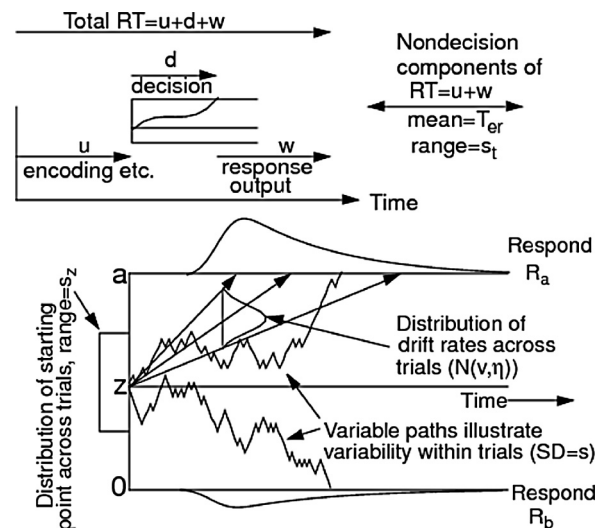


Figure 2 Diffusion model. The process starting at z accumulates evidence between absorbing boundaries at 0 and a . Moment-to-moment stochastic variability means the process can sometimes terminate at the correct response boundary, R_a , and sometimes at the error response boundary, R_b . Drift rate, ξ , is normally distributed between trials with mean ν and standard deviation η . Starting point is assumed to be rectangularly distributed with range s_z . RT is the sum of the stimulus encoding time, u , the decision time, d , and the response output time, w . The nondecision time, $T_{er} = u + w$, is rectangularly distributed with range s_t .

the process of Feller (1967), in which the lower barrier is set to 0, the starting point is set to z , and the upper barrier is set to a . Evidence for one response drives the process in an upward direction; evidence for the other response drives it in a downward direction, with the process continuing until one of the absorbing boundaries is crossed and a response is made. Unlike those models, evidence in his model accumulates continuously rather than at discrete points in time.

Like relative judgment theory, the evidence accumulated in the diffusion model is assumed to be a sequence of noisy sensory or cognitive states, rather than a likelihood ratio transformation of those states. Because the accumulation process, X_t , in the model is Gaussian, and because the likelihood ratio of a Gaussian process itself has a Gaussian distribution, the accumulating evidence in the diffusion model is proportional to the log-likelihood ratio. Consequently, although the model does not assume a likelihood ratio transformation of evidence states, it is SPRT optimal in the same way as are the likelihood ratio models of Laming (1968) and earlier researchers (Bogacz et al., 2006).

Ratcliff (1978) emphasized the importance of other sources of trial-to-trial variability in the model, in addition to diffusive noise arising from moment-to-moment variability in the evidence entering the decision process. The effects of diffusive noise are shown as short-time-scale random fluctuations in the accumulating evidence in Figure 2. The other sources of variability reflect cognitive processes that affect the decision process on individual trials. The first of these, shown in Figure 2, is variability in the rate of evidence accumulation, or the *drift rate*, of the process, which is assumed to be normally distributed. Drift variability represents trial-to-trial variation in the average quality of the information entering the decision process and arises due to variations in the quality of the match between encoded stimuli and the decision alternatives in memory. The presence of drift variability means that on some trials the process will drift in the wrong direction (i.e., presentation of stimulus s_a will cause the process to drift toward the R_b boundary or vice versa). This allows the model to predict slow errors, which are often found experimentally when stimulus discriminability is low and accuracy is stressed.

Ratcliff et al. (1999) added another source of trial-to-trial variability, first proposed by Laming (1968), in the starting point of the process. Starting point variability also allows the model to predict fast errors, which are often found experimentally when stimulus discriminability is high and speed is stressed. Fast errors arise because responses on trials on which the process has started near the error response boundary are more likely to be made rapidly and incorrectly.

Ratcliff (1978) also initiated the study of RT distributions as a method for testing decision models. Most early model comparisons focused on the ordering of mean RTs for correct responses and errors and the relationship between mean RT and accuracy. Ratcliff showed that the properties of empirical RT distributions impose strong constraints on the class of plausible models and, moreover, that the Wiener diffusion model predicts the families of RT distributions that are found experimentally. This has been confirmed in numerous studies of speeded decision making in simple perceptual and

cognitive tasks in a variety of settings (Ratcliff and Smith, 2004; Ratcliff et al., 2005).

