Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Mathematical Psychology 47 (2003) 304–322
Journal of Mathematical Psychology
http://www.elsevier.com/locate/jmp

Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time

Adele Diederich[a,*] and Jerome R. Busemeyer[b]

[a] School of Humanities and Social Sciences, International University Bremen, P.O. Box 750 561, D-28725 Bremen, Germany
[b] Psychology Department, Indiana University, Bloomington, IN 47405, USA

Received 19 January 2002; revised 8 January 2003

Abstract

Diffusion processes (e.g., Wiener process, Ornstein–Uhlenbeck process) are powerful approaches to model human information processes in a variety of psychological tasks. Lack of mathematical tractability, however, has prevented broad applications of these models to empirical data. This tutorial explains step by step, using a matrix approach, how to construct these models, how to implement them on a computer, and how to calculate the predictions made by these models. In particular, we present models for binaries choices for unidimensional and multiattribute choice alternatives; for simple reaction time tasks; and for three alternatives choice problems.

Keywords: Diffusion models; Wiener process; Ornstein–Uhlenbeck process; Markov chains

1. Introduction

Over the last 40 years there have been many systematic attempts to model the processes underlying human decision making in a wide variety of psychological tasks (for reviews see, Luce, 1986; Townsend & Ashby, 1983). One line of thought is based on the notion of sequential sampling according to which information in sensory or cognitive systems is accumulated sequentially over time until a preset criterion is reached and a response is initiated. *Information* in this context usually means any changes in the central nervous system that translate perception and cognition into action (Luce, 1986; Smith, 2000). Characteristics, both of the stimulus (e.g., intensity) and the subject (e.g., strategy), may influence the information accumulation process.

Sequential sampling models seek to account for both response time and accuracy data. All approaches assume that the stimuli (or choice alternatives) can be mapped onto a hypothetical numerical dimension representing the instantaneous level of activation, evidence, or preference. Further, they assume some random fluctuation of this value over time in the course of information accumulation. Therefore, sequential sampling can be described as a stochastic process. Two quantities are of foremost interest: (1) the probability that the process eventually reaches one or the other boundary for the first time (the criterion to initiate a response), called *first passage probability*; (2) the time it takes for the process to reach one of the boundaries for the first time, called *first passage time*. The former quantity is related to the observed relative frequencies, the latter usually to the observed mean choice response times.

Traditionally, sequential sampling models have been applied to account for response time and accuracy data in identification and discrimination tasks using choice response time paradigms (e.g., Audley (1960), Audley and Pike (1965); Ashby (1983), Edwards (1965), Heath (1981), Laming (1968), Link and Heath (1975), Pike (1966); Ratcliff (1978, 1981, 1985, 1988), Ratcliff and McKoon (1982), Ratcliff and Rouder (1998, 2000), Ratcliff, Van Zandt, and McKoon (1999), Stone (1960), Van Zandt, Colonius, and Proctor (2000), Vickers (1970); for reviews see Luce (1986), Townsend and Ashby (1983)). They were also used for classification tasks (e.g., general recognition theory, Ashby (2000); exemplar-based random walk models of classification, Nosofsky and Palmeri (1997)) and decision making tasks (e.g., criterion-dependent choice model,

*Corresponding author. Fax: +49-421-200-49-3431.
E-mail address: a.diederich@iu-bremen.de (A. Diederich).

0022-2496/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0022-2496(03)00003-8

Aschenbrenner, Albert, and Schmalhofer (1984); decision field theory, Busemeyer and Townsend (1993); multiattribute dynamic decision model, Busemeyer and Diederich (2002), Diederich (1996, 1997, 2003a,b); Diederich and Busemeyer (1999)). Moreover, they have been applied to account for observed response time patterns in detection tasks using *simple* response time paradigms (e.g., Diederich (1992, 1995); Pacut (1980); Schwarz (1994); Smith (1995)). Most recently, these models have been applied to multialternative decision problems (Roe, Busemeyer, & Townsend, 2001; Usher & McClelland, 2001).

However, lack of mathematical tractability has prevented broad applications of these models to empirical data, especially when the models go beyond utilizing simple random walks. The present tutorial shows how to construct these models, how to implement them on a computer, and how to calculate the predictions made by these models. We only include basic mathematics in the tutorial; derivations and additional comments are put in the appendix. As will be shown all quantitative predictions are calculated rather than simulated, using Markov chain approximations. The tutorial is organized as follows. We start with a simple numerical example and introduce the simple random walk model. Then we show how the simple random walk can approximate the Wiener process. Third, we introduce the birth–death chain and show how this chain can approximate a diffusion process. As a special diffusion process we develop the Ornstein–Uhlenbeck process. We show how to extend the approach to multiattributive stimuli and, for the first time, to three choice alternatives and to processes with variable drift. We included mathematical derivations in the text only if absolutely necessary. Details are put in the appendix. Finally, we show how to implement the models on the computer by providing two Matlab programs.

2. Simple random walk

Intuitively, a stochastic process is an entity that evolves randomly in time or space. In physics or biology this entity may be particles or neural activation. Here, we will refer to it as amount of information throughout the present tutorial although in psychology it is also called activation, pieces of evidence, preference and the like.

Consider the following very simple example in the context of signal detection. Suppose a person is trying to make the decision whether a signal or noise trial is presented, based on the stimulus presented on a trial.

At any moment in time, the person may sample an amount of information that provides evidence for a signal or noise response. For this example, the probability of sampling evidence in favor of the signal response equals 0.4; the probability of sampling evidence favoring the noise response is 0.6. Amounts of information are coded in units of one and minus one, and the accumulated information is labeled X for the *state* of the process. Assume that information in favor of the signal is added $(+1)$ and information in favor of the noise is subtracted (-1). Let X_n denote the amount of information accumulated up to time unit n, $n = 0, 1, 2, 3, \ldots$. (X_n is also called the value of the process at time n.) Furthermore, assume that at the beginning of the trial no amount of information has been sampled yet, $X_0 = 0$, and that the sampling process stops as soon as a critical value (state) is reached for either initiating a response in favor of the signal or in favor of the noise. Let us assume that the critical value (state) for the signal response is 3 and for the noise response is -3. Three possible realizations of this information accumulation process are

$$(0, +1, 0, -1, -2, -3) \quad \text{trial } \# \ 1,$$
$$(0, +1, +2, +1, +2, +3) \quad \text{trial } \# \ 2,$$
$$(0, -1, 0, -1, -2, -1, -2, -3) \quad \text{trial } \# \ 3.$$

In the first trial, the first amount of information gives evidence for the signal and the state changes from neutral (state 0) to a state that moderately favors the signal (state $+1$). But the evidence is not sufficiently strong to decide. The next amount of information favors noise, producing a step back to neutral. The process continues until one of the critical values is reached, at which point the respective response is chosen, i.e., as soon as $X_n = +3$ or $X_n = -3$. In the case of trial 1, the evidence is sufficiently strong to make a decision in favor of the noise after the 6th sample. In the second trial there was enough information accumulated in favor of the signal after the 6th sample, in the last trial there was sufficiently strong evidence to make a decision in favor of the noise after the 8th sample. For the last example, the cumulative amount of information at each time is $X_0 = 0; X_1 = -1; X_2 = 0; X_3 = -1; X_4 = -2; X_5 = -1; X_6 = -2; X_7 = -3$.

This is an example of a very basic stochastic process, called a *simple random walk with two absorbing states*. It is *simple* because steps (amounts of information) take only values of $+1$ and -1. Random walk indicates that the steps up $(+1)$ and down (-1) occur according to a probability distribution, here $Pr(step = +1) = 0.4$ and $Pr(step = -1) = 0.6$. An absorbing state means that the process stops, or is absorbed as soon as it reaches this state, in this example $+3$ or -3. Absorbing states are also called *absorbing boundaries* or *absorbing barriers*. With respect to psychological applications this is called *decision criterion*, *decision boundary*, or simply *criterion*. The remaining intermediate states $(-2, -1, 0, +1, +2)$ are called *transient states*. Eventually, the state of

information leaves the transient states and is captured by one of the absorbing states. That is, the probability of absorbtion is 1. The set of all states of the process is called the *state space* and is labeled S.

As we have indicated at the beginning, two quantities are of foremost interest: (1) the probability that the process eventually reaches one or the other boundary for the first time (the criterion to initiate a response), which is then related to the observed relative frequencies; (2) the time it takes for the process to reach one of the boundaries, which is usually related to the observed mean choice response times.

According to this simple model, the probability of choosing a signal is determined by the probability of eventually reaching the $+3$ absorbing state before reaching the -3 absorbing state. The mean time for initiating a "signal" response is simply the mean number of steps required to reach the absorbing state for signal, given that "signal" is chosen. Both quantities can be computed in a simple way by using Markov Chain theory as follows.

First, we construct a matrix representing all the probabilities of the states. The rows of the state matrix are organized with states $-3, -2, -1, 0, +1, +2, +3$ represented by rows 1, 2, 3, 4, 5, 6, and 7, respectively; and similarly for the columns. This matrix (the entries within the bracket) is called *transition probability matrix* and labeled **P**.

$$\mathbf{P} = \begin{array}{c|c} \textit{index} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \textit{state} & \begin{array}{ccccccc} -3 & -2 & -1 & 0 & +1 & +2 & +3 \end{array} \\ \hline \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} & \begin{array}{c} -3 \\ -2 \\ -1 \\ 0 \\ +1 \\ +2 \\ +3 \end{array} \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right] \end{array}.$$

As we will see, this matrix plays an important role for determining the predicted choice probabilities and choice response times for all processes discussed in the following.

Each cell, p_{ij}, of this matrix represents the probability of going from state i to state j and is called *transition probability*. For example, from column 4 to column 5 reflects the transition probabilities from the neutral state (state 0) to either the state one step up, (to state $+1$ from row 4 to column 5 with probability 0.4); or remaining in neutral (with probability 0); or to the state one step down (to state -1 from column 4 to column 3 with probability 0.6). The remaining rows are defined in a similar manner.

For computational convenience the transition probability matrix can be rearranged according to the transition states. The absorbing states are -3 and $+3$ with transition probabilities $p_{11} = 1$ and $p_{77} = 1$, i.e., the process is absorbed with probability 1. The matrix containing these probabilities is labeled $\mathbf{P_I}$. The size of this submatrix is determined by the number of absorbing states, here a 2×2 matrix. The second class of states are those that eventually lead to the absorbing states in one-step transition, here from $+2$ to $+3$ with transition probability $p_{67} = 0.4$ and from -2 to -3 with transition probability $p_{21} = 0.6$. The matrix containing these probabilities is labeled **R**. The size of this submatrix is determined by the number of absorbing states and the number of the transient states, here a 5×2 matrix. The matrix of the probabilities for the remaining transient states is labeled **Q**. Its size is determined by the number of transient states, here a 5×5 submatrix. Obviously, the number of transient states is the total number of states minus the number of absorbing states.

$$\mathbf{P} = \begin{array}{c|c} \textit{state} & \begin{array}{ccccccc} -3 & +3 & -2 & -1 & 0 & +1 & +2 \end{array} \\ \hline \begin{array}{c} -3 \\ +3 \\ -2 \\ -1 \\ 0 \\ +1 \\ +2 \end{array} & \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0.4 & 0 & 0 & 0 & 0.6 & 0 \end{array} \right] \end{array} = \left[\begin{array}{c|c} \mathbf{P_I} & \mathbf{0} \\ \hline \mathbf{R} & \mathbf{Q} \end{array} \right]. \quad (1)$$

Furthermore, we construct a vector, labeled **Z**, representing the starting position of the process over the transient states, before the decision process begins. That is, at time $t = 0$ the process is set in motion either by starting it at a fixed state i_0, called the *initial state*, or by randomly locating it in the state space according to a probability distribution **Z** on S, called the *initial distribution*. In the former case, **Z** is the distribution concentrated at the state i_0, i.e., $Z_j = 1$ if $j = i_0$, $Z_j = 0$ if $j \neq i_0$. In the latter case, the probability is z_i that at time zero the process will be found in state i, where $0 \leqslant z_i \leqslant 1$ and $\sum_i z_i = 1$. For example, $\mathbf{Z} = (0\,0\,1\,0\,0)$ guarantees that the process starts at the neutral state (at state 0 with probability 1), as for the example above; with $\mathbf{Z} = (0\,0\,0\,1\,0)$ the process starts at state $+1$, indicating a bias towards signal. In both cases the process starts at a *fixed state*. Setting $\mathbf{Z} = (0.05\ 0.10\ 0.70\ 0.10\ 0.05)$ means, there is a 0.70 probability of starting the decision process in the neutral state, and a 0.1 probability of starting in state $+1$, etc. In this case we have a *initial probability distribution* over all transient states.

