A Comparison of Sequential Sampling Models for Two-Choice Reaction Time

Roger Ratcliff
The Ohio State University

Philip L. Smith
University of Melbourne

The authors evaluated 4 sequential sampling models for 2-choice decisions—the Wiener diffusion, Ornstein–Uhlenbeck (OU) diffusion, accumulator, and Poisson counter models— by fitting them to the response time (RT) distributions and accuracy data from 3 experiments. Each of the models was augmented with assumptions of variability across trials in the rate of accumulation of evidence from stimuli, the values of response criteria, and the value of base RT across trials. Although there was substantial model mimicry, empirical conditions were identified under which the models make discriminably different predictions. The best accounts of the data were provided by the Wiener diffusion model, the OU model with small-to-moderate decay, and the accumulator model with long-tailed (exponential) distributions of criteria, although the last was unable to produce error RTs shorter than correct RTs. The relationship between these models and 3 recent, neurally inspired models was also examined.

A common feature of many tasks studied by experimental psychologists is that they involve a simple decision about some feature of a stimulus that is expressed as a choice between two alternative responses. Because decisions of this type are so fundamental to theory development and evaluation, their study has been an important part of cognitive psychology for many years.

Among the models that have been proposed to account for simple two-choice decisions, sequential sampling models are unique in providing a way to understand both the speed and accuracy of performance within a common theoretical framework. These models are based on the premise that the representation of stimuli in the central nervous system is inherently variable or noisy and to make a decision about a stimulus, one must accumulate successive samples of this noisy stimulus representation until a criterion quantity of evidence is obtained. The particular criterion that is attained determines which of the two responses is made; the time taken to attain it determines the response time (RT).

In the sequential sampling framework, performance in an experimental task depends on two main factors: the quality of the information derived from processing the stimulus and the quantity of information required before a response is made. The quality of the information from the stimulus depends jointly on the objective properties of the stimulus and on the inherent variability of the stimulus processing mechanisms in the central nervous system. The quantity of information required for a response can be controlled by the subject, who can adjust the decision criteria. The interaction of the quality of the information and decision criteria allows sequential sampling models to account for the main relationship between accuracy and RT in two-choice decisions: RTs are longer and accuracy is lower in response to more difficult stimuli than in response to less difficult stimuli (Luce, 1986; Pachella, 1974). In contrast, there are many models that provide an account of either RT or accuracy, but not of the relationship between them. For example, stage theory models (Townsend & Ashby, 1983; Sternberg, 1969) provide an account only of RT; signal detection theory models (Green & Swets, 1966) provide an account only of accuracy.

Sequential sampling models provide a straightforward account of the speed–accuracy trade-off phenomena that are often observed in cognitive tasks (Luce, 1986; Wickelgren, 1977). Although such effects are widespread, neither stage theories nor signal detection theories provide a way of explaining why they occur or a way of predicting their magnitude. Sequential sampling models attribute speed–accuracy trade-off effects to changes in the amount of evidence needed for a response, represented in the models by changes in the values of the decision criteria.

Sequential sampling models also provide precise quantitative predictions of the relationships between mean RTs and the probabilities of correct responses and errors and of the shapes of the associated RT distributions. Because different sequential sampling mechanisms predict different relationships among these features of performance, it should be possible to determine which models can account for experimental data reasonably well and which models can be ruled out.

In this article, we begin by evaluating the four most developed of the sequential sampling models for two-choice decisions. We carry out a detailed qualitative investigation of the RT and accu-

racy properties of the models and perform comparative fits of the models to three sets of experimental data, examining whether, and under what circumstances, the models mimic each other. We then compare the best-fitting models to a recent model by Usher and McClelland (2001) and two new models closely related to Usher and McClelland's model. These newer models combine features of the various sequential sampling models, and they have been argued to be more compatible with neurally inspired theoretical frameworks.

Sequential Sampling Models

Within the sequential sampling framework, models may differ on whether evidence is sampled and accumulated at discrete equally spaced time intervals, discrete randomly spaced time intervals, or continuously through time; whether it is accrued in fixed-sized chunks or chunks of varying sizes; and whether the decision is based on an absolute stopping rule, such that the amount of evidence must reach a particular criterion value for one or the other of the response alternatives, or a relative stopping rule, such that the evidence for one of the alternatives must exceed the other by a criterion amount. Variations on these dimensions produce the range of models shown in Figure 1, which exhibits relationships among the main models that have been historically influential in the sequential sampling literature. We examine the four models in bold first in this article.