Mathematical Methods for Diffusion Processes

Diffusion processes are continuous time, continuous state space, Markov processes. Members of the class of diffusion processes are defined by specifying the *infinitesimal moments* of the process, together with initial and boundary conditions. The infinitesimal moments are the *drift* and *diffusion* coefficients, which define, respectively, the mean rate of change and the variance in the rate of change in a unit interval. In a general diffusion process, the drift and diffusion can both depend on time and on the state of the process. For such a process, the drift, $\mu(x, t)$, and diffusion coefficient, $\sigma^2(x, t)$, can be continuous functions of the state variable, x , and the time variable, t . In Ratcliff's (1978) model, the drift and diffusion coefficients are both constant and are denoted ξ and s^2 , respectively, where $\xi \sim N(\nu, \eta)$ and where $N(\dots)$ is the normal density function and the tilde means 'is distributed as'.

Mathematically, diffusion processes can be defined via either partial differential equations or stochastic differential equations. If $f(s, y; t, x)$ is the transition density of the process, that is, $f(s, y; t, x) dx$ is the probability that a process starting at time s in state y will be found at time t in the interval $(x, x + dx)$, then the general diffusion, X_t , satisfies the partial differential equation

$$-\frac{\partial f}{\partial s} = \frac{1}{2}\sigma^2(y, s)\frac{\partial^2 f}{\partial y^2} + \mu(y, s)\frac{\partial f}{\partial y}. \quad [3]$$

This is the Kolmogorov backward equation, so called, because its variables are the starting time and state, s and y . The process also satisfies the forward equation, which is an equation in t and x . The backward equation is useful for deriving RT distributions; the forward equation is useful for studying evidence accumulation by a process unconstrained by absorbing boundaries (Ratcliff, 1978). Alternatively, X_t can be defined as satisfying the stochastic differential equation

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t, \quad [4]$$

where W_t is the Brownian motion process.

The key properties of the process in modeling decision making are the first passage time probability and the first passage time distributions. The first passage time probability is the probability that the process reaches a specified one of the absorbing boundaries before reaching the other one. In the decision model, the probability of making a particular response, say R_a , is given by the probability that the process crosses the absorbing boundary a before it crosses the boundary b . If T_a and T_b are the first passage times for the boundaries a and b , respectively, then the probability of making response R_a is $P(R_a) = P[T_a < T_b]$. The first passage time distributions, denoted $G_a(t)$ and $G_b(t)$, are the distributions of T_a and T_b . In the model, the first passage times are the decision times for the two responses. If the stimulus presented was s_a , then $G_a(t)$ and $G_b(t)$ are the distributions of correct responses and errors, respectively. The converse holds if the stimulus presented was s_b .

A variety of methods exist in the literature for obtaining first passage time statistics for diffusion processes. Classically, partial differential equations like eqn [3] are solved as algebraic eigenvalue problems. When the diffusion is time homogeneous (i.e., the drift and diffusion coefficients do not vary with time), these methods yield the first passage time distributions in the form of infinite series, which can be approximated by truncation (Ratcliff, 1978). For a diffusion with drift ξ , diffusion coefficient s^2 , and starting point z , where $b < z < a$, the probabilities of responding at the upper and lower barriers are

$$P(R_a) = \frac{\exp(-2\xi z/s^2) - \exp(-2\xi b/s^2)}{\exp(-2\xi a/s^2) - \exp(-2\xi b/s^2)}, \quad [5]$$

and

$$P(R_b) = 1 - P(R_a), \quad [6]$$

respectively. The first passage time densities, $g_a(t)$ and $g_b(t)$, defined as $g_a(t) = G'_a(t)$ and $g_b(t) = G'_b(t)$, where primes denote differentiation with respect to time (Smith, 1990), are

$$g_a(t) = \frac{1}{P(R_a)} \exp\left[\frac{\xi(a-z)}{s^2} - \frac{\xi^2 t}{2s^2}\right] \frac{\pi s^2}{(a-b)^2} \times \sum_{k=1}^{\infty} k \exp\left[-\frac{k^2 \pi^2 s^2 t}{2(a-b)^2}\right] \sin\left[\frac{k\pi(a-z)}{a-b}\right] \quad [7]$$

and

$$g_b(t) = \frac{1}{P(R_b)} \exp\left[\frac{\xi(z-b)}{s^2} - \frac{\xi^2 t}{2s^2}\right] \frac{\pi s^2}{(a-b)^2} \times \sum_{k=1}^{\infty} k \exp\left[-\frac{k^2 \pi^2 s^2 t}{2(a-b)^2}\right] \sin\left[\frac{k\pi(z-b)}{a-b}\right]. \quad [8]$$