Finally, we need to build a matrix called *identity matrix*, labeled **I**, which has ones in the diagonal and zeros in the off diagonal. It is only for computational

reasons and has the same size (number of rows and columns) as the matrix **Q**. For this example, it is a 5×5 matrix defined as follows:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The probability of choosing the signal response, $Pr(S)$, and the probability of choosing the noise, $Pr(N)$, can be computed according to the matrix equation (e.g., Bhat, 1984):

$$[Pr(N), Pr(S)] = \mathbf{Z} \cdot (\mathbf{I} - \mathbf{Q})^{-1} \cdot \mathbf{R}, \tag{2}$$

where $(\mathbf{I} - \mathbf{Q})^{-1}$ is the inverse of the matrix $(\mathbf{I} - \mathbf{Q})$ and $(\mathbf{I} - \mathbf{Q})^{-1}(\mathbf{I} - \mathbf{Q}) = \mathbf{I}$. (For derivation see Appendix A.)

The mean number of steps required to reach the absorbing state either for signal or for noise can be computed in a similar way. Define $E(T|S)$ as the mean number of time steps required to make a choice, given that a signal response is chosen; and similarly, $E(T|N)$ as the mean number of time steps, given a noise response is chosen. The conditional expected values are

$$[E(T|N), E(T|S)] = [\mathbf{Z} \cdot (\mathbf{I} - \mathbf{Q})^{-2} \cdot \mathbf{R}]./[Pr(N), Pr(S)], \tag{3}$$

where $(\mathbf{I} - \mathbf{Q})^{-2}$ indicates the matrix product $[(\mathbf{I} - \mathbf{Q})^{-1} \cdot (\mathbf{I} - \mathbf{Q})^{-1}]$. (For derivation see Appendix A.) In this formula, the numerator produces a row vector of two numbers, and the denominator is also a row vector of two numbers, and the symbol ./ means an element-wise division (divide the first element in the numerator by the first in the denominator, and the second element in the numerator by the second in the denominator). These equations can be computed almost directly as shown using commonly available matrix language programs such as Gauss, Matlab, Mathematica, or Proc IML in SAS.

The choice probabilities and conditional expectations for the above example are:

Initial distribution	$Pr(S)$	$Pr(N)$	$E(T\|S)$	$E(T\|N)$
$\mathbf{Z} = (0\ 0\ 1\ 0\ 0)$	0.23	0.77	8.14	8.14
$\mathbf{Z} = (0\ 0\ 0\ 1\ 0)$	0.39	0.61	5.93	9.77
$\mathbf{Z} = (0.05\ 0.10\ 0.70\ 0.10\ 0.05)$	0.25	0.75	7.26	7.77

In the above formulas, there is nothing special about the particular probabilities, and any transition probabilities may be substituted. Also, there is nothing special about the number of transient states, and the same formula applies to an arbitrary number of transient states. Finally, as we have seen there is nothing special about the initial probability vector, and any distribution is permitted.

3. Wiener process

Before discussing the Markov-chain method to compute the distribution statistics for the Wiener process we should mention that there are other solutions to the same problem.[1] However, the solution presented, for example, by Ratcliff (1978) is limited only to the Wiener process, whereas the Markov chain method can be used for other, more complex processes as discussed later.

Consider the simple random walk again. The state space of the process is

$$S = \{-\theta, \ldots, -2, -1, 0, +1, +2, \ldots, +\theta\},$$

where θ denotes the absorbing states (the decision criteria). The parameter set (time space) is $\{0, 1, 2, 3, \ldots\}$. Now suppose we have a time interval $[0, t]$ and divide this into subintervals of length τ so that we have t/τ such subintervals. That is, the process makes a step at $\tau, 2\tau, 3\tau, \ldots$. Further, instead of steps up and down of unit magnitude assume that the size of the steps is either $+\Delta$ or $-\Delta$ with probability 0.5, so that the move is up or down. The state space is now

$$S = \{-k\Delta, \ldots, -2\Delta, -\Delta, 0, +\Delta, +2\Delta, \ldots, +k\Delta\}, \tag{4}$$

where $\theta = k\Delta$ and $-\theta = -k\Delta$, and k equals the number of steps to reach the criterion from the neutral starting point. The amount of information accumulated at time t (the value of the process at time t) is a random walk after t/τ steps. This depends obviously on the choice of Δ and τ which can get smaller and smaller. Setting $\Delta = \sqrt{\tau^2}$ and letting τ approach 0 the random walk model converges to a process called the *standard*[3] *Wiener process* which has continuous time set and continuous state space and we will label it $W(t)$. Markov processes with continuous time set and continuous state space are called *diffusion processes* and the standard Wiener process is the simplest diffusion process.

For example, suppose that we have the same criteria (absorbing states) as before, i.e. $\theta = +3$ for a signal response and $-\theta = -3$ for a noise response. Further,

[1] For example, suppose there are two boundaries, θ_S and θ_N, one for choosing S and the other for choosing N and the process starts at z then the probability that the process reaches boundary $\theta_S > z$ before hitting $\theta_N < z$ is given by $Pr(S|\text{process starts at } z) = \frac{\exp(-2\mu z/\sigma^2) - \exp(-2\mu\theta_N/\sigma^2)}{\exp(-2\mu\theta_S/\sigma^2) - \exp(-2\mu\theta_N/\sigma^2)}$ (see e.g., Karlin & Taylor, 1975, p. 361.) μ and σ are defined and explained in the following section.

[2] This is only a convenient choice for which the variance of the process becomes t for all values of τ, see below.

[3] "Standard" means that the probability of moving up or down is 0.5, and this expression is used in the mathematical literature. In psychology this process is often labeled *unbiased process with 0 drift*.

assume that the step size is now $\Delta = 0.5$, and therefore $k = 3/0.5 = 6$. The state space becomes $S = \{-6 \cdot 0.5, -5 \cdot 0.5, -4 \cdot 0.5, -3 \cdot 0.5, -2 \cdot 0.5, -1 \cdot 0.5, 0, +1 \cdot 0.5, +2 \cdot 0.5, +3 \cdot 0.5, +4 \cdot 0.5, +5 \cdot 0.5, +6 \cdot 0.5\}$. For this process it would take, on average, 36 steps of size $\Delta = 0.5$ to reach the criterion θ. This is compared to the 9 steps, when assuming a step size of $\Delta = 1$. However, if we want to relate the expected number of steps to decision time, we have to adjust Eq. (3) by the time unit τ, i.e.,

$$[(E(T|N), E(T|S)] = \tau \cdot [\mathbf{Z} \cdot (\mathbf{I} - \mathbf{Q})^{-2} \cdot \mathbf{R}]./[Pr(N), Pr(S)]. \quad (5)$$

Fig. 1 shows these ideas for $\Delta = 1$, 0.5, and 0.25. For the $\Delta = 0.25$ case we have 4 small steps instead of one big step and within the original one time unit we have 16 subintervals of length τ equal to $1/16$.

The simple random walk and the Wiener process seem to be very similar. Note however a very important theoretical difference. The random walk moves one step up or down from one time unit to the next with probability p and $q = 1 - p$. The increment of the process (a step up or down) from time unit n to $n + 1$ is the difference between two instances of the process, (e.g., Tuckwell, 1995) i.e.,

$$X_{n+1} - X_n = Z_{n+1}, \quad n = 0, 1, 2, \ldots$$

and the Z_n are independent identically distributed with $Pr\{Z_1 = +1\} = p$ and $Pr\{Z_1 = -1\} = q$; and $X_0 = 0$. For $p = q$ it follows that $E(Z_1) = p - q = 0$ and therefore $E(X_n) = E(\sum_{i=1}^{n} Z_i) = nE(Z_1)$. The variance $Var(Z_1) = 4pq = 1$ and therefore $Var(X_n) = n\,Var(Z_1) = n$.

For the approximation of the standard Wiener process we take $X_{t/\tau} = \sum_{i=1}^{t/\tau} Z_i$. Since the Z_i are independent and identically distributed with $Pr[Z_i = +\Delta] = Pr[Z_i = -\Delta] = 0.5$ we get $E(Z_1) = 0$ and $Var(Z_1) = \Delta^2$. Therefore, the expected value is $E(X_{t/\tau}) = 0$ and, with $\Delta = \sqrt{\tau}$, the variance is $Var(X_{t/\tau}) = (t/\tau)\,Var(Z_1) = t\Delta^2/\tau = t$. If we let $\tau \to 0$, then $X_{t/\tau}$ converges in distribution to a random variable $W(t)$ which we called the standard Wiener process. From the central limit theorem we know that $W(t)$ is normally distributed.

Thus, for the standard Wiener process the increment (a very small step moving up or down) in the small time interval $(t, t + \tau)$ is

$$dW(t) = W(t + \tau) - W(t), \quad (6)$$

and $W(t)$ is normally distributed with mean zero and variance t. The *Wiener process with drift* is obtained with some algebraic operations on a standard Wiener process. The random variable $W(t)$ is multiplied by a constant, σ, $\sigma > 0$, and a linear function of time, μt, is added, where μ, can be positive, negative or zero, resulting in

$$V(t) = \mu t + \sigma W(t), \quad (7)$$

or in a difference equation form

$$dV(t) = V(t + \tau) - V(t) = \mu\tau + \sigma dW(t). \quad (8)$$

$V(t)$ is normally distributed with mean μt and variance $\sigma^2 t$.

The Wiener process with drift is related to more interesting cases when $p \neq q \neq 0.5$. We show next how we determine the choice probabilities and choice response time for this situation. To do so we first have to define two functions more precisely.

The properties of an increment $dX(t) = X(t + \tau) - X(t)$ in a small time interval $(t, t + \tau)$ may depend on the time t and the value x of the process at the beginning of the small time interval. Therefore, conditioned on $X(t) = x$, two functions are defined now as infinitesimal mean $\mu(x, t)$ and infinitesimal variance $\sigma^2(x, t)$.

(a) The infinitesimal mean (or infinitesimal first moment or instantaneous mean or drift coefficient) is

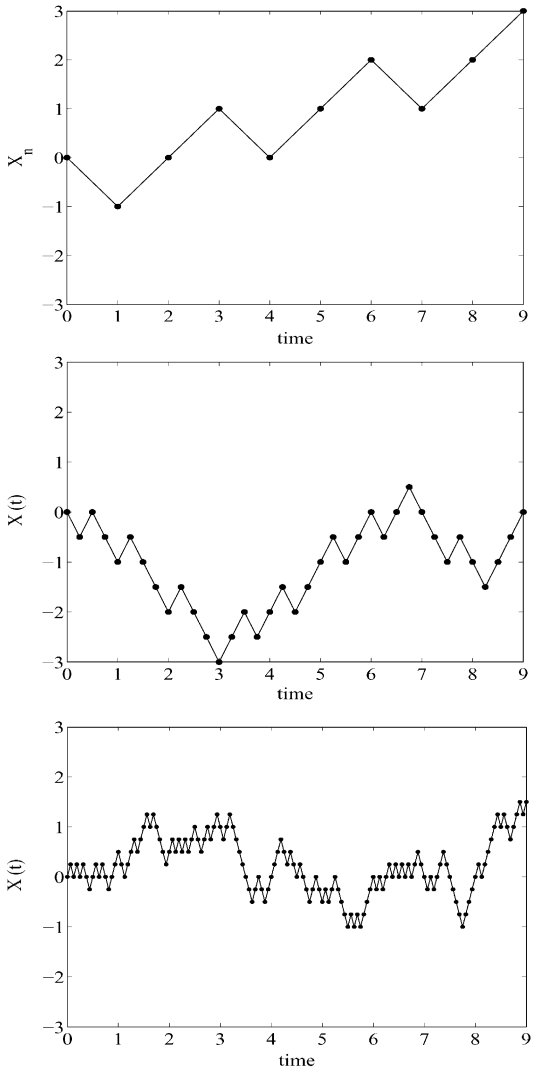

Fig. 1. Sample path for the random walk with step size $\Delta = 1$ (top); $\Delta = 0.5$ (middle); and $\Delta = 0.25$(bottom).

defined as

$$\mu(x,t) = \lim_{\tau \to 0} \frac{E[dX(t)|X(t)=x]}{\tau}.$$

(b) The infinitesimal variance (or infinitesimal second moment or instantaneous variance or diffusion coefficient) is defined as

$$\sigma^2(x,t) = \lim_{\tau \to 0} \frac{E[(dX(t)^2|X(t)=x)]}{\tau}.$$

Note that $\mu(x,t)$ and $\sigma^2(x,t)$ are deterministic functions of x and t. Once the drift and diffusion terms are specified, the transition probability density function can be determined and all required information about the process is provided. These two coefficients are quite important for modeling in psychological tasks, especially the drift coefficient $\mu(x,t)$. For example, a higher intensity stimulus may provide more information than a weaker one and the amount of accumulated information needed to evoke a response may reach the boundary sooner. Therefore, the drift coefficient of the process may be related to stimulus intensity as we will see in more detail below.