The division between a relative stopping rule and an absolute stopping rule appeared early in the evolution of sequential sampling models. The general class of models with a relative stopping rule is labeled *random walk models.* In these models, developed by Stone (1960), Laming (1968), and Link and Heath (1975; Link, 1975), evidence from the stimulus in favor of one response alter-

native is evidence against the other alternative. The amount of evidence accumulated at each interval is sampled from a continuous distribution at equally spaced, discrete time steps. More recent investigations have focused on diffusion process models, in which evidence accumulates continuously in time. In the Wiener diffusion model, the rate of accumulation of evidence is constant, and in the Ornstein–Uhlenbeck model, the rate of accumulation decreases as the amount of accumulated evidence increases.

The most influential of the early models with an absolute stopping rule was LaBerge's (1962) recruitment model. In this model, evidence in favor of one alternative is accumulated in one response counter, and evidence in favor of the other alternative is accumulated in another response counter. The counter that first reaches its criterion amount of evidence determines the response. The stopping rule is termed *absolute* because an increase in the amount of evidence for one response does not change the amount of evidence for the other. This model has the serious problem that it cannot correctly predict the shapes of RT distributions. Two successors to this model have been proposed: the accumulator model and the Poisson counter model.[1] In the accumulator model, evidence arrives at discrete, equally spaced time steps, and the amount that arrives at each step is sampled from a continuous distribution (like the diffusion models; Smith & Vickers, 1988; Vickers, 1970, 1978, 1979; Vickers, Caudrey, & Willson, 1971). In the Poisson counter model, the same amount of evidence is accumulated at each step, and it arrives at times sampled from a continuous distribution (LaBerge, 1994; Pike, 1966, 1973; Townsend & Ashby, 1983).

Despite the large amount of research on sequential sampling models (see Luce, 1986; Townsend & Ashby, 1983; and Vickers, 1979, for reviews), there has been a lack of systematic comparative evaluations of the classes of models shown in Figure 1. The earliest comparison was by Vickers et al. (1971), who compared the behavior of correct and error RTs and accuracy across several models, including random walk, recruitment, and accumulator models. Apart from this study, most published articles have evaluated either a single model (see Luce, 1986) or, occasionally, have compared a pair of competing models (e.g., Smith & Vickers, 1989; Van Zandt, Colonius, & Proctor, 2000). Moreover, many of the evaluations have focused on a restricted range of properties such as the relationship between mean RTs for correct responses and errors or the relationship between mean RT and response probability. The early investigations yielded many useful insights, but their main results were to reject the simplest versions of the
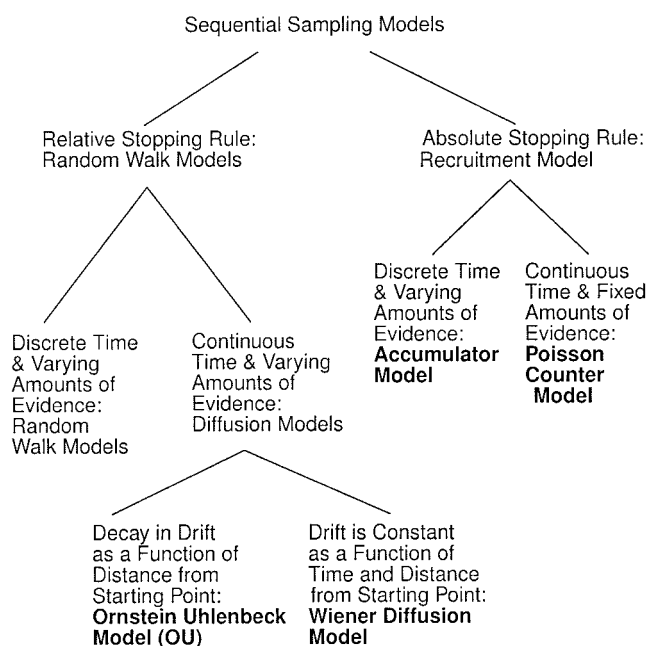

Sequential Sampling Models

Relative Stopping Rule: Random Walk Models

Absolute Stopping Rule: Recruitment Model

Discrete Time & Varying Amounts of Evidence: Random Walk Models

Continuous Time & Varying Amounts of Evidence: Diffusion Models

Discrete Time & Varying Amounts of Evidence: **Accumulator Model**

Continuous Time & Fixed Amounts of Evidence: **Poisson Counter Model**

Decay in Drift as a Function of Distance from Starting Point: **Ornstein Uhlenbeck Model (OU)**

Drift is Constant as a Function of Time and Distance from Starting Point: **Wiener Diffusion Model**

Figure 1. The relationship between the various stochastic reaction time models. The models evaluated in this article are in bold.