In applications, it is often more convenient to work with first passage time distribution functions rather than densities. Ratcliff (1978) gave the distribution functions using Feller's (1967) gambler's ruin form of the process, in which the lower boundary is set to zero ($0 < z < a$).

Several other methods have been investigated for obtaining first passage time statistics for diffusion processes. These methods can be used for more complex problems, such as, for example, when either the drift and diffusion coefficients or the absorbing boundaries vary with time, for which classical methods for solving partial differential equations often do not yield tractable solutions. These include integral equation methods, originally developed to study models of integrate-and-fire neurons in mathematical biology (Smith, 2000), numerical methods for solving partial differential equations (Voss and Voss, 2008), and matrix methods (Diederich and Busemeyer, 2003). The matrix approach, which approximates diffusion processes with finite state Markov chains, can be applied to very general problems of the kind that sometimes arise in applications, such as multidimensional diffusions with correlated increments (Ditterich, 2006). Although such problems can be made theoretically tractable via the use of finite state approximations, the size of the transition matrix required to represent the process and the associated computational demands rapidly become large. Tuerlinckx et al. (2001) have investigated methods for simulating diffusion processes.

Ornstein-Uhlenbeck Models

Since Ratcliff's (1978) article, researchers have investigated a number of variations on the underlying model. Busemeyer and Townsend (1993) proposed a model based on the Ornstein-Uhlenbeck (OU) diffusion process. This process was proposed in 1930 by Uhlenbeck and Ornstein as an alternative model for Brownian motion to that of Einstein and Wiener (Gardiner, 2004). It sought to correct a perceived shortcoming of that model, namely, it implies the velocity of a diffusing particle is almost everywhere infinite. Rather than modeling the displacement of the particle, the OU process models its velocity. Mathematically, it differs from the Wiener process by the presence of a decay term in the drift coefficient, which acts as a restoring force that pulls the particle back toward its starting point. For a particle starting at the origin ($X_0 = 0$), the OU process is defined by the stochastic differential equation,

$$dX_t = (\mu - \beta X_t)dt + \sigma dW_t. \quad [9]$$

In this equation, μ is the external component of drift, which in decision-making models depends on the stimulus, $-\beta X_t$ is the restoring force, and σ^2 is the diffusion coefficient. When $\beta = 0$, the process becomes the Wiener diffusion process of Ratcliff's model (with $\mu \equiv \xi$ and $\sigma \equiv s$). Busemeyer and Townsend (1993) interpreted the OU decay term in their model as representing avoidance in a system of approach-avoid dynamics.

Most applications of diffusion models before Busemeyer and Townsend (1993) had been to brief two-choice decisions based on a single, 'one shot' cognitive process, for which RTs do not average much more than a second. They proposed that diffusion models could also be applied to more complex tasks, in which decisions are made after a period of deliberation. They showed their OU model could provide an account of otherwise puzzling findings in the judgment and decision-making literature, such as preference reversals, which are difficult to explain using traditional algebraic models (Roe et al., 2001).

Other investigators have proposed OU models in which the decay term represents a bounded evidence accumulation process. Smith (1995) proposed a model of simple RT with time-varying drift, $\mu(t) - \beta x$, in which the decision process is driven by sustained and transient visual channels. Usher and McClelland (2001) proposed a neurally inspired model of choice RT, the *leaky competing accumulator model* (LCA), based on a pair of coupled OU processes. Their model combined the properties of earlier accumulator models, such as that of Vickers (1970) – in which evidence for competing responses is accumulated in parallel by racing evidence totals – and diffusion process models, in which the accumulation process is diffusive. Denoting the evidence totals by X_1 and X_2 (and suppressing the dependence on the time variable, t), the growth equations in the LCA model are

$$dX_i = (\mu_i - \beta X_i - kX_j)dt + \sigma dW_i, \quad i, j \in \{1, 2\}, \quad i \neq j. \quad [10]$$

In these equations, k is a coupling coefficient that represents the amount of mutual inhibition between the accumulating totals X_1 and X_2 , and W_1 and W_2 are independent Brownian

motion processes. The relative optimality properties of the model were analyzed by Bogacz et al. (2006).