In particular, the drift and diffusion coefficients for the Wiener process are $\mu(x,t) = \mu$ and $\sigma^2(x,t) = \sigma^2$, respectively. Now, the transition probabilities $p_{i,i-1}$ and $p_{i,i+1}$ can be written as functions of these coefficients (see the appendix). With $\sqrt{\tau} = \Delta/\sigma$,

$$p_{i,i-1} = \frac{1}{2}\left(1 - \frac{\mu}{\sigma}\sqrt{\tau}\right), \tag{9}$$

$$p_{i,i+1} = \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\tau}\right). \tag{10}$$

These transition probabilities determine the probability matrix in exactly the same way as we have already seen. Note that μ/σ is restricted to the range $(-1/\sqrt{(\tau)}, 1/\sqrt{(\tau)})$ to guarantee that $0 \leqslant p_{i,i-1}$, $p_{i,i+1} \leqslant 1$. For example, with $\mu = 0$, the probability is 0.5 of a step in either direction. The sign of the drift coefficient determines the direction the process takes, on average. For $\mu > 0$ the process moves, on average, up. For $\mu < 0$ the process moves, on average, in the opposite direction. For example, setting σ and step size to one and $\mu = -0.2$ yields $p_{i,i-1} = 0.6$ and $p_{i,i+1} = 0.4$, the transition probabilities for the introductory example. The process stops and a response is initiated as soon as an absorbing state is reached. In particular, "signal", or in general, alternative A, is chosen as soon as $V(t) \geqslant \theta$ and "noise", or in general, alternative B, is chosen as soon as $V(t) \leqslant -\theta$.

To see how well the continuous process can be mimicked by a discrete process we determine the mean and probability as a function of the step size Δ. In Eq. (4) we set the criteria to $\theta = k\Delta$ and $-\theta = -k\Delta$ and expressed the state space as a function of the step size Δ and the criteria, i.e.,

$$S = \{-k\Delta, -(k-1)\Delta, \ldots, -2\Delta, -\Delta, 0, \Delta, 2\Delta, \ldots, (k-1)\Delta, k\Delta\}. \tag{11}$$

The total number of states in S is $m = 2 \cdot k + 1$ or, equivalently, $m = 2 \cdot \theta/\Delta + 1$. Thus, decreasing the step size Δ leads to an increase of the matrix size m, holding all other factors fixed.

Fig. 2 shows the mean and probability for choosing one alternative plotted as a function of the matrix size m for different decision criteria, $\theta = 3$ and $\theta = 5$, with three different drift coefficients μ (see Diederich, 1995, for a different range of parameters).

As can be seen, mean and probability reach an asymptote already for matrix of size 7×7. This corresponds to a step size $\Delta = 1$ for $\theta = 3$ and a step size $\Delta = 1.67$ for $\theta = 5$. Thus, with a relatively small matrix size and a relative large step size the continuous process is mimicked by a discrete process.

4. Birth–death process

Let us stay for another moment with our example. The person is trying to make a decision whether a signal

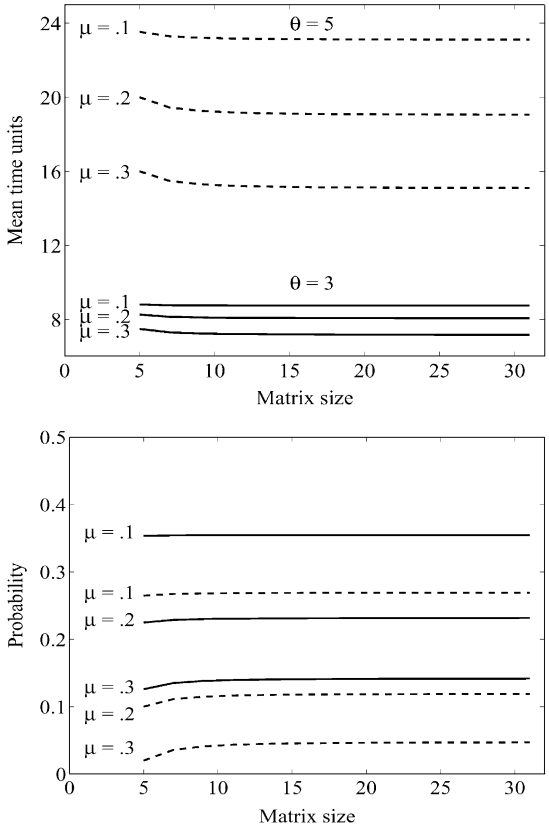

Fig. 2. Mean and probability of the discrete process as a function of the matrix size m.

or noise trial is presented, based on the stimulus presented on a trial. However, suppose that no diagnostic information can be sampled from one moment to the next, i.e., there is a positive probability that the process remains in the same state. Assume that the probability for a signal response is 0.2; the probability for a noise response is 0.3; and the probability for non-diagnostic information is $1-(0.2+0.3)=0.5$. With the same state space as before the transition probability matrix becomes

$$\begin{array}{c|ccccccc} \textit{state} & -3 & -2 & -1 & 0 & +1 & +2 & +3 \\ -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0.3 & 0.5 & 0.2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0.3 & 0.5 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0.5 & 0.2 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0.3 & 0.5 & 0.2 & 0 \\ +2 & 0 & 0 & 0 & 0 & 0.3 & 0.5 & 0.2 \\ +3 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}. \tag{12}$$

Following the same procedure as before, the choice probabilities and conditional expectations for this example are

Initial distribution	$Pr(S)$	$Pr(N)$	$E(T\|S)$	$E(T\|N)$
$\mathbf{Z}=(0\,0\,1\,0\,0)$	0.23	0.77	16.29	16.29
$\mathbf{Z}=(0\,0\,0\,1\,0)$	0.39	0.61	11.86	19.55
$\mathbf{Z}=(0.05\,0.10\,0.70\,0.10\,0.05)$	0.25	0.75	14.51	15.54

This describes a simple random walk that never skips states in its evolution, i.e., a Markov chain with state space $S=\{0,\pm1,\pm2,\ldots\}$ for which transitions from state n can only move to state $n+1$ or to state $n-1$ or stay in the same state. It is called a *birth–death chain*. In addition let the time unit τ between two steps decrease so that a step occurs within the time interval $(t,t+\tau)$. As τ approaches 0 the Markov chain with these properties is called a *birth–death process*. The state space can be seen as the size of a population, the increase of the population size by one as birth and the decrease by one as death. The time set is continuous as τ approaches 0.

These ideas lead to a process that allows us to approximate a broad class of stochastic processes, called *diffusion processes*, one of which is the Wiener process.

5. Ornstein–Uhlenbeck process (OUP)

The transition probabilities p_{ij} of going from state i to state j for the birth–death chain are

$$p_{ij}=\begin{cases}\alpha_i & \text{if } j=i-1,\\ \beta_i & \text{if } j=i+1,\\ 1-\alpha_i-\beta_i & \text{if } j=i,\\ 0 & \text{otherwise.}\end{cases}$$

Note, that the transition probabilities may depend on the state of the process.

Decreasing also the step size Δ for the birth–death chain with transition probabilities p_{ij} approximates a diffusion process $X(t)$ with drift coefficient $\mu(x,t)$ and diffusion coefficient $\sigma^2(x,t)$ as $\Delta\to0$. (See Convergence Theorem, e.g., Bhattacharya and Waymire (1990, p. 387), Karlin and Taylor (1975, p.169), and the appendix).

One of the diffusion processes becomes increasingly interesting for cognitive science research, namely the Ornstein–Uhlenbeck (OU) process, $X(t), t\geqslant0$, with drift. The process, developed in physics, assumes that the *velocity* of a particle is a stochastic process (rather than its position as in the Wiener process). As we will see the OUP has properties that are interesting for model building in psychology. For example, some properties account for assumptions concerning leakage or decay in information accumulation (Busemeyer & Townsend, 1992; Diederich, 1992, 1995, 1997; Smith, 1995; Usher & McClelland, 2001); primacy and recency effects in memory (Busemeyer & Townsend, 1993); approach and avoidance conflicts in decision making (Busemeyer & Townsend, 1993, Diederich, 2003a).

In a small time interval the change in velocity of information accumulation is assumed to be influenced by two factors: the specific tasks with their cognitive interpretation which slows down the motion of the process and some random fluctuation due to new information. Assuming that the damping effect is directly proportional to the velocity, leads to the equation

$$\begin{aligned}X(t+\tau)&=(1-\tau\gamma)X(t)+V(t+\tau)\\&=\tau\delta+(1-\tau\gamma)X(t)+\sigma W(t+\tau)\end{aligned} \tag{13}$$

or in a difference equation form

$$\begin{aligned}dX(t)&=X(t+\tau)-X(t)=-\tau\gamma X(t)+V(t+\tau)\\&=\tau\delta-\tau\gamma X(t)+\sigma W(t+\tau).\end{aligned} \tag{14}$$

$W(t)$ refers to the standard Wiener process (Eq. (6)) and $V(t)$ to the Wiener process with drift δ and diffusion coefficient σ^2 (Eq. (7)), as we discussed above. $X(t)$ is normally distributed with mean $\frac{\delta}{\gamma}(1-e^{-\gamma t})$ and variance $\frac{\sigma^2}{2\gamma}(1-e^{-2\gamma t})$. The drift coefficient of the OUP (more precisely the OUP with drift for $\delta\neq0$) is $\mu(x,t)=\delta-\gamma x$. One part of it, $-\gamma x$, is proportional to the value x of the process and the other part of the drift, δ, is a constant. As before, δ determines the direction of the process' motion and $-\gamma x$ causes the decay of the process depending on the state x in the state space. For example,

if $\gamma>0$ the process takes longer to reach a preset criterion because the increments decrease with increasing position in the state space, i.e. in a sense it slows down the process. This is applicable to model 'leakage of information', or avoidance behavior. For $\gamma<0$, it takes less time for the process to reach a preset criterion since the increments increase with increasing position in the state space. This is applicable for "the winner takes all" assumption; to model facilitation effects; or speeded responses in approach–conflict situations. However, the process might become unstable and, strictly speaking, it is not called OUP for $\gamma<0$.

Fig. 3 shows three sample paths (trajectories) of the standard Wiener process ($\mu=0$); the Wiener process with drift $\mu=0.2$; the OUP with $\delta=0.2$ and $\gamma=0.005$; and the OUP with $\delta=0.2$ and $\gamma=-0.01$. The time units (x-axis) are $t=n\cdot 0.1$ for all four figures. The solid lines are the respective drift rates.

The process stops and a response is initiated as soon as an absorbing state is reached. In particular, alternative A is chosen as soon as $X(t)\geqslant\theta$ and alternative B is chosen as soon as $X(t)\leqslant -\theta$.