---

[1] Historically, there has been some ambiguity in terminology in relation to models of this class. The name *accumulator model* was introduced by Audley and Pike (1965) to refer to the discrete-time, unit-increment model that LaBerge (1962) had previously called the *recruitment model.* In Vickers's (1970) original presentation of his model, he referred to it as "an accumulator model" to indicate the model was one member of a general class. However, subsequent usage has tended to follow LaBerge and call the unit-increment model a recruitment model, leaving the title *the accumulator model* (with definite article) for Vickers's model. The general class of models in which evidence for the two responses accrues in parallel, of which these models are both members, is usually referred to as *counter models.* This is the terminology we have used, although it is not an accurate description of the accumulator model because the underlying stochastic process is not a counting process.

sequential sampling models. As discussed in detail later, the simplest sequential sampling models predict overly restrictive relationships between mean RTs for correct responses and errors, which are inconsistent with experimental data. Van Zandt et al.'s results were somewhat different, and we discuss these in the penultimate section.

We have restricted our evaluation to models in which the rate of accumulation of the evidence derived from the stimulus and the amount of information required for a decision are both stationary, that is, they do not change over the time course of an experimental trial. We have imposed these restrictions because the class of nonstationary models is potentially very large and because nonstationarity gives a considerable amount of model freedom. Also, nonstationary assumptions can be best tested with experimental manipulations in which either the amount or the kind of stimulus information entering the decision process is varied over the time course of evidence accumulation, a type of data outside the scope of this article (e.g., Diederich, 1995, 1997; Heath, 1992; Ratcliff, 1980; Ratcliff & McKoon, 1982; Ratcliff & Rouder, 2000; Smith, 1995, 1998, 2000; Smith & Van Zandt, 2000). We have also restricted our evaluation to two-choice tasks that are rapid, one-process decisions (e.g., less than 1,000–1,500-ms mean RT at a maximum). Slower decisions can induce multiple or repeated decision processes, which are currently outside the domain of application of the models examined here.

The Data To Be Explained

An attractive feature of sequential sampling models is their ability to deal with RT and accuracy data simultaneously. The main data in two-choice tasks are the proportions of correct and error responses, their mean RTs, and the shapes of their RT distributions. In themselves, these are difficult for a model to fit, but the real challenge comes in explaining how they jointly vary as a function of experimental independent variables. There are two major types of experimental manipulations that are used to test the models.

The first type is any *within-block* manipulation that affects task difficulty but does not allow a change in decision criteria. For example, in lexical decision tasks, low-frequency words are more difficult than high-frequency words. If both kinds of words are mixed together within a block of trials, subjects cannot know at the beginning of a trial whether the test item is a high- or low-frequency word (or a nonword); so, they cannot set their criteria on the basis of frequency.

Within-block manipulations of difficulty are especially constraining for the models when they lead to large changes in accuracy across conditions. The models have to fit the changes in accuracy and concomitant changes in mean RTs for correct and error responses and also fit the shapes of the RT distributions over the whole range of accuracy values. Moreover, they must do this with only a single parameter, the rate of accumulation of evidence, varying. Generally, as the difficulty of the stimuli increases, accuracy decreases and RTs for correct responses increase. The shapes of the RT distributions are positively skewed, with most of the increase in mean RT for correct responses coming from an increase in the skew of the RT distribution coupled with a much smaller increase in the leading edge of the distribution. Depending on the experimental manipulation, RTs for errors are sometimes

shorter than RTs for correct responses, sometimes longer, and sometimes there is a crossover in which errors are slower than correct responses when accuracy is low and faster than correct responses when accuracy is high. The models must be capable of capturing all these aspects of a data set.

The second type of manipulation is a *between-blocks* manipulation, that is, a manipulation that allows changes in decision criteria between blocks of trials. For example, in speed–accuracy manipulations, when the instructions for a block of trials emphasize speed, RTs are shorter, often by several hundred milliseconds, and responses are less accurate, often by a few percentage points, compared with blocks for which accuracy is emphasized. According to the sequential sampling models, subjects adapt to this manipulation by adjusting their decision criteria such that they require less evidence in conditions for which speed is stressed than in conditions for which accuracy is stressed. Another example of a between-blocks manipulation is one in which the proportion of stimuli for which a particular response is correct is varied; for example, the proportion of words and nonwords in a lexical decision experiment is varied. This manipulation favors one response over the other, and in the models, the decision criterion for the favored alternative is set lower than the other criterion. Between-blocks manipulations are less constraining for the models than within-block manipulations because the decision criteria, and other criteria to be discussed later, can be changed between conditions.

Overall, there are four main aspects of data against which the models are tested: the positively skewed shapes of the RT distributions, the effects of experimental variables on the leading edges and degrees of skewness of the distributions, the accuracy rates, and the relative speeds of correct and error responses. All of these aspects of data must be accommodated simultaneously by the models, for both within-block and between-blocks manipulations.