Diffusion Models and Empirical Data

Scaling and Distribution Shape

The two dependent variables, RT and accuracy, have different scale properties. RT has a minimum value, and its variance increases as RT increases. Accuracy is bounded on the range 0.5–1.0, and its variance decreases as it approaches 1.0. Diffusion models account for these scale properties automatically as a result of the geometry of the diffusion process. When mean drift rate is high, the probability of a correct response is near 1.0, and decision processes approach the correct boundary with little spread in arrival times. When mean drift rate is nearer zero, variability leads some processes to hit one boundary, other processes to hit the other boundary, and accuracy nears 0.5; the arrival times at boundaries are long and highly variable. In addition to these scale properties, the geometry of the diffusion process also gives RT distributions that are skewed to the right; as mean RT increases, due to a reduction in drift rate, the fastest responses slow a little and the slowest responses slow a lot.

Speed Versus Accuracy

People can manipulate their speed performance relative to their accuracy performance. If they are instructed in some blocks of an experiment to respond as quickly as possible, and in other blocks to respond as accurately as possible, mean RTs can vary between the two kinds of blocks by as much as 500 ms and mean accuracy can vary by 10% (Ratcliff and Rouder, 1998). Ratcliff's diffusion model can account for these differences with only the boundary separation parameter: moving the boundaries close together produces fast responses for which variability in drift can cause large numbers of errors; moving the boundaries further apart produces slower responses that are more likely to be accurate.

Error Response Times

Historically, comparisons of sequential-sampling models focused on their ability to predict empirical patterns of mean RTs for correct responses and errors. A problem with early random walk models was that they predicted that mean RTs for correct responses and errors would be the same. Specifically, the models predicted that $M(T_a|s_a) = M(T_a|s_b)$ and $M(T_b|s_b) = M(T_b|s_a)$, where $M(T_i|s_j)$ denotes the mean time to make response R_i when stimulus s_j is presented. This prediction is made by both SPRT random walks and Gaussian random walks in which the drift rates for the stimuli s_a and s_b are of equal magnitude and opposite sign: μ and $-\mu$, respectively (Luce, 1986; Townsend and Ashby, 1983).

A claimed advantage of the class of accumulator models (the right branch of Figure 1), which includes the recruitment model (LaBerge, 1962), the Poisson counter model (Townsend and Ashby, 1983), and the Vickers accumulator model (Vickers, 1970), is they predict that mean RTs for error responses will be slower than mean RTs for correct responses. But this represents

only a partial solution to the problem of RT ordering, as errors can be faster or slower than correct responses depending on stimulus discriminability and instructions (Luce, 1986). Laming (1968) showed that variability in the starting point of a SPRT random walk allowed the model to predict fast errors, but this too represented only a partial solution.

One solution to the RT ordering problem was proposed in Link and Heath's (1975) relative judgment theory, which allows the distribution of evidence states, Z_j , in eqn [2] to be non-Gaussian. They showed that the ordering of mean RTs for correct responses and errors depends on the relative sizes of the steps taken by the walk toward the correct and error boundaries, which is characterized mathematically by asymmetry in the moment generating function of the evidence states. By allowing the asymmetry to vary freely, the model can predict either fast or slow errors. Although this yielded a formal solution to the problem of mean RT ordering, the theory did not provide an account of why the evidence distribution should vary with conditions that produce changes in ordering, such as instructions to respond rapidly or slowly.