As mentioned, once the drift and diffusion terms are specified, the transition probability density function can be determined and all required information about the process is provided. The drift and diffusion coefficients for the OUP are $\mu(x,t)=\delta-\gamma x$ and $\sigma^2(x,t)=\sigma^2$, respectively. The transition probabilities $p_{i,i-1}$, p_{ii}, and $p_{i,i+1}$ can be written as functions of these coefficients exactly as we have done before (see the appendix).

Again, we start with defining the state space of the process, S, which consists of all states of information (states of activation, states of preference). We assume that these entities change by very small steps of size $\Delta>0$. The information accumulation process is bounded by θ and $-\theta$, the criteria to initiate a response for A and B, respectively. Setting $(\theta=k\Delta)$ and $(-\theta=-k\Delta)$ the state spaces can be expressed as a function of the step size Δ and the criteria (Eq. (11)). The total number of states in S is $m=2\cdot k+1$ or, equivalently, $m=2\cdot\theta/\Delta+1$. For convenience the states are indexed by $i=1,\ldots,m$ according to magnitude, i.e.,

$$S=\begin{array}{c}\{-k\Delta,\quad -(k-1)\Delta,\quad \ldots,\quad -\Delta,\quad 0\quad \Delta,\quad \ldots,\quad (k-1)\Delta,\quad k\Delta\}.\\ 1\qquad\quad 2\qquad\quad \ldots\qquad (m-1)/2\qquad \ldots\qquad m-1\qquad m\end{array}$$

As before, the transition probabilities for the m states of the Markov chain are arranged in an $m\times m$ transition probability matrix $\mathbf{P}$, with the elements $p_{11}=1$ and $p_{mm}=1$, and, for $1<i<m$,

$$p_{i,j}=\begin{cases}\frac{1}{2\alpha}\left(1-\frac{\mu(-k\Delta+(i-1)\Delta)}{\sigma^2}\sqrt{\tau}\right) & \text{if } j-i=-1,\\ \frac{1}{2\alpha}\left(1+\frac{\mu(-k\Delta+(i-1)\Delta)}{\sigma^2}\sqrt{\tau}\right) & \text{if } j-i=+1,\\ 1-p_{i,i-1}-p_{i,i+1}=1-\frac{1}{\alpha} & \text{if } j=i,\\ 0 & \text{otherwise}\end{cases}\tag{15}$$

and with $\mu(x)=\delta-\gamma\cdot x$ for the OUP

$$p_{i,j}=\begin{cases}\frac{1}{2\alpha}\left(1-\frac{\delta-\gamma\cdot(-k\Delta+(i-1)\Delta)}{\sigma^2}\sqrt{\tau}\right) & \text{if } j-i=-1,\\ \frac{1}{2\alpha}\left(1+\frac{\delta-\gamma\cdot(-k\Delta+(i-1)\Delta)}{\sigma^2}\sqrt{\tau}\right) & \text{if } j-i=+1,\\ 1-\frac{1}{\alpha} & \text{if } j=i\\ 0 & \text{otherwise.}\end{cases}\tag{16}$$

Here Δ is set to $\Delta=\alpha\sigma\sqrt{\tau}$. The parameter, $\alpha>1$, is a free parameter and is chosen to improve the approximation to a continuous time process. It has no effect on choice probabilities.

The transition probability matrix $\mathbf{P}=||p_{ij}||$ is presented in its canonical form:

$$\mathbf{P}=\left[\begin{array}{c|c}\mathbf{P_I} & \mathbf{0}\\ \hline \mathbf{R} & \mathbf{Q}\end{array}\right]=\begin{array}{c|c|c|ccccc}
 & 1 & m & 2 & 3 & \cdots & m-2 & m-1\\ \hline
1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0\\
m & 0 & 1 & 0 & 0 & \cdots & 0 & 0\\ \hline
2 & p_{21} & 0 & p_{22} & p_{23} & \cdots & 0 & 0\\
3 & 0 & 0 & p_{32} & p_{33} & \cdots & 0 & 0\\
4 & 0 & 0 & 0 & p_{43} & \cdots & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots\\
m-3 & 0 & 0 & 0 & 0 & \cdots & p_{m-3,m-2} & 0\\
m-2 & 0 & 0 & 0 & 0 & \cdots & p_{m-2,m-2} & p_{m-2,m-1}\\
m-1 & 0 & p_{m-1,m} & 0 & 0 & \cdots & p_{m-1,m-2} & p_{m-1,m-1}
\end{array}\tag{17}$$

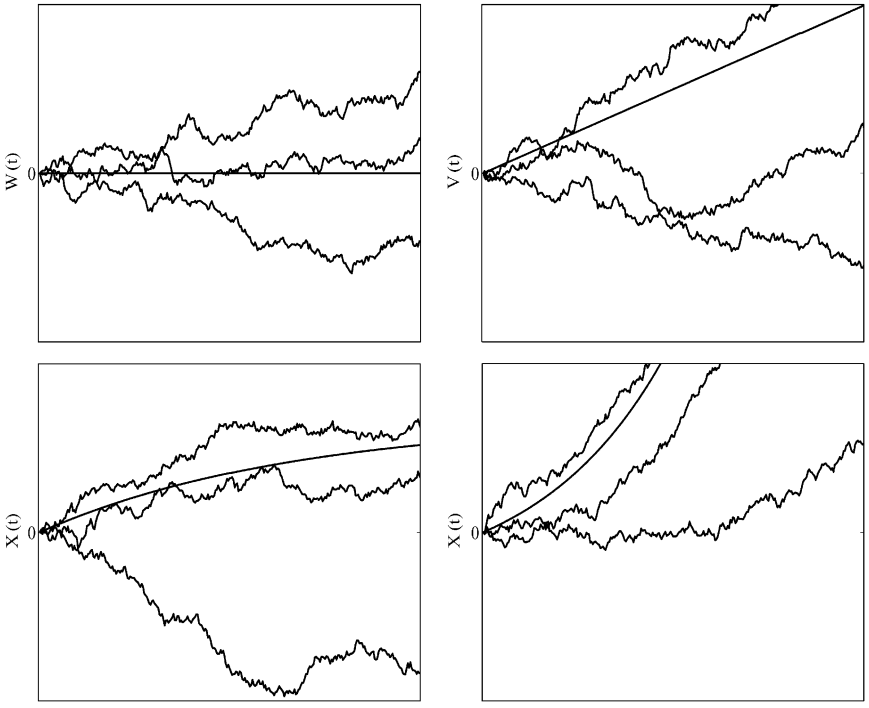

Fig. 3. Three sample paths (trajectories) and drift rate (solid line) of the standard Wiener process (upper left); the Wiener process with drift (upper right); the OUP with $\gamma > 0$ (lower left); and the OUP with $\gamma < 0$ (lower right).

The number of states in the state space determine the size of the transition matrix and its submatrices. In particular, $\mathbf{R}$ is a matrix of size $(m-2) \times 2$ and $\mathbf{Q}$ is of size $(m-2) \times (m-2)$. With these submatrices we proceed as we did before to determine the probability for reaching a particular boundary (Eq. (2)) and the mean time to reach that absorbing boundary (Eq. (5)). Note, that for $\gamma = 0$ the OUP becomes the Wiener process with drift. The first passage time distribution for reaching the boundaries for alternatives A and B can also be determined by using this approach, i.e.,

$$\begin{aligned}&[Pr(T = t|B), Pr(T = t|A)]\\ &\quad = \frac{[Pr(T = t \cap \text{choose } B), Pr(T = t \cap \text{choose } A)]}{[Pr(\text{choose } B), Pr(\text{choose } A)]}\\ &\quad = \mathbf{Z}\mathbf{Q}^{n-1}\mathbf{R}./\mathbf{Z}(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{R}, \qquad (18)\end{aligned}$$

where $t = n\tau$. (The symbol ./ refers to dividing element-wise, see above. $\mathbf{Q}$ to the nth power can be computed easily by decomposing the matrix into its eigenvectors and eigenvalues, as shown in Appendix A.)

To summarize the steps to determine the choice probabilities, the mean decision times, and the probability distribution of decision times:

1. Define the process you want to use, e.g., Wiener, OUP.
2. Depending on your choice, the parameters for the drift and diffusion coefficients are specified. For the Wiener process, these are μ and σ; for the OUP, they are δ, γ, σ. For most applications σ can be set to one since it is only a scaling factor and not observable.
3. Set the decision criterion θ to establish the time constant τ and, for the OUP, α. This gives you immediately the step size Δ and the matrix size m. ($\Delta = \alpha\sigma\sqrt{\tau}; m = 2 \cdot \theta/\Delta + 1$).
4. With these parameters determine the transition probabilities and construct the transition probability matrix $\mathbf{P}$.
5. Partition the matrix $\mathbf{P}$ according to Eq. (1) (Eq. (17)); you only need $\mathbf{R}$ and $\mathbf{Q}$.
6. Decide about the initial distribution, i.e., the probabilities at which states the process starts. These probabilities are assigned to the vector $\mathbf{Z}$. The length of $\mathbf{Z}$ is the same as for the matrix $\mathbf{Q}$.
7. Calculate the choice probabilities (Eq. (2)), mean times (Eq. (5)) and distributions (Eq. (18)) using the vector $\mathbf{Z}$ and the matrices $\mathbf{Q}$ and $\mathbf{R}$.

τ and α are set depending on how good the approximation to the continuous time process is desired. The parameters μ (δ and γ for the OUP), σ and θ can be estimated from the data. Furthermore, depending of the choice of the initial distribution the parameters for this distribution are estimated as well. The minimum number of parameters for the Wiener process is three (drift, boundary, starting point) and for the OUP four. If we assume the special case of symmetric boundaries (no bias), i.e., if the starting point is exactly between the boundaries, we can even reduce the number of parameters to two and three, respectively. Note, the Wiener process with two parameters often cannot account for observed data since it predicts the same

RT for correct and incorrect responses (see e.g., Luce, 1986; Townsend & Ashby, 1983; for a review). However, the OUP with three parameters predicts different RT for correct and incorrect response.

We can extend these models to a broader frame which will be shown next.

6. Diffusion processes with variable drift rates

The unbiased Wiener process (process with symmetric boundaries around the starting position) predicts the same reaction time both for the more frequently and the less frequently chosen alternative, regardless of the drift. However, response time and accuracy data often show a speeded response for the more frequently chosen alternative (see for a review Luce, 1986; Townsend & Ashby, 1983). There are at least three possibilities to modify the Wiener process to account for those data: (1) the Wiener process with bias (e.g., Link & Heath, 1975), i.e., the process does not start half-way between the decision boundaries but closer to either one or with a distribution over starting values as shown for the random walk in the previous sections; (2) the Wiener process with changing drift within a trial (e.g., Diederich, 1992, 1995, 1997; Diederich, 1996, Heath, 1981; Ratcliff, 1981); (3) the Wiener process with variable drift across trials (e.g., Laming, 1968; Ratcliff, 1978, 1981). That is, on each trial the information is accumulated according to a Wiener process with a different drift drawn from a distribution of drifts. How to incorporate these ideas in the matrix approach is shown next. Note that this approach is not limited to a Wiener process but it can be applied to the Ornstein–Uhlenbeck process as well. For simplicity, we keep the illustration limited to the Wiener process.

6.1. A method for generating a distribution of drift rates

A commonly used assumption (cf. Ratcliff, 1978) is that the drift rate is approximately Gaussian distributed with mean $\mu_G = \mu$ and variance σ_G^2. Use an approximation of the Gaussian distribution over discrete intervals and use the mid interval as single drifts $\mu_1, \mu_2 \ldots, \mu_n$. Note that $-1/\sqrt{(\tau)} \leqslant \mu_i/\sigma \leqslant 1\sqrt{(\tau)}$ must be obtained to provide transition probabilities between 0 and 1 (see above). Choose a large number of drift values to form a finite distribution over drift values that closely approximates a Gaussian distribution ($n = 20$ is usually adequate, R.Ratcliff, pers. comm.). For each single drift μ_i determine the probability and expected time for choosing the alternatives N and S of our example. The overall mean and probability for each choice alternative is a mixture of the individual mean and probabilities. Let $Pr(*|\mu_i)$ denote the probability for choosing alternative $*$ (S or N) given drift μ_i, $i = 1, \ldots, n$ and $Pr(\mu_i)$ the probability for μ_i according to the Gaussian distribution. The overall probability for choosing alternative $*$ is the sum of this product, i.e., $Pr(*) = \sum_{i=1}^{n} Pr(*|\mu_i)Pr(\mu_i)$ (according to the law of total probability). The expected time for each choice alternative is the sum over weighted expected time given drift μ_i, i.e., $E[T|*] = \sum_{i=1}^{n} Pr(*|\mu_i)Pr(\mu_i)/Pr(*) \cdot E[T| * \cap \mu_i]$. Table 1 shows how the mean times and choice probabilities change as a function of the number of individual μ_i, $i = 1, \ldots, n$ with $\mu_G = 0.2$, $\sigma_G^2 = ((1/4) \cdot (1/\sqrt{(\tau)} - \mu_G))^2$, and $\theta = 3$ as before.