Variability in Processing and Criteria Across Trials

The earliest sequential sampling models, such as the simple random walk model and the recruitment model, had serious limitations. For example, the random walk model with two stimuli with equal and opposite accumulation rates made the prediction that the mean RT for a given response made correctly was the same as the mean RT for that response made in error, whereas data usually show unequal RTs. The recruitment model predicted negatively skewed RT distributions in some cases, and it predicted that RT distributions should become more symmetrical as criteria increase, whereas empirical RT distributions are positively skewed and become more skewed as criteria increase. The solution to problems like these came with the incorporation into the models of variability in cognitive processing across trials and the realization that the inclusion of variability has strong, unexpected consequences for the behaviors of the models.

The assumption of trial-to-trial variability in processing is a cornerstone of a number of models in psychology. In signal detection theory, signal and noise strengths are assumed to vary across trials. In some applications, across-trial variability in the criterion setting has also been considered. However, in signal detection theory, the differences between stimulus variability and criterion variability have typically not been emphasized because their effects on performance are mathematically equivalent, that is,

they are not separately identifiable. If the signal, noise, and criterion distributions are all normally distributed, then the combined effects of the criterion and a stimulus can be represented by a single normal distribution whose variance is equal to the sum of the variances of the stimulus distribution and the criterion distribution; so, the components cannot be separated. Between-trials variability in the representations of study items is also fundamental for the global memory models (Gillund & Shiffrin, 1984; Hintzman, 1986; Murdock, 1982).

In signal detection theory and in the global memory models, adding new sources of variability would change the models only in minor ways. In contrast, in sequential sampling models, different sources of variability and different combinations of sources of variability have qualitatively different effects on predictions compared with the models without such variability. One source of across-trial variability in the sequential sampling models is the same as for signal detection theory and the global memory models, namely, across-trial variability in the information extracted from nominally equivalent stimuli. In sequential sampling models, this is represented by variation in the rate at which evidence accumulates toward one or the other of the response criteria. For some of the models, across-trial variability is also found in the values of the response criteria, and for other models, there is an equivalent source of variability in the position of the starting point of accumulation of evidence. For all the sources of across-trial variability, the primary motivation is the belief that subjects cannot set components of cognitive processing at exactly the same values from trial to trial.

The assumption that starting points or decision criteria and accumulation rates vary across trials is consistent with a view that comes from the literature on sequential effects (Falmagne, 1965, 1968; Ollman, 1966; Ratcliff, 1985; Ratcliff, Van Zandt, & McKoon, 1999; Remington, 1969; see also Luce, 1986, and Kirby, 1980, for reviews) and from the related literature on error monitoring (Rabbitt, 1979). These literatures suggest that trial-by-trial variation in RT is partly determined by the prior stimulus and the prior response. Adaptive regulatory mechanisms to account for trial-by-trial effects have been proposed by a number of investigators, including Laming (1969), Rabbitt and Rogers (1977), and Vickers (1978).

Without multiple sources of across-trial variability, it is clear that the sequential sampling models cannot fit experimental data. For an early random walk model, Laming (1968) showed that variability in the starting point was necessary to predict shorter error RTs than correct RTs, which is a frequent finding in choice RT paradigms. For the Wiener diffusion model, Ratcliff (1978) showed that variability in the accumulation rate enabled it to predict longer error RTs than correct RTs, as is typical in recognition memory paradigms (Ratcliff, 1978). Smith and Vickers (1988) used variability in both criteria and accumulation rate for the accumulator model, and Ratcliff et al. (1999; Ratcliff & Rouder, 1998) used variability in accumulation rate and starting point to fit crossover patterns in which errors are sometimes faster than correct responses and sometimes slower, in the same experiment. More generally, Van Zandt and Ratcliff (1995) have argued that some methods developed to discriminate between model architectures fail when variability in processing is allowed across trials; in fact, sometimes models that failed tests passed them when variability in processing across trials was added to the models. For

each of the models we evaluate, we provide the same qualitative sources of between-trials variability. For some of the models, some of the sources of between-trials variability have not previously been explored or evaluated against empirical data.

Overview

Our aim in this article is to carry out a systematic evaluation of the most recent forms of traditional sequential sampling models and an evaluation of three new neurally inspired models. In the sections below, we present the four traditional models: Ratcliff's Wiener diffusion model (Ratcliff, 1978, 1980, 1981, 1985, 1988; Ratcliff & Rouder, 1998, 2000; Ratcliff et al., 1999), the Ornstein–Uhlenbeck (OU) model (Busemeyer & Townsend, 1992, 1993; Roe, Busemeyer, & Townsend, 2001; Smith, 1995), the accumulator model (Smith & Vickers, 1988; Vickers, 1970; Vickers et al., 1971), and the Poisson counter model (Pike, 1966, 1973; Townsend & Ashby, 1983). Following presentation of the models, we apply them to three sets of experimental data, investigating which of the models provide adequate accounts of the data sets and whether and under what circumstances the models are empirically distinguishable from one another. The first issue has two parts: Which models provide the correct qualitative patterns of performance, and which of these provide adequate and approximately equivalent quantitative fits to the data? On the basis of these evaluations, we arrive at conclusions about which of the four traditional models provide the best current account of two-choice RT data. We then present the neurally inspired models and evaluate whether they also provide adequate quantitative fits to the data.