Ratcliff's diffusion model, with trial-to-trial variability in drift rate and starting point, provides an alternative solution to the ordering problem. As a Gaussian process, it predicts equal mean times for correct responses and errors when there is no trial-to-trial variability. When there is variability in both drift and starting point, it can predict fast errors and slow errors, and the crossover pattern found in some experiments in which errors are faster in some conditions and slower in others (Ratcliff and Smith, 2004). As well as predicting the ordering of mean RTs, it also predicts the ordering of correct and error RT distributions. It therefore gives a complete account of the data from simple two-choice decision experiments. Unlike moment-generating function asymmetry, the sources of trial-to-trial variability in the model have clear psychological interpretations.

Quantile Probability Plot of RT Distributions

An experimental task in which the difficulty of the stimulus discrimination is varied yields a family of RT distribution pairs and choice probabilities. Each stimulus condition yields a distribution of correct responses and a distribution of errors, and the probability of a correct response. These data, which impose strong constraints on models of the decision process, can be represented in the form of a quantile probability plot, as shown in Figure 3. The RT distributions are summarized by their .1, .3, .5, .7, and .9 quantiles. The .1 quantile summarizes the leading edge of the distribution; the .5 quantile summarizes the 'typical' (median) response; and the .9 quantile summarizes the distribution's tail. For a given experimental condition, i , the quantiles of the distribution of RTs for correct responses are plotted against the probability of a correct response, q_i , and the quantiles of the distribution of RTs for errors are plotted against the probability of an error response ($1 - q_i$). The distributions on the right of the .5 point on the x -axis are the distributions of correct responses, and the distributions on the left are the distributions of errors.

The main features of performance are shown on this plot. The spacing of the RT quantiles on the y -axis shows the unimodal, positively skewed form that is typical of RTs in

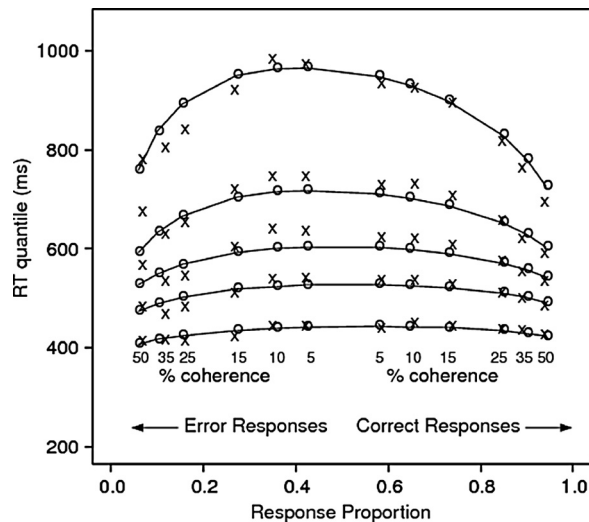


Figure 3 Quantile probability plot. Quantiles of the RT distributions are plotted on the y -axis against choice probabilities on the x -axis. The experimental task was a motion discrimination task in which discriminability was manipulated by varying the coherence of the motion in six steps. The x 's are the experimental data; the o 's connected by continuous lines are the fitted values of the Ratcliff diffusion model. Source: Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks, *Neural Computation*, 20, 873–922, Figure 7. Reprinted by permission.

simple cognitive tasks. The mean and standard deviation of the RT distributions increase, roughly in proportion to one another (Wagenmakers and Brown, 2007), as the discriminability of the stimulus is reduced. Most of the change in the RT distribution with changes in stimulus discriminability occurs in the tail; there is little change in the leading edge. In the data set shown, error RTs are typically slower than correct responses, which appears as an asymmetry of the plot across its vertical midline. Figure 3 also shows fits of the Ratcliff diffusion model to the RT distributions and choice probabilities. The figure shows that the model captures the shapes of the RT distributions for correct responses and errors and their relationship with the proportions of correct and error responses.

Comparison of Diffusion and Accumulator Models

Ratcliff and Smith (2004) compared the performance of diffusion models, accumulator models, and models combining the attributes of both classes, in which evidence for competing responses accumulates in separate totals that are modeled as diffusion process. One such model was the LCA model (eqn [10]); the other, the *dual diffusion model*, assumed a pair of independent parallel diffusion processes. This model is like the LCA model but omits the mutual inhibition term, $-kX_j$, of eqn [10].