Table 1
Choice probabilities and conditional expectations for a Wiener process with variable drift

Number of μ_i, n	$Pr(S)$	$Pr(N)$	$E(T\|S)$	$E(T\|N)$
11	0.28	0.72	8.36	7.43
21	0.28	0.72	8.37	7.47
31	0.28	0.72	8.37	7.47

7. Mean response time and response time distribution for simple response time tasks

Up to here we have considered two alternative choice response time tasks and proposed a model to account for choice probabilities, mean choice response times and response time distributions. Consider now a simple response time paradigm in which a subject gives a response as soon as a stimulus is detected.

As before we assume that as soon as a stimulus is presented to the subject a mechanism stochastically accumulates very small amounts of information continuously over time until the *criterion* is reached and a response is initiated. It is assumed that the criterion is set by the subject prior to the stimulus presentation and does not change during the accumulation process.

In contrast to choice response paradigms we only assume a single criterion where the process is absorbed. Further we assume that information never drops below its initial level. That is, we assume that the process is reflected on its lower boundary. This idea fits nicely with neuronal activity dynamics studies in neurobiology (e.g., Ricciardi, 1977). Note, that assuming the reflecting boundary placed far away from the starting position, this boundary has no effect and the process approximates a process with an unrestricted boundary.[4] The state space for the process with one reflecting and one absorbing boundary is

$$S = \{0, \Delta, 2\Delta, \ldots, k\Delta\},$$

[4] The density of the decision latencies to this problem is known as Wald density or inverse Gaussian, see e.g., Luce (1986, pp. 145, 333ff).

with $\theta = k\Delta$ and $m = k + 1$. Here $k\Delta$ is an absorbing state and 0 is a reflecting state. The transition probability matrix (in brackets) is accordingly

index		1	2	3	$\cdots$	$m-2$	$m-1$	m
	state	0	Δ	2Δ	$\cdots$	$(k-2)\Delta$	$(k-1)\Delta$	$k\Delta$
1	0	p_{11}	p_{12}	0	$\cdots$	0	0	0
2	Δ	p_{21}	p_{22}	p_{23}	$\cdots$	0	0	0
3	2Δ	0	p_{32}	p_{33}	$\cdots$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
$m-2$	$(k-2)\Delta$	0	0	0	$\cdots$	$p_{m-2,m-2}$	$p_{m-2,m-1}$	0
$m-1$	$(k-1)\Delta$	0	0	0	0	$p_{m-1,m-2}$	$p_{m-1,m-1}$	$p_{m-1,m}$
m	$k\Delta$	0	0	0	0	0	0	1

Note that there is no absorption at state 0. Further, if the process reaches state 0 the probability is nonzero to make a step Δ. That is what is meant by classifying 0 as a reflecting state.

The transition probability matrix $\mathbf{P} = ||p_{ij}||$ is presented in its canonical form:

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{P_I} & \mathbf{0} \\ \hline \mathbf{R} & \mathbf{Q} \end{array}\right] =$$

	m	1	2	3	$\cdots$	$m-2$	$m-1$
m	1	0	0	0	$\cdots$	0	0
1	0	p_{11}	p_{12}	0	$\cdots$	0	0
2	0	p_{21}	p_{22}	p_{23}	$\cdots$	0	0
3	0	0	p_{32}	p_{33}	$\cdots$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
$m-3$	0	0	0	0	$\cdots$	$p_{m-3,m-2}$	0
$m-2$	0	0	0	0	$\cdots$	$p_{m-2,m-2}$	$p_{m-2,m-1}$
$m-1$	$p_{m-1,m}$	0	0	0	$\cdots$	$p_{m-1,m-2}$	$p_{m-1,m-1}$

With the matrix $\mathbf{Q}$ and the vector $\mathbf{R}$ we proceed as before, i.e. the first passage time distribution for reaching the boundary with $t = n\tau$, $n = 1, 2, \ldots, \infty$, is

$$Pr[T = t] = \mathbf{Z}\mathbf{Q}^{n-1}\mathbf{R}. \tag{19}$$

The expected time for the entire process to reach the boundary is

$$E[T] = \tau\mathbf{Z}(\mathbf{I} - \mathbf{Q})^{-2}\mathbf{R}. \tag{20}$$

Note that the first passage time distribution and the expected value are not conditioned on a choice anymore since we only have one absorbing state, i.e., one response in the simple response time paradigm. The vector $\mathbf{Z}$ is the initial state or initial distribution. For example, assuming that the process starts accumulating information only if a signal is presented it is also reasonable to assume that $\mathbf{Z}$ has a 1 at index 1 and zeros otherwise. This process can be extended to multi-dimensional stimuli. Consider an experimental setup in information about the stimulus changes during one trial. For example, in trials with different stimulus onset times, light followed by a tone, information is presented successively during one trial (Diederich, 1992, 1995). This model is presented in the appendix.

8. Three choice alternatives

Here we develop the ternary choice model for the optional stopping time task. Before we do this consider Eq. (13) again, i.e., the information accumulation process for two choice alternatives was described by the diffusion process

$$X(t + \tau) = (1 - \tau\gamma)X(t) + V(t + \tau) \tag{21}$$

for $\tau \to 0$. The process stops and a response for A is initiated as soon as $X(t) \geqslant \theta$ and a response for B is initiated as soon as $X(t) \leqslant -\theta$. This assumption, however, is equivalent to the following. Assume that the accumulation process is represented by a two-dimensional diffusion process of the form (cf. Heath, 1981)

$$X_1(t+\tau) = [(1-\tau\gamma_1)X_1(t) - \tau\gamma_2 X_2(t)] + V_1(t+\tau) \quad (22)$$

for choosing A and

$$X_2(t+\tau) = [-\tau\gamma_2 X_1(t) + (1-\tau\gamma_1)X_2(t)] + V_2(t+\tau) \quad (23)$$

for choosing B, with $V_1 + V_2 = 0$.

Writing these equations in matrix form we get

$$\begin{bmatrix} X_1(t+\tau) \\ X_2(t+\tau) \end{bmatrix} = \begin{bmatrix} (1-\tau\gamma_1) & -\tau\gamma_2 \\ -\tau\gamma_2 & (1-\tau\gamma_1) \end{bmatrix} \cdot \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} + \begin{bmatrix} V_1(t+\tau) \\ V_2(t+\tau) \end{bmatrix},$$

$$\mathbf{X}(t+\tau) = [\mathbf{I} - \tau\boldsymbol{\Gamma}] \cdot \mathbf{X}(t) + \mathbf{V}(t+\tau), \quad (24)$$

where $\mathbf{I}$ is the identity matrix and $\boldsymbol{\Gamma}$ the matrix containing the γ's. The diagonal elements of $[\mathbf{I} - \tau\boldsymbol{\Gamma}]$ provide memory for previous states of the system. The off-diagonal values allow for competitive interactions among competing alternatives.

Note that if the sum $X_1(0) + X_2(0) = 0$ then $X_1(t) + X_2(t) = 0 \Leftrightarrow X_1(t) = -X_2(t)$ for all t, and therefore the processes in Eqs. (22) and (23) are represented by the single process in Eq. (21) with $\gamma = (\gamma_2 - \gamma_1)$.

For a choice from a set $\{A, B, C\}$ let us first assume three processes, one for each choice alternative, similar to Eq. (24). The change in evidence within the interval $(t, t+\tau)$ is

$$d\mathbf{X}(t) = \mathbf{X}(t+\tau) - \mathbf{X}(t) = -\tau\boldsymbol{\Gamma}(t)\mathbf{X}(t) + \mathbf{V}(t+\tau).$$

The moments of the change in evidence are given by

$$E[d\mathbf{X}(t)|\mathbf{X}(t) = \mathbf{x}] = \tau\mu(\mathbf{x}) = \tau(\boldsymbol{\delta} - \boldsymbol{\Gamma}\mathbf{x}) = \tau \begin{bmatrix} \delta_1 - (\gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3) \\ \delta_2 - (\gamma_4 x_1 + \gamma_5 x_2 + \gamma_6 x_3) \\ \delta_3 - (\gamma_7 x_1 + \gamma_8 x_2 + \gamma_9 x_3) \end{bmatrix} \quad (25)$$

and

$$Cov[d\mathbf{X}(t)|\mathbf{X}(t) = \mathbf{x}] = \tau\boldsymbol{\Sigma} = \tau \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}. \quad (26)$$

If we restrict the matrix $\boldsymbol{\Gamma}$ to the following: (1) the elements in the diagonal are the same; (2) the matrix is symmetric; (3) the columns sum up to the same constant, i.e.

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_2 \\ \gamma_2 & \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_2 & \gamma_1 \end{bmatrix}$$

and if we assume that $V_1 + V_2 + V_3 = 0$ and $X_1(0) + X_2(0) + X_3(0) = 0$, then $X_3(t) = -[X_1(t) + X_2(t)]$. In this case, option A is chosen as soon as $X_1(t) \geqslant \theta$, option B is chosen as soon as $X_2(t) \geqslant \theta$, and option C is chosen as soon as $X_3(t) = -[X_1(t) + X_2(t)] \geqslant \theta$. Thus, the diffusion process is restricted to a two-dimensional simplex (triangular plane). The following developments are limited to diffusions on this simplex. Markov chains can be used to approximate multidimensional diffusion processes (see Stroock & Varadhan, 1979, Theorem 11.2.3). In the case of the diffusion on the simplex, we approximate the stochastic change in evidence $d\mathbf{X}(t)$ by a random increment, $[dX_1(t), dX_2(t), dX_3(t)]$. On each step, we assume that these three increments sum to zero, so that only the first two coordinates need to be included in the model. Each coordinate, $dX_i(t)$, is a small positive step $\Delta = \alpha\sigma\sqrt{\tau}$ or a small negative step $-\Delta = -\alpha\sigma\sqrt{\tau}$, or zero; and we set $\sigma = \sqrt{(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)/2}$ (i.e., the diffusion coefficients for $X_1(t), X_2(t)$, and $X_3(t)$), and we require $\alpha \geqslant 1$. The state space of the Markov chain is defined by a set of equally spaced discrete states that form a grid of points on the simplex (see Fig. 4):

$$\mathbf{S} = \{[i\Delta, j\Delta], (i = -(k+j), \ldots, k), (j = -2k, -(2k-1), \ldots, -1, 0, 1, \ldots, k)\}.$$

The total number of states is therefore $m = (3k+2)(3k+1)/2$. The threshold bound for option A is defined by the right vertical line of points $[\theta = k\Delta, j\Delta]$ for $j = -2k, -(2k-1)\ldots, k$; the threshold bound for option B is defined by the top horizontal line of points $[i\Delta, \theta = k\Delta]$ for $i = -2k, -(2k-1)\ldots, k$; and the threshold bound for option C is defined by the negative diagonal line of points $[i\Delta, -(k+i)\Delta]$ for $i = -2k, -(2k-1)\ldots, k$. The remaining $((3k+1)-3)((3k+1)-2)/2$ states represent transient states. The transition probabilities for the transient states are