Random Walk Models

As mentioned above, the early random walk models had the problem that errors and correct responses were predicted to have equal RTs. Early solutions to this problem were aimed at producing errors faster than correct responses, a pattern frequently observed in the choice RT tasks that were a major focus of research at that time. Laming (1968) introduced variability in starting point to allow predictions of errors faster than correct responses, which he motivated by supposing that on some trials subjects begin the decision process before the stimulus is presented, which results in variability around the starting point when the stimulus finally becomes available for processing.

To enable predictions of either errors faster than correct responses or errors slower than correct responses, Link and Heath (1975; Link, 1975) proposed that the distribution of increments to the random walk could vary with changes in the experimental task. Although this provided a formal solution to the problem of ordering mean error RTs and correct RTs, it did not address the shapes of RT distributions. Also, Link and Heath (1975) did not develop an explanation of how the properties of the increment distribution depended on the particular task being modeled.

Early random walk models also had the problem that they predicted that accuracy grows without bound as response boundaries are moved away from the starting point. In other words, by moving the boundaries far from the starting point, a subject can approach close to perfect accuracy. Ratcliff (1978) avoided this problem in the Wiener diffusion model with the introduction of

variability in the rate of accumulation of evidence across trials. This was introduced because, it was argued, it is implausible that the information content of a studied item in a recognition memory experiment from, for example, Study Position 10, would be identical on every trial of an experiment. Across-trial variability in the rate of accumulation of evidence also has the effect of limiting the growth of accuracy as a function of time in response signal and deadline experiments (Ratcliff, 1978, but see 1988), and it allows the model to predict longer error RTs than correct RTs.

The methods that are used to derive predictions for the early discrete random walk models yield results that are only approximate. As the time steps are made small, the approximation becomes better, and it becomes exact when evidence is accumulated continuously in time. Smith (1990a) showed that in the case for which the amount of evidence at each step is sampled from a normal distribution, the use of such methods is formally equivalent to approximating a discrete-time random walk model with a continuous-time diffusion process. Because the similarities between the two classes of model are sufficiently great, we do not believe it to be possible to distinguish between them on the basis of experimental data (see Smith, 1990a), and we prefer to work with the continuous-time diffusion models.

Wiener Diffusion Model

The Wiener diffusion model is depicted in the top panel of Figure 2. Noisy information is accumulated continuously over time from a starting point z to decision criteria (response boundaries) at 0 and a. As shown in Figure 2, the path of the amount of accumulated evidence varies over time during the course of a trial; its mean (illustrated with the arrow in the top panel of the figure) is called drift rate, ξ, and its variance is s^2 (termed the *diffusion coefficient*). The path is a highly irregular function, illustrated by the three sample paths in Figure 2, that results from the cumulative effect of a large number of small, independent statistical perturbations. If the response boundaries are removed, the population of sample paths is normally distributed with mean (ξt) and variance ($s^2 t$) that increase linearly with time. The parameter s is a scaling parameter for the model; that is, if the parameter were doubled, other parameters of the model could be doubled to produce exactly the same predictions. In applications of this model reported by Ratcliff (e.g., Ratcliff, 1978, 1988, 2002; Ratcliff & Rouder, 1998, 2000; Ratcliff, Thapar, & McKoon, 2001; Ratcliff et al., 1999) and in the applications presented later in this article, s is set to a fixed value, 0.1.

Because of the variability (s) in the path of evidence accumulation, decision processes with the same drift rate ξ hit the boundaries at different times, and a decision process that drifts toward one boundary can hit the wrong boundary by mistake, producing an error. If the response boundaries are moved farther away from the starting point, the probability that a process that is drifting toward the correct response boundary will hit the other boundary by chance is reduced, thus increasing accuracy (and RT).