Ratcliff and Smith found that model architecture – whether a model assumed one evidence total or two – was less important than the process used to model evidence accumulation. Models with diffusive evidence accumulation, in either

a single evidence total or two racing totals, were able to account for the main features of the experimental data, whereas the accumulator models were not. Accumulator models predict RT distributions that become less skewed and more symmetrical with increases in the mean and variance of RT, which is not what is found experimentally (e.g., Figure 3). The successful models assume that accumulating evidence can both increase and decrease, whereas in the unsuccessful models it only increases. Models with diffusive accumulation predict the shapes of RT distributions that are found empirically, whereas models with other kinds of accumulation process do not.

For the diffusion and hybrid accumulator–diffusion models, neither decay nor mutual inhibition between evidence totals was needed to account for empirical data – although these features are often thought to confer biological plausibility. With small values of decay, β , the OU model (eqn [9]) was indistinguishable from the Ratcliff diffusion model. With large values of decay, it predicted RT distributions that were more skewed than those found experimentally. The LCA model, which assumes mutual inhibition between evidence totals, and the dual diffusion model, which does not, gave similarly good accounts of the data. Mutual inhibition appears to be needed in models in which stimulus inputs only produce increases in evidence totals; in such models, mutual inhibition can provide the fluctuations in evidence needed to correctly predict RT distributions (e.g., Purcell et al., 2012).

Neural Studies of Diffusion Making

Recently, there has been considerable interest in neural analogs of diffusion processes. This has been stimulated by the emerging subdiscipline of neuroeconomics, which focuses on the computation of value, reward, and uncertainty in the brain (Glimcher et al., 2009), and by single-cell recordings from awake behaving monkeys performing eye-movement analogs of human-speeded decision tasks (Hanes and Schall, 1996; Roitman and Shadlen, 2002). These studies have reported patterns of firing rates in decision-related brain structures that appear to be neural correlates of the accumulation process hypothesized in the psychological models (Ratcliff and Smith, 2004). Cells in brain structures like the frontal eye fields, lateral intraparietal area, and superior colliculus, which respond selectively to the presence of one of the decision alternatives in their receptive fields, show a progressive increase in firing rates to a constant level that is time locked to the production of a response. This characteristic pattern of firing rates appears to be the neural signature of the process of accumulating evidence to a criterion.

In the past decade, there has been a proliferation of research investigating the connection between the psychological processes of evidence accumulation and the neural processes of decision making. Ratcliff et al. (2007) showed that the parameters of a dual-diffusion model estimated from the RT distributions and choice probabilities of monkeys performing an eye-movement decision task correctly predicted the time course and relative firing rates in build-up cells in the superior colliculus in the period preceding the response. Ratcliff et al. (2009) identified a late

component of the human electroencephalogram (EEG) that indexes decision making and showed that trial-to-trial variations in the amplitude of this component predicted estimates of drift rates in a diffusion model fitted to the RT distributions and accuracy data. Mulder et al. (2012) linked estimates of bias (starting point, z) in a diffusion process to activity in frontoparietal brain areas involved in the processing of probabilities and rewards.

Smith (2010) investigated the Poisson *shot noise process* (Gardiner, 2004) as an idealized model of neural information coding. He showed that when evidence states are described by the difference of two shot-noise processes, the resulting accumulation process approximates an integrated OU process, or *OU displacement process*. This process, which is the time integral of the OU velocity process in eqn [9], has similar properties to the Wiener diffusion process in Ratcliff's model and, consequently, makes similar RT distribution and accuracy predictions. It thus provides a theoretical account of how diffusive information at a behavioral level can arise from the statistical coding of stimuli in the underlying neural population.

Conclusions

Many researchers have contributed to the development of sequential-sampling models since they were introduced into psychology in the 1960s, and many different models have been proposed of how evidence is accumulated to make a decision. Models that represent evidence accumulation as a diffusion process have been particularly successful in accounting for performance in a wide variety of speeded decision tasks. Neuroscientists have begun to study the brain processes involved in decision making and have reported patterns of activity that appear to be correlates of the psychological processes of evidence accumulation hypothesized in the models. The models thus appear to have the potential to help us understand the theoretical link between behavior and its neural underpinnings.

See also: Computational Neuroscience; Decision and Choice: Economic Psychology; Decision and Choice: Sequential Decision Making; Markov Decision Processes; Prefrontal Cortex.