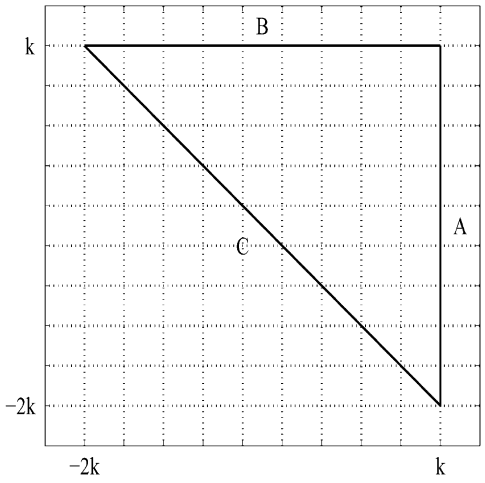

Fig. 4. The state space $\mathbf{S}$ of the Markov chain is defined by a set of equally spaced discrete states that form a grid of points on the simplex.

symbolized as

$$Pr[dX_1(t) = +\Delta, dX_2(t) = 0|\mathbf{X}(t) = \mathbf{x}] = p_{+0}(\mathbf{x}),$$
$$Pr[dX_1(t) = +\Delta, dX_2(t) = -\Delta|\mathbf{X}(t) = \mathbf{x}] = p_{+-}(\mathbf{x}),$$
$$Pr[dX_1(t) = 0, dX_2(t) = +\Delta|\mathbf{X}(t) = \mathbf{x}] = p_{0+}(\mathbf{x}),$$
$$Pr[dX_1(t) = 0, dX_2(t) = -\Delta|\mathbf{X}(t) = \mathbf{x}] = p_{0-}(\mathbf{x}),$$
$$Pr[dX_1(t) = -\Delta, dX_2(t) = 0|\mathbf{X}(t) = \mathbf{x}] = p_{-0}(\mathbf{x}),$$
$$Pr[dX_1(t) = -\Delta, dX_2(t) = +\Delta|\mathbf{X}(t) = \mathbf{x}] = p_{-+}(\mathbf{x}),$$

and we define $p_{00}(\mathbf{x}) = 1 - [p_{+0}(\mathbf{x}) + p_{+-}(\mathbf{x}) + p_{0+}(\mathbf{x}) + p_{0-}(\mathbf{x}) + p_{-0}(\mathbf{x}) + p_{-+}(\mathbf{x})]$ as the probability of no change in state. Note that we assume $p_{++}(\mathbf{x}) = p_{--}(\mathbf{x}) = 0$, that is no simultaneous changes in the same direction. The transition probabilities are defined by equating the moments of the discrete and continuous state processes shown in the appendix. Solving the equations yields the transition probabilities for each transient state:

$$p_{+0}(\mathbf{x}) = \tau\left(\frac{\mu_1(\mathbf{x})}{2\Delta} + \frac{\sigma_1^2}{2\Delta^2} + \frac{\sigma_{21}}{\Delta^2}\right) + \beta, \tag{27}$$

$$p_{+-}(\mathbf{x}) = -\tau\frac{\sigma_{21}}{\Delta^2} - \beta, \tag{28}$$

$$p_{0+}(\mathbf{x}) = \tau\left(\frac{\mu_2(\mathbf{x})}{2\Delta} + \frac{\sigma_2^2}{2\Delta^2}\right) - \beta, \tag{29}$$

$$p_{0-}(\mathbf{x}) = -\tau\left(\frac{\mu_2(\mathbf{x})}{2\Delta} - \frac{\sigma_2^2}{2\Delta^2} - \frac{\sigma_{21}}{\Delta^2}\right) + \beta, \tag{30}$$

$$p_{-0}(\mathbf{x}) = -\tau\left(\frac{\mu_1(\mathbf{x})}{2\Delta} - \frac{\sigma_1^2}{2\Delta^2}\right) - \beta, \tag{31}$$

$$p_{-+}(\mathbf{x}) = \beta. \tag{32}$$

The parameter β reflects indeterminacy in this linear system of equations. Any value of β will reproduce the moments of the continuous state model. The only constraints on β are that $\beta > 0$, it yields transition probabilities between zero and one, and the transition probabilities sum to less than or equal to one (thus allowing for $p_{00}(\mathbf{x}) \geqslant 0$). Constraints produced by the moments of the third coordinate do not provide linearly independent equations. The solutions for $p_{-+}(\mathbf{x})$ and $p_{+-}(\mathbf{x})$ require $\sigma_{21} < 0$, which will generally be true as a consequence of the fact that the evidence sums to zero. The transition probabilities defined in Eqs. (27)–(32) are then inserted into the Markov chain model to approximate the continuous state diffusion process.

Consider the following example. With $k = 3$ we have $m = (3k + 2)(3k + 1)/2 = 55$ states that are arranged as a grid of points on a simplex in the following way:

	−6	−5	−4	−3	−2	−1	0	1	2	3
3	C|B	B	B	B	B	B	B	B	B	B|A
2		C	22	23	24	25	26	27	28	A
1			C	16	17	18	19	20	21	A
0				C	11	12	13	14	15	A
−1					C	7	8	9	10	A
−2						C	4	5	6	A
−3							C	2	3	A
−4								C	1	A
−5									C	A
−6										C|A

Analogous to the binary choice problem, the transition probabilities defined in Eqs. (27)–(32) are arranged into four matrices: the first is a $(3k + 1 - 3)(3k + 1 - 2)/2 \times (3k + 1 - 3)(3k + 1 - 2)/2$ matrix $\mathbf{Q}$ containing the probabilities of transiting from one transient state to another. For our example with $k = 3$ we have a 28×28 matrix (first index: right(+)/left(−); second index: up(+)/down(−)), i.e.,

$$\mathbf{Q} = \begin{array}{c|cccccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & \cdots & 24 & 25 & 26 & 27 & 28 \\ \hline 1 & p_{00} & p_{-+} & p_{0+} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 2 & p_{+-} & p_{00} & p_{+0} & p_{-+} & p_{0+} & p_{++} & \cdots & 0 & 0 & 0 & 0 & 0 \\ 3 & p_{0-} & p_{-0} & p_{00} & 0 & p_{-+} & p_{0+} & \cdots & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & p_{+-} & 0 & p_{00} & p_{+0} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & p_{0-} & p_{+-} & p_{-0} & p_{00} & p_{+0} & \cdots & 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & p_{--} & p_{0-} & 0 & p_{-0} & p_{00} & \cdots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 24 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & p_{00} & p_{+0} & 0 & 0 & 0 \\ 25 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & p_{-0} & p_{00} & p_{+0} & 0 & 0 \\ 26 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & p_{-0} & p_{00} & p_{+0} & 0 \\ 27 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & p_{-0} & p_{00} & p_{+0} \\ 28 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & p_{-0} & p_{00} \end{array} \tag{33}$$

The second is a $(3k - 2)(3k - 1)/2 \times 1$ vector $\mathbf{R}_A$ containing the probabilities of transiting from a transient state to an absorbing state for option A; the third is a $(3k - 2)(3k - 1)/2 \times 1$ vector $\mathbf{R}_B$ containing the probabilities of transiting from a transient state to an absorbing state for option B; and the fourth is a $(3k - 2)(3k - 1)/2 \times 1$ vector $\mathbf{R}_C$ containing the probabilities of transiting from a transient state to an absorbing state for option C. Let $p_{A_i}, p_{B_i}, p_{C_i}$ denote the probabilities for entering the absorbing states

for A, B, and C, respectively from row i. Then

$$\mathbf{R}_A = \begin{array}{c|c} index & \\ \hline 1 & p_{A_1} \\ 2 & 0 \\ 3 & p_{A_3} \\ \vdots & 0 \\ 6 & p_{A_6} \\ \vdots & 0 \\ 10 & p_{A_{10}} \\ \vdots & 0 \\ 15 & p_{A_{15}} \\ \vdots & 0 \\ 21 & p_{A_{21}} \\ \vdots & 0 \\ 28 & p_{A_{28}} \end{array}, \quad \mathbf{R}_B = \begin{array}{c|c} index & \\ \hline 1 & 0 \\ \vdots & 0 \\ 21 & 0 \\ 22 & p_{B_{22}} \\ 23 & p_{B_{23}} \\ 24 & p_{B_{24}} \\ 25 & p_{B_{25}} \\ 26 & p_{B_{26}} \\ 27 & p_{B_{27}} \\ 28 & p_{B_{28}} \end{array}, \quad \mathbf{R}_C = \begin{array}{c|c} index & \\ \hline 1 & p_{C_1} \\ 2 & p_{C_2} \\ 3 & 0 \\ 4 & p_{C_4} \\ \vdots & 0 \\ 7 & p_{C_7} \\ \vdots & 0 \\ 11 & p_{C_{11}} \\ \vdots & 0 \\ 16 & p_{C_{16}} \\ \vdots & 0 \\ 22 & p_{C_{22}} \\ \vdots & 0 \\ 28 & 0 \end{array} \tag{34}$$

As before, the $\mathbf{Z}$ vector contains the initial distribution and is of length $(3k-2)(3k-1)/2$. The neutral starting position, for example, is determined by assigning a one at index $(2k-1)(2k)/2-(k-1)$.

After making these assignments, Eqs. (2) and (18) are used to calculate the choice probabilities and mean decision times for each of the three options.

For example, let $\mu_1 = 0.5$ and $\mu_2 = 0.3$. ($\mu_3 = -(\mu_1+\mu_2) = -0.8$) and the covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 4 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 4 \end{bmatrix}.$$

Set $\tau = 0.01$ and $\alpha = 1$. Thus the step size becomes $\Delta = \alpha\sqrt{(\sigma_1^2+\sigma_2^2+\sigma_3^2)/2\tau} = 1\cdot\sqrt{3\cdot 4/2\cdot 0.01}$. Inserting these values in Eqs. (27)–(32) and further assuming that $p_{+-} = p_{-+}$ and $p_{00} = 0$ yields the transition probabilities $p_{+0} = 0.1769$, $p_{+-} = 0.1667$, $p_{0+} = 0.1728$, $p_{0-} = 0.1605$, $p_{-0} = 0.1565$, $p_{-+} = 0.1667$. With these probabilities the respective matrices are constructed. Let the process start at a neutral position, i.e., $Z_{13} = 1$. Then $Pr(A) = 0.3636$; $Pr(B) = 0.3508$; and $Pr(C) = 0.2856$. The expected times are $E(T|A) = 0.0897$; $E(T|B) = 0.0897$; and $E(T|C) = 0.0897$.

Note, that this can also be used to compute the distribution statistics for the leaky competing accumulator model recently proposed by Usher and McClelland (2001).

9. Concluding remarks

Diffusion processes are very powerful approaches to model human information processing in a variety of psychological tasks. We believe that lack of mathematical tractability has prevented broad applications of these models to empirical data. We hope that this tutorial helps to spread this approach by showing how to construct models, how to implement them on a computer, and how to calculate predictions made by these models.

A few words about why we chose the Markov chain approximation approach. As we have mentioned above that once the drift and diffusion coefficient are specified, all the information about the process can be determined via the transition probability density function. Traditionally, the transition probability density function can be found by solving partial differential equations known as forward equation (Fokker–Planck equation) and backward equation (Chapman–Kolmogorov equation). We did not elaborate on these equations since the major concern of this tutorial was just to avoid them and not to scare users away. Instead, we used the discrete state and time transition probabilities as approximations to the continuous transition densities. We only include basic mathematics in the tutorial. Second, for processes with boundaries, which are of special interest for psychologists, often no closed solutions are available when trying to solve partial differential equations. Third, any type of simulation of continuous time processes usually requires digital computers. That means all calculations are carried out discretely. Thus,

the difference between continuous time and space models and the matrix approach is no longer valid. Fourth, following the discussion of Thompson (1999), differential–integral–difference equation type of formulations are an invention of the 19th and early 20th century in lack of powerful computers. Researchers had to content themselves with these formulations as means for summary information about the effect of trillions of gas molecules.

The main advantage of the Markov chain approach, as compared to Smith's (2000) approach, is simplicity. For the Markov chain approach, the same simple matrix equations (e.g., Eqs. (2) and (3)) can be used for any model, and there is no need to derive any new formulas when transition probability assumptions are changed. Smith's (2000) approach requires the derivation of new formulas for the distribution statistics for each new diffusion process. However, the main advantage of Smith's (2000) approach is that it can be used to model continuous changes in drift rate over time, whereas the Markov chain methods are limited to a finite number of changes in drift rate (Diederich, 1995, 1997, 2003a, b).