The drift rate for stimuli in difficult conditions is smaller than the drift rate for stimuli in easier conditions, and a smaller drift rate results in longer RTs and a decrease in accuracy because processes are more likely to hit the wrong boundary. RT distributions are predicted to be right skewed by the geometry of the decision

Diffusion and OU Models
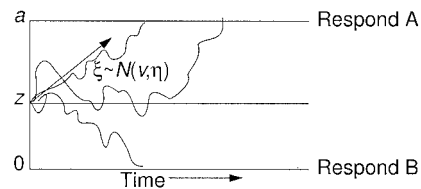
Parameters of the Diffusion and OU Models
1. a Boundary separation
2. z Starting point
3. T_{er} Mean of the nondecision component of RT
4. s_t Range in rectangular distrib. of nondecision component of RT
5. η S.D. in drift rate (ξ) across trials (normal distribution)
6. s_z Range in starting point across trials (uniform distribution)
7. s S.D. in mean drift rate within a trial (scaling parameter = 0.1)
8. v Mean drift rate across trials (one for each condition)

In Addition, for the OU Model,
9. β = decay term; size of decay=$-\beta\times$(distance from z).
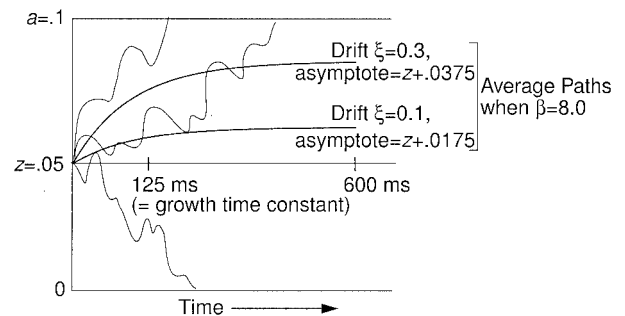
Figure 2. Illustration of the Wiener and Ornstein–Uhlenbeck (OU) diffusion models with a list of parameters. RT = response time; distrib. = distribution; S.D. = standard deviation.

process. If drift rate decreases, RTs increase with a relatively small change in the leading edge of the RT distribution and a larger change in the tail of the distribution.

Not shown in Figure 2 is the drift criterion, which serves the same function as the criterion in signal detection theory: It separates stimuli into those with positive drift rates and those with negative drift rates, just as the signal detection criterion separates stimuli into signal and noise. Like the criterion in signal detection theory, the value of the drift criterion may vary with experimental manipulations such as payoffs or the proportions of the two stimuli (Ashby, 1983; Link, 1975; Link & Heath, 1975; Ratcliff, 1978, 1985, 2002; Ratcliff et al., 1999). Changing the drift criterion from one block of trials to another is equivalent to adding or subtracting a constant to the drift rates for all stimuli in one block relative to another (Ratcliff, 2002).

Variability in processing across trials is implemented in several components of the diffusion model. First, the starting point z varies across trials with a rectangular distribution with mean z and range s_z. This allows the model to predict errors faster than correct responses because when the process starts near the incorrect boundary, correct responses will be slower and occur with lower probability, and error responses will be faster and occur with higher probability than if the process started near the correct boundary. The mixture of a larger proportion of fast errors when

the process starts near the incorrect boundary and a smaller proportion of slow errors when the process starts near the correct boundary averages overall to a pattern of error responses faster than correct responses (see Ratcliff & Rouder, 1998).

Second, the drift rate for nominally equivalent stimuli, that is, stimuli that are all in the same experimental condition, is not fixed across trials but instead varies with a normal distribution with mean v and standard deviation η. Because of this across-trial variability, the actual drift rates for some stimuli in a given experimental condition can be quite different from the mean drift rate (v) for the stimuli in that condition and can even have the opposite sign. With the opposite sign, the decision process will terminate with probability greater than .5 at the incorrect response boundary, regardless of how far the boundaries are placed from the starting point; this ensures that accuracy asymptotes as a function of time. The proportion of processes with drift of the opposite sign is usually relatively small, and their mean drift is closer to zero than the majority of the processes; so, their RTs will tend to be longer than those for correct responses.

With both variability in drift rate and starting point across trials, Ratcliff and Rouder (1998) and Ratcliff et al. (1999; see also Ratcliff, 1981; Smith, 1994; Van Zandt & Ratcliff, 1995) showed that the Wiener diffusion model can predict all of the patterns of relative speeds of correct and error responses that have been observed empirically. Error responses are sometimes faster than correct responses (this usually occurs when accuracy is high), sometimes they are slower (usually when accuracy is low), and sometimes there is a crossover pattern, within an experiment, such that errors are faster than correct responses in high-accuracy conditions and slower in low-accuracy conditions (Ratcliff & Rouder, 2000; Ratcliff et al., 1999; Smith & Vickers, 1988). Which pattern is observed in an experiment depends on the magnitudes of the two sources of variability. The model cannot predict a crossover pattern such that errors are slower than correct responses in high-accuracy conditions and faster in low-accuracy conditions with only drift rate varying, a pattern which has not been obtained experimentally to our knowledge.