Bibliography

- Audley, R.J., Pike, A.R., 1965. Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology* 18, 207–225.
- 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.
- 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.
- 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.
- Ditterich, J., 2006. Evidence for time-variant decision making. *European Journal of Neuroscience* 24, 3628–3641.

- Edwards, W., 1965. Optimal strategies for seeking information: models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology* 2, 312–329.
- Feller, W., 1967. *An Introduction to Probability Theory and Its Applications*, third ed. Wiley, New York.
- Gardiner, C.W., 2004. *Handbook of Stochastic Methods*, third ed. Springer, Berlin.
- Glimcher, P.W., Camerer, C.F., Fehr, E., Poldrack, R.A., 2009. *Neuroeconomics: Decision Making and the Brain*. Elsevier, Amsterdam.
- Gold, J.I., Shadlen, M.N., 2002. Banburismus and the brain: decoding the relationship between sensory stimulus, decisions and reward. *Neuron* 36, 299–308.
- Hanes, D.P., Schall, J.D., 1996. Neural control of voluntary movement initiation. *Science* 274, 427–430.
- LaBerge, D.A., 1962. A recruitment theory of simple behavior. *Psychometrika* 27, 375–396.
- Laming, D.R.J., 1968. *Information Theory of Choice Reaction Times*. Wiley, New York.
- Link, S.W., Heath, R.A., 1975. A sequential theory of psychological discrimination. *Psychometrika* 40, 77–105.
- Luce, R.D., 1986. *Response Times*. Oxford University Press, New York.
- Mulder, M.J., Wagenmakers, E.J., Ratcliff, R., Boekel, W., Forstmann, B.U., 2012. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience* 32, 2335–2343.
- Purcell, B.A., Schall, J.D., Logan, G.D., Palmeri, T.J., 2012. From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. *Journal of Neuroscience* 32, 3433–3446.
- Ratcliff, R., 1978. A theory of memory retrieval. *Psychological Review* 85, 59–108.
- Ratcliff, R., Rouder, J.N., 1998. Modeling response times for two-choice decisions. *Psychological Science* 9, 347–356.
- Ratcliff, R., Smith, P.L., 2004. A comparison of sequential sampling models for two-choice reaction time. *Psychological Review* 111, 333–367.
- Ratcliff, R., Philastides, 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 of the United States of America* 106, 6539–6544.
- Ratcliff, R., Van Zandt, T., McKoon, G., 1999. Connectionist and diffusion models of reaction time. *Psychological Review* 106, 261–300.
- Ratcliff, R., Thapar, A., Smith, P.L., McKoon, G., 2005. Aging and response times. In: Duncan, J., Phillips, L., McLeod, P. (Eds.), *Measuring the Mind*. Oxford University Press, Oxford, UK.
- Ratcliff, R., Hasegawa, Y., Hasegawa, R., Smith, P.L., Segraves, M., 2007. A dual diffusion model for single cell recording data from the superior colliculus in a brightness discrimination task. *Journal of Neurophysiology* 97, 1756–1774.
- 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. Responses of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience* 22, 9475–9489.
- Smith, P.L., 1990. A note on the distribution of response time for a random walk with Gaussian increments. *Journal of Mathematical Psychology* 34, 445–459.
- Smith, P.L., 1995. Psychologically principled models of visual simple reaction time. *Psychological Review* 102, 567–591.
- 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 information accumulation in decision-making and response time. *Journal of Mathematical Psychology* 54, 266–283.
- Stone, M., 1960. Models for choice-reaction time. *Psychometrika* 25, 251–260.
- Townsend, J.T., Ashby, F.G., 1983. *Stochastic Modeling of Elementary Psychological Processes*. CUP, Cambridge.
- Tuerlinckx, F., Maris, E., Ratcliff, R., De Boeck, P., 2001. A comparison of four methods for simulating the diffusion process. *Behavior Research Methods, 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, 550–559.
- Vickers, D., 1970. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics* 13, 37–58.
- Voss, A., Voss, J., 2008. A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology* 52, 1–9.
- Wagenmakers, E.-J., Brown, S., 2007. On the linear relationship between the mean and standard deviation of a response time distribution. *Psychological Review* 114, 830–841.
- Wald, A., 1947. *Sequential Analysis*. Wiley, New York.