Note that both the Markov chain approach and Smith's (2000) approach require the use of finite approximations to a continuous process since both approaches make use of digital computers.

Acknowledgments

This research was supported by Deutsche Forschungsgemeinschaft Grants Di 506/5-1 and Di 506/6-1.

Appendix A. Definitions

A.1. Stochastic process

Given an index set I, a *stochastic process* indexed by I is a collection of random variables $\{X_\lambda : \lambda \in I\}$ on a probability space taking values in a set S, the *state space* of the process. The state space in this context contains possible information states of the process and may be discrete or continuous. The index set or parameter set, often also called time space, contains the indices for which the process is defined. Again, the index set may be discrete or continuous. For example, information may be accumulated at discrete points in time or continuously over time.

A.2. Transition probability matrix

A *transition probability matrix* is a square matrix $\mathbf{P} = (p_{ij})$, where i and j vary over a finite enumerable set S, satisfying $p_{ij} \geqslant 0$ for all i and j and $\sum_{j\in S} p_{ij} = 1$ for all i. The set S is called the *state space* and its elements are *states*.

A.3. Deriving Eqs. (2) and (3)

The probability of choosing P at time step n is

$$Pr(P \text{ at } n) = \mathbf{Z} \cdot \mathbf{Q}^n \cdot \mathbf{R}_P,$$

where $\mathbf{R}_P$ refers to the vector of matrix $\mathbf{R}$ with probabilities that lead to the absorbing state for alternative P. The probability of choosing P is obtained by summing up over discrete time, i.e.,

$$Pr(P) = \mathbf{Z} \cdot \sum_{n=0}^{\infty} \mathbf{Q}^n \cdot \mathbf{R}_P.$$

Realizing that

$$\sum_{n=0}^{\infty} \mathbf{Q}^n = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots = (\mathbf{I} - \mathbf{Q})^{-1}$$

as $n \to \infty$ (for a proof see, e.g., Bhat, 1984, p. 71) yields

$$Pr(P) = \mathbf{Z} \cdot (\mathbf{I} - \mathbf{Q})^{-1} \cdot \mathbf{R}_P.$$

If we define T as the random number of time steps to reach the boundary then the first moment for the distribution of time steps to choose P is

$$E[T|P] = \frac{\mathbf{Z} \sum_{n=1}^{\infty} n \cdot \mathbf{Q}^{n-1} \cdot \mathbf{R}_P}{Pr(P)}.$$

Realizing that

$$\sum_{n=1}^{\infty} n \cdot \mathbf{Q}^{n-1} = \mathbf{I} + 2\mathbf{Q} + 3\mathbf{Q}^2 + \cdots = (\mathbf{I} - \mathbf{Q})^{-2}$$

as $n \to \infty$ (for a proof see, e.g. Diederich, 1995, Appendix) yields

$$E[T|P] = \frac{\mathbf{Z}(\mathbf{I} - \mathbf{Q})^{-2} \cdot \mathbf{R}_P}{Pr(P)}.$$

A.4. Approximation of the diffusion process

Suppose there are two real-valued functions $\mu(x)$ and $\sigma^2(x)$ on $(-\infty, \infty)$ that are continuously differentiable and that σ'' exists and is continuous, and that $\sigma^2(x) > 0$ for all x. Also assume that $\mu(x)$ and $\sigma^2(x)$ are bounded. Consider now a birth–death chain with state space $S = \{0, \pm\Delta, \pm 2\Delta, \ldots\}$ with step size $\Delta > 0$ and having transition probabilities p_{ij} of going from $i\Delta$ to $j\Delta$ in one step given by

$$p_{i,i-1} = \alpha_i^{(\Delta)} := \frac{\sigma^2(i\Delta)\tau}{2\Delta^2} - \frac{\mu(i\Delta)\tau}{2\Delta},$$

$$p_{i,i+1} = \beta_i^{(\Delta)} := \frac{\sigma^2(i\Delta)\tau}{2\Delta^2} + \frac{\mu(i\Delta)\tau}{2\Delta},$$

$$p_{ii} = 1 - \frac{\sigma^2(i\Delta)\tau}{\Delta^2} = 1 - \alpha_i^{(\Delta)} - \beta_i^{(\Delta)},$$

where τ is given by

$$\tau = \frac{\Delta^2}{\sup_x \sigma^2(x)}$$

and represents the actual time in between two successive transitions. With the conditions imposed on $\mu(x)$ and $\sigma^2(x)$ a sufficiently small Δ guarantees the non-negativity of the transition probabilities $p_{ii}, p_{i,i-1}, p_{i,i+1}$ (see Bhattacharya & Waymire, 1990, p. 386).

Given that the process is at $x = i\Delta$, the mean displacement in a single step of size Δ in time τ is

$$\Delta\beta_i^{(\Delta)} + (-\Delta)\alpha_i^{(\Delta)} = \mu(i\Delta)\tau = \mu(x)\tau. \tag{A.1}$$

Therefore, the instantaneous rate of mean displacement per unit time, when the process is at x, is $\mu(x)$. The mean squared displacement in a single step of size Δ in time τ is

$$\Delta^2\beta_i^{(\Delta)} + (-\Delta)^2\alpha_i^{(\Delta)} = \sigma^2(i\Delta)\tau = \sigma^2(x)\tau. \tag{A.2}$$

That is, the instantaneous rate of mean squared displacement per unit time, when the process is at x, is $\sigma^2(x)$.

Hence, the transition probabilities $p_{i,i-1}$ and $p_{i,i+1}$ can be written as functions of the drift and diffusion coefficient $\mu(x)$ and $\sigma^2(x)$, respectively. With $\Delta^2 = \tau\sigma^2(x)$

$$p_{i,i-1} = \frac{1}{2}\left(1 - \frac{\mu(x)}{\sigma(x)}\sqrt{\tau}\right), \tag{A.3}$$

$$p_{i,i+1} = \frac{1}{2}\left(1 + \frac{\mu(x)}{\sigma(x)}\sqrt{\tau}\right). \tag{A.4}$$

Thus, the birth–death chain with transition probabilities p_{ij} approximates the diffusion process with drift coefficient $\mu(x)$ and diffusion coefficient $\sigma^2(x)$ as $\Delta \to 0$ (see Convergence Theorem, e.g. Bhattacharya & Waymire, 1990, p. 387; Karlin & Taylor, 1981, p. 169).

To see how a differential equation can be approximated by a difference equation let $p_{ij}^{(n)}$ be the n-step transition probabilities going from i to j. By definition $p_{ij}^{(n)} = \sum_{k\in S} p_{ik}p_{ik}^{n-1}$, $n = 2, 3, \ldots$ and S is the state space. Therefore, with $p_{ii} = 1 - \beta_i^{(\Delta)} - \alpha_i^{(\Delta)}$, we get for the birth–death process

$$\begin{aligned} p_{ij}^{(n+1)} &= p_{ii}p_{ij}^{(n)} + p_{i,i+1}p_{i+1,j}^{(n)} + p_{i,i-1}p_{i-1,j}^{(n)}, \\ p_{ij}^{(n+1)} &= (1 - \beta_i^{(\Delta)} - \alpha_i^{(\Delta)})p_{ij}^{(n)} + \beta_i^{(\Delta)}p_{i+1,i}^{(n)} + \alpha_i^{(\Delta)}p_{i-1,i}^{(n)} \end{aligned}$$

or

$$\begin{aligned} p_{ij}^{(n+1)} - p_{ij}^{(n)} &= \beta_i^{(\Delta)}(p_{i+1,j}^{(n)} - p_{ij}^{(n)}) - \alpha_i^{(\Delta)}(p_{ij}^{(n)} - p_{i-1,j}^{(n)}) \\ &= \frac{\mu(i\Delta)\tau}{2\Delta}\left((p_{i+1,j}^{(n)} - p_{ij}^{(n)}) + (p_{ij}^{(n)} - p_{i-1,j}^{(n)})\right) \\ &\quad + \frac{1}{2}\frac{\sigma^2(i\Delta)\tau}{\Delta^2}(p_{i+1,j}^{(n)} - 2p_{ij}^{(n)} + p_{i-1,j}^{(n)}). \end{aligned}$$

With $t = n\tau$ and states $x = i\Delta$ and $y = j\Delta$ and $\Delta \to 0$, defining as approximate density

$$p^{(\Delta)}(n\tau; i\Delta, j\Delta) := \frac{p_{ij}^{(n)}}{\Delta},$$

we get

$$\begin{aligned} &p^{(\Delta)}((n+1)\tau; i\Delta, j\Delta) - p^{(\Delta)}(n\tau; i\Delta, j\Delta)) \\ &\quad = \mu(i\Delta)(p^{(\Delta)}(n\tau; (i+1)\Delta, j\Delta) \\ &\qquad - p^{(\Delta)}(n\tau; (i-1)\Delta, j\Delta))\tau/2\Delta \\ &\qquad + \tfrac{1}{2}\sigma^2(i\Delta)(p^{(\Delta)}(n\tau; (i+1)\Delta, j\Delta) - 2p^{(\Delta)}(n\tau; i\Delta, j\Delta) \\ &\qquad + p^{(\Delta)}(n\tau; (i-1)\Delta, j\Delta))\tau/\Delta^2 \end{aligned}$$

which is the difference equation version of the partial differential equation

$$\frac{\partial}{\partial t}p(t; x, y) = \mu(x)\frac{\partial}{\partial x}p(t; x, y) + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}p(t; x, y),$$
$$\text{for } t > 0, \quad -\infty < x, \quad y < \infty.$$

That is, the transition probability density $p(t; x, y)$ can be computed approximately by determining the n-step transition probabilities $p_{ij}^{(n)}$. This can be done by constructing a transition probability matrix and raising it to the nth power shown below.

A.5. Decomposing Q

To raise $\mathbf{Q}$ to the nth power it is convenient to decompose $\mathbf{Q}$ into its eigenvectors and eigenvalues. Proceed as follows:

1. Determine the eigenvalues and eigenvectors of $\mathbf{Q}$, a $k \times k$ matrix, by using a mathematical program, e.g., MATLAB with $[\mathbf{X}, \mathbf{v}] = eig(\mathbf{Q})$. The program solves the determinant equation $|\mathbf{vI} - \mathbf{Q}| = \mathbf{0}$ to get the v eigenvalues $\mathbf{v}$ and solves $\mathbf{QX} = \mathbf{vX}$ to get the k eigenvectors of length k, i.e., $\mathbf{X} = ||\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k||$.
2. Determine the inverse of $\mathbf{X}$, i.e., $\mathbf{X}^{-1}$, e.g., $inv\ \mathbf{X} = inv(\mathbf{X})$.
3. $\mathbf{Q}^n$ can be written as

   $$\mathbf{Q}^n = \mathbf{X}\mathbf{V}^n\mathbf{X}^{-1}$$

   where

   $$\mathbf{V}^n = \begin{bmatrix} v_1^n & & 0 \\ & \ddots & \\ 0 & & v_k^n \end{bmatrix}.$$

4. Conveniently the MATLAB program gives already the eigenvalues in form of a diagonal matrix. Thus

   $$\mathbf{Q}^n = \mathbf{Xv}.^n * inv\ \mathbf{X},$$

   where $.^n$ means elementwise raising to the nth power.

A.6. Three alternatives

The transition probabilities are defined by equating the moments of the discrete and continuous state

processes:

$$E[dX_1(t)|X(t)=x]/\tau = [p_{+0}(x)+p_{+-}(x)-p_{-0}(x) - p_{-+}(x)]\Delta/\tau = \mu_1(x),$$

$$E[dX_2(t)|X(t)=x]/\tau = [p_{0+}(x)+p_{-+}(x)-p_{0-}(x) - p_{+-}(x)]\Delta/\tau = \mu_2(x),$$

$$Var[dX_1(t)|X(t)=x]/\tau = [p_{+0}(x)+p_{+-}(x)+p_{-0}(x) p_{-+}(x)]\Delta^2/\tau - \mu_1(x)^2\tau = \sigma_{11}^2,$$

$$Var[dX_2(t)|X(t)=x]/\tau = [p_{0+}(x)+p_{-+}(x)+p_{0-}(x) + p_{+-}(x)]\Delta^2/\tau - \mu_2(x)^2\tau = \sigma_{22}^2,$$

$$Cov[dX_1(t), dX_2(t)|X(t)=x] = -[p_{-+}(x)+p_{+-}(x)]\Delta^2/\tau - \mu_1(x)\mu_2(x)\tau = \sigma_{21}.$$

As $\tau \to 0$, then $\mu_i(x)\mu_j(x)\tau \to 0$, and so they can be ignored for small τ.