Besides the decision process, which is shown in Figure 2, there are other nondecisional components of processing, such as stimulus encoding and response execution, which are represented in the model by a single random variable. The nondecision time varies across trials, with values coming from a rectangular distribution with mean T_{er} and range s_r. The predicted mean RT is therefore the mean time for the decision process to terminate plus T_{er}. In practice, the standard deviation of the distribution of decision times is much larger than that of the distribution of nondecision times; so, the shape of the RT distribution is determined almost completely by the shape of the distribution of decision times (Ratcliff & Tuerlinckx, 2002). Such variability is included in all of the models considered here. However, variability in the nondecision components does have two effects on model predictions: The leading edge of the RT distribution has greater variability across conditions than would otherwise be the case, and the rise in the leading edge of the RT distribution is more gradual than it would otherwise be. This latter effect was crucial to a Wiener diffusion model account of lexical decision data in Ratcliff, Gómez, and McKoon (2004).

OU Model

The OU model is identical to the Wiener diffusion model except that the drift of the process depends on two opposing quantities: ξ, the rate of accumulation of evidence from the stimulus, and β, a decay force that moves the process back toward its starting point (see the bottom panel of Figure 2). The value of ξ is constant over the course of evidence accumulation, whereas decay increases as a function of the amount of evidence already accumulated. Drift rate is equal to $\xi - \beta x$, where x is the distance from the starting point.[2] At some distance from the starting point, ξ and βx become equal, and so the mean drift rate becomes zero, although individual paths still vary randomly around the mean with drift rate $\xi - \beta x$. For the two average paths in Figure 2, β is 8 and ξ is either 0.3 or 0.1. The average paths of the process asymptote at the value of x at which $\xi - \beta x = 0$. Both of these paths have a time constant of 125 ms; that is, after 125 ms they are about two thirds of the way to asymptote, which is well within the usual range of RTs.

The OU model accounts for right-skewed RT distributions and errors in the same ways as does the Wiener diffusion model. Also, in our applications, variability in processing across trials is implemented in drift rate, starting point, and the nondecision components of RT just as in the Wiener model, and the implications for the relative speeds of correct and error responses are the same as for the Wiener model.

The decay mechanism in the OU model has been promoted as an alternative to variability in drift across trials as a way of limiting asymptotic accuracy in diffusion models for data from response signal procedures (e.g., Usher & McClelland, 2001). In response signal procedures, in which subjects are asked to respond at experimenter-determined times (e.g., Dosher, 1976, 1984; Ratcliff, 1978, 1980; Ratcliff & McKoon, 1982; Reed, 1973; Wickelgren, 1977), the growth of accuracy as a function of time reaches an asymptote. The Wiener diffusion model's account of this depends on variability in drift across trials (as mentioned above), whereas the OU model produces this behavior even without variability in drift across trials (Smith, 2000; Usher & McClelland, 2001). However, later we show that if boundaries are increased without limit, the OU model without across-trial variability in drift predicts no asymptote on accuracy.

As we show later in the results, when decay is large in the OU model, it does not fit the experimental data well. With moderate decay it does, but not as well as the Wiener diffusion model. When the decay parameter is free to vary, the best fits are obtained when the parameter approaches zero, making the model identical to the Wiener diffusion model.

---

[2] In our version of the OU model, we assumed that the decay parameter represents a restoring force attracting the process back to a starting point that may vary randomly between trials. An alternative interpretation of decay is that it attracts the process back toward zero, irrespective of starting point. We preferred the former interpretation because when the process decays toward zero, the effects of starting point decay exponentially and have only a transient effect on the subsequent dynamics. This property conflicts with the evidence that starting point variability is needed in diffusion process models to predict fast errors. This allows the model to parallel the Wiener diffusion model, which allowed us to evaluate the effects of decay without variation in any of the model's other assumptions.

The Accumulator and Poisson Counter Models

The accumulator model (Smith & Vickers, 1988, 1989; Vickers, 1970, 1978, 1979; Vickers et al., 1971) and the Poisson counter model (LaBerge, 1994; Pike, 1966, 1973; Smith & Van Zandt, 2000; Townsend & Ashby, 1983) have absolute stopping rules. Just as for their predecessor, LaBerge's (1962) recruitment model, evidence in favor of one response is accumulated in one counter, evidence for the other response is accumulated in a second counter, and the decision is determined by the first counter to reach its criterion. However, in the recruitment model, evidence is accumulated at discrete time steps in unit increments, resulting in incorrect RT distribution predictions. In the accumulator model, the assumption of sampling at discrete time steps is retained, but the amount of evidence accumulated on each step is drawn from a continuous distribution. In the Poisson counter model, the evidence is accumulated in unit increments but is sampled at random, continuously distributed times. These assumptions allow the models to avoid problems with distribution shape that occurred in the recruitment model.

The Accumulator Model

In the accumulator model (see the top panel of Figure 3), a value is sampled from a normal distribution of amounts of evidence at each of a sequence of equally spaced time steps. As in signal detection theory, the distribution has a standard deviation of 1.0 and a mean, μ, that depends on the quality of the information from the stimulus. The mean would be larger for easy stimulus conditions and smaller for difficult conditions. A criterion, termed the *sensory referent,* is set on the underlying evidence dimension. Like the drift criterion in the diffusion models, this criterion represents a point of zero stimulus information. If the amount of evidence sampled falls above the criterion, an amount equal to the difference between that amount and the criterion is added to one counter. If the amount falls below the criterion, the difference is added to the other counter. Because evidence is accumulated at discrete time steps, a parameter, λ, is required to convert time steps to continuous time.

In this article, we assume variability across trials in three components of the model (cf. Smith & Vickers, 1988). First, equivalent to variability in mean drift rates across trials in the diffusion models, the means of the evidence distributions are normally distributed with standard deviation σ_μ. Second, the nondecision component of RT varies across trials with a rectangular distribution with mean T_{er} and range s_t, exactly as in the diffusion models.

Third, the values of the response criteria vary across trials. Without this source of variability, the accumulator model, as described so far, has the problem that as the response criteria (K_A and K_B) increase, that is, as RTs become longer and accuracy increases, the RT distribution becomes more symmetric. For fitting the model to the data described later, we tried several possible distributions for variability in the criteria, namely, rectangular, normal, and Weibull distributions—the last to allow a range of criterion distribution shapes (see also Smith, 1989, and Smith & Vickers, 1989, for other proposals). The best fitting was the Weibull with an exponential form (i.e., Weibull shape parameter equal to 1). The values of the criteria on each trial (see Figure 3) are calculated by adding a value obtained from an exponential with mean κ to two base values, k_A and k_B (the same value added to each), to obtain the values of K_A and K_B (i.e., the mean values of the criteria are $k_A + \kappa$ and $k_B + \kappa$). Using two independent values from the exponential for the two criteria did not alter the qualitative fits of the model to data. We also found it necessary for the mean of the exponential to become larger in experimental conditions with accuracy instructions compared with conditions with speed instructions to produce RT distributions as skewed as data. This adds an additional parameter to the model for each additional speed–accuracy condition tested. The reason why the exponential works well is that empirical RT distributions are approximately exponential in the extreme tail (Burbeck & Luce, 1982; Luce, 1986; Ratcliff et al., 1999; Van Zandt & Ratcliff, 1995).

The Poisson Counter Model

The top panel of Figure 4 shows the arrival times of counts at the two counters in the Poisson counter model, and the bottom panel shows how they are accumulated. The times between counts are exponentially distributed with rate α for Counter A and rate β for Counter B. With exponentially distributed times between counts, the evidence streams are Poisson processes with means $1/\alpha$ and $1/\beta$ for Counters A and B, respectively. The quality of the information in the stimulus is represented by the relative accumulation rates for the two counters. Increasing the quality of the stimulus information causes an increase in the accumulation rate for one counter and a decrease for the other, such that the sum of the two rates, $\alpha + \beta$, is constant. This constraint means that the
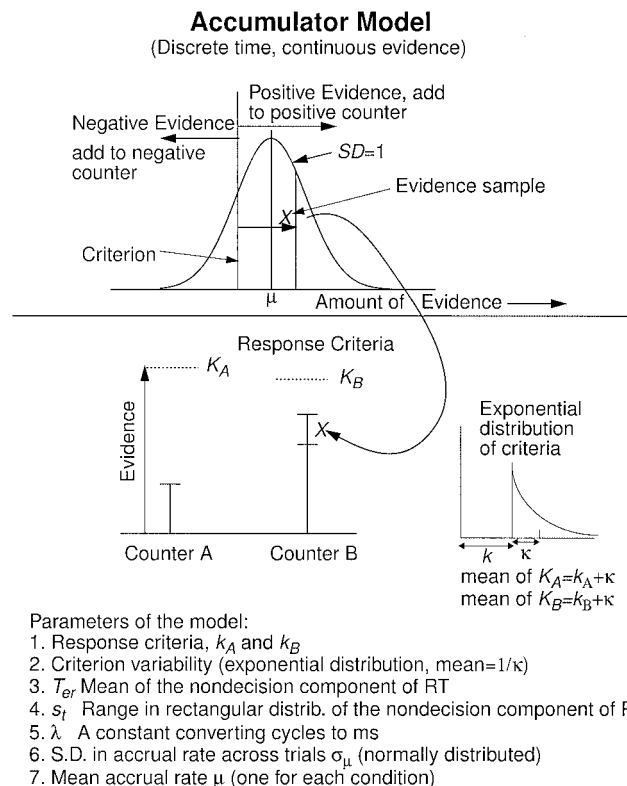