Appendix B. Fixed t_*

The simplest case how to combine different attributes is to assume that at a particular *point in time* attention switches from one attribute of the stimulus to the next and, therefore, the information accumulation process for the first attribute transforms into an information accumulation process for the second attribute *at time* t_*. t_* may have to be estimated from the data or is determined by the experimental setup (e.g., SOA). That is, the basic assumption proposed here is that attributes are processed serially. Each attribute is considered for a certain amount of time and generates an information accumulation process for making a decision. The subprocesses are concatenated sequentially over time.

As before we assume that as soon as the stimulus is presented, an information accumulation process is initiated accumulating very small amounts of information in very small time intervals. After some time t_* new information is available, e.g., the subject has switched attention to another attribute of the stimulus, or new information is externally added. The information accumulation process may change, i.e., the second subprocess (after t_*) may have a different drift and/or diffusion coefficient than the first subprocess. Therefore, the entire process may reach the boundary, on average, sooner, or later, or may take an opposite direction towards the alternative boundary.

The entire process is considered piecewise up to time t_*; and then starting from there with a new process, we determine the probability of the second part. Eq. ?? shows that a different drift coefficient influences the **Q** and **R** matrices. Note, that the entire process is not time-homogeneous any more, i.e., the process changes at time t_*.

Some notion has to be introduced. Note that the first column of **R** is associated with choosing alternative B, the second column with choosing A. For convenience we label these vectors as $\mathbf{R}_A$ and $\mathbf{R}_B$, respectively. To keep it simpler we show how to calculate the quantities for choosing a particular alternative, say, A. The calculations for B are analogous. Further, the subprocesses are indicated by 1,2, etc. Thus, the **Q** matrix for the first subprocess is labeled $\mathbf{Q}_1$, for the second subprocess $\mathbf{Q}_1$. The **R** vector associated with choosing alternative A for the first process is labeled $\mathbf{R}_{A_1}$, the one for the second process $\mathbf{R}_{A_2}$.

The probability that the process reaches the boundary for choosing A before time t_*, $t_* = n_1\tau$, is

$$Pr[T \leqslant t_* \cap \text{ choose } A] = \mathbf{Z}\sum_{i=1}^{n_1} \mathbf{Q}_1^{i-1}\mathbf{R}_{A_1}. \tag{B.1}$$

Z contains the probability of the starting position of the process. $\mathbf{Q}_1$ denotes the transition probability matrix for non-absorbing (transient) states for the first part of the process up to time t_*. The vector $\mathbf{R}_{A_1}$ contains the transition probabilities from the transient states to absorbing states for the first part of the process.

The probability that the process reaches the boundary for choosing A after time $t_* = n_1\tau$ is

$$Pr[t_* < T \cap \text{ choose } A] = \mathbf{Z}\mathbf{Q}_1^{n_1}\sum_{i=n_1+1}^{\infty} \mathbf{Q}_2^{i-(n_1+1)}\mathbf{R}_{A_2}. \tag{B.2}$$

$\mathbf{Z}\mathbf{Q}_1^{n_1}$ contains the defective initial distribution (that is, it does not add up to 1) for the second part of the process. $\mathbf{Q}_2$ and $\mathbf{R}_{A_2}$ contain the respective probabilities for the second part of the process.

The probability that the process reaches the boundary for choosing A is therefore

$$Pr[\text{choose } A] = \mathbf{Z}\sum_{i=1}^{n_1} \mathbf{Q}_1^{i-1}\mathbf{R}_{A_1} + \mathbf{Z}\mathbf{Q}_1^{n_1}\sum_{i=n_1+1}^{\infty} \mathbf{Q}_2^{i-(n_1+1)}\mathbf{R}_{A_2}, \tag{B.3}$$

The first passage time distribution for the A boundary is determined as

$$Pr[T=t|\text{choose } A] = \begin{cases} \mathbf{Z}\mathbf{Q}_1^{n-1}\mathbf{R}_{A_1}/Pr[\text{choose } A] & : \text{for } t = n\tau \leqslant t_* = n_1\tau, \\ & \quad n = 1, 2, \ldots, n_1 \\ \mathbf{Z}\mathbf{Q}_1^{n_1}\mathbf{Q}_2^{n-(n_1+1)}\mathbf{R}_{A_2}/Pr[\text{choose } A] & : \text{for } t = n\tau > t_* = n_1\tau, \\ & \quad n = n_1+1, \ldots . \end{cases} \tag{B.4}$$

The mean time for the process to reach the boundary for choosing A is

$$E[T|\text{choose } A] = \tau\left[\mathbf{Z}\sum_{i=1}^{n_1} i\mathbf{Q}_1^{i-1}\mathbf{R}_{A_1} + \mathbf{Z}\mathbf{Q}_1^{n_1}\sum_{i=n_1+1}^{\infty} i\mathbf{Q}_2^{i-(n_1+1)}\mathbf{R}_{A_2}\right] \Big/ Pr[\text{choose } A]. \quad \text{(B.5)}$$

The probability, the first passage time distribution, and the mean time of the process for choosing B are determined by substituting vector $\mathbf{R}_A$ by vector $\mathbf{R}_B$, containing the probability from the transient to the absorbing state of the boundary for B.

Appendix C. Matlab programs

In the following we use sparse matrices to run programs with large matrices. The MATLAB command "SPARSE" converts a matrix to sparse form by squeezing out any zero elements.

C.1. Simple random walk

```
clear;
% Simple Random walk
%criterion value
theta = 3;
%total number of states
m = 2 * theta + 1;
% Probabilities
p = 0.6;
q = 1 - p;
% Constructing the transition probability matrix
for i = 1 : m - 2
i = i + 1;
j = [(i - 1)  i  (i + 1)];
tm(i - 1, j) = [p  (1 - (p + q))  q];
end
Q = sparse(tm(:, 2 : m - 1));
R = sparse([tm(:, 1)    tm(:, m)]);
% three different initial distribution
Z1 = [0 0 1 0 0];
Z2 = [0 0 0 1 0];
Z3 = [0.05 0.1 0.7 0.1 0.05];
% Identity matrix
I = speye(m - 2);
% Choice probabilities
PrBPrA = Z1 * (I - Q) ∧ (-1) * R
% mean decision times
EtBEtA = (Z1 * (I - Q) ∧ (-2) * R)./PrBPrA
```

C.2. OUP

```
% Ornstein–Uhlenbeck process for two choice alternatives
clear;
% model parameters
% criterion value
theta = 4;
% drift rate
delta = 0.2;
% decay
gamma = 0.01;
% diffusion coefficient
sigma = 1;
% parameters set to approximate continuous process
% continuous time adjustment
alpha = 1.5;
% time unit
tau = 0.01;
% step size
Delta = alpha * sqrt(sigma) * sqrt(tau);
%total number of states
m = 2 * round(theta/Delta) + 1;
%states – exclusive the two absorbing boundaries
x = -(m - 1)/2 : (m - 1)/2;

%transition probabilities
p1 = 1/(2 * alpha) * (1 - (delta - gamma * x) *
sqrt(tau)/sigma);
p2 = 1/(2 * alpha) * (1 + (delta - gamma * x) *
sqrt(tau)/sigma);
p3 = 1 - (1/alpha) - 0 * x;

% constructing the transition probability matrix
for i = 1 : m - 2
i = i + 1;
j = [(i - 1)  i  (i + 1)];
tm(i - 1, j) = [p1(i)  p3(i)  p2(i)];
end

Q = sparse(tm(:, 2 : m - 1));
R = sparse([tm(:, 1)    tm(:, m)]);
% initial distribution, here the process starts unbiased
with probability 1
Z1 = zeros(1, m - 2);
Z1(1, (m - 3)/2 + 1) = 1;
% identity matrix
I = speye(m - 2);
% choice probabilities
PrBPrA = Z1 * inv(I - Q) * R
% mean decision times
EtBEtA = tau * (Z1 * inv(I - Q) * inv(I_Q) *
R)./PrBPrA
```

References

Aschenbrenner, K. M., Albert, D., & Schmalhofer, F. (1984). Stochastic choice heuristics. *Acta Psychologica*, *56*, 153–166.

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

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*, 310–329.

Audley, R. J. (1960). A stochastic model for individual choice behavior. *The Psychological Review*, *67*(1), 1–15.

Audley, R. J., & Pike, A. R. (1965). Some alternative stochastic models of choice. *The British Journal of Mathematical and Statistical Psychology*, *18*(2), 207–225.

Bhat, U. N. (1984). *Elements of applied stochastic processes*. New York: Wiley.

Bhattacharya, R. N., & Waymire, E. C. (1990). *Stochastic processes with applications*. Wiley: New York.

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, *43*, 345–370.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, *23*(3), 255–282.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision-making in an uncertain environment. *Psychological Review*, *100*(3), 432–459.

Diederich, A. (1992). *Intersensory facilitation: Race, superposition, and diffusion models for reaction time to multiple stimuli*. Frankfurt: Verlag Peter Lang.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*(2), 197–215.

Diederich, A. (1996). Multi-attribute Dynamic Decision Model: A Cognitive Dynamic Approach for Multiattribute Binary Choice Tasks. Universität Oldenburg: Habilitationsschrift.

Diederich, A. (1997). Dynamic stochastic models for decision making with time constraints. *Journal of Mathematical Psychology*, *41*(3), 260–274.

Diederich, A. (2003a). Choice under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin and Review*, to appear.

Diederich, A. (2003b). MDFT account of decision making under time pressure. *Psychonomic Bulletin and Review*, to appear.

Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. *Psychological Science*, *10*(4), 353–359.

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.

Heath, R. A. (1981). A tandem random walk model for psychological discrimination. *British Journal of Mathematical and Statistical Psychology*, *34*, 76–92.

Karlin, S., & Taylor, H. M. (1975). *A first course in stochastic processes*, Vol. I. New York: Academic Press.

Karlin, S., & Taylor, H. M. (1981). *A second course in stochastic processes*. New York: Academic Press.

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

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

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

Pacut, A. (1980). Mathematical modelling of reaction latency: The structure of the models and its motivation. *Acta Neurobiologiae Experimentalis*, *40*, 199–215.

Pike, A. R. (1966). Stochastic models of choice behavior: Response probabilities and latencies of finite Markov chain systems. *British Journal of Mathematical and Statistical Psychology*, *19*, 15–32.

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

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

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

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, *92*(2), 212–225.

Ratcliff, R. (1988). Continuous versus discrete information processing: modeling accumulation of partial information. *Psychological Review*, *95*(2), 238–255.

Ratcliff, R., & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 16–36.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127–140.

Ratcliff, R., VanZandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*(2), 261–300.

Ricciardi, L. M. (1977). *Lecture notes in biomathematics*. Berlin: Springer.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multi-alternative decision field theory. A dynamic connectionist model of decision making. *Psychological Review*, *108*, 370–392.

Schwarz, W. (1994). Diffusion, superposition, and the redundant-target effect. *Journal of Mathematical Psychology*, *38*, 504–520.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*(3), 408–463.

Smith, P. L. (1995). Psychophysical principled models of visual simple reaction time. *Psychological Review*, *102*, 567–591.

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

Stroock, D. W., & Varadhan, S. R. S. (1979). *Multidimensional diffusion processes*. Berlin: Springer.

Thompson, J. R. (1999). *Simulation: A modeler's approach*. New York: Wiley.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

Tuckwell, H. C. (1995). *Elementary applications of probability theory* (2nd ed.). London: Chapman & Hall.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550–592.

Van Zandt, T., Colonius, H., & Proctor, R. (2000). A comparison of two reaction time models applied to perceptual matching. *Psychonomic Bulletin & Review*, *7*, 208–256.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*, 37–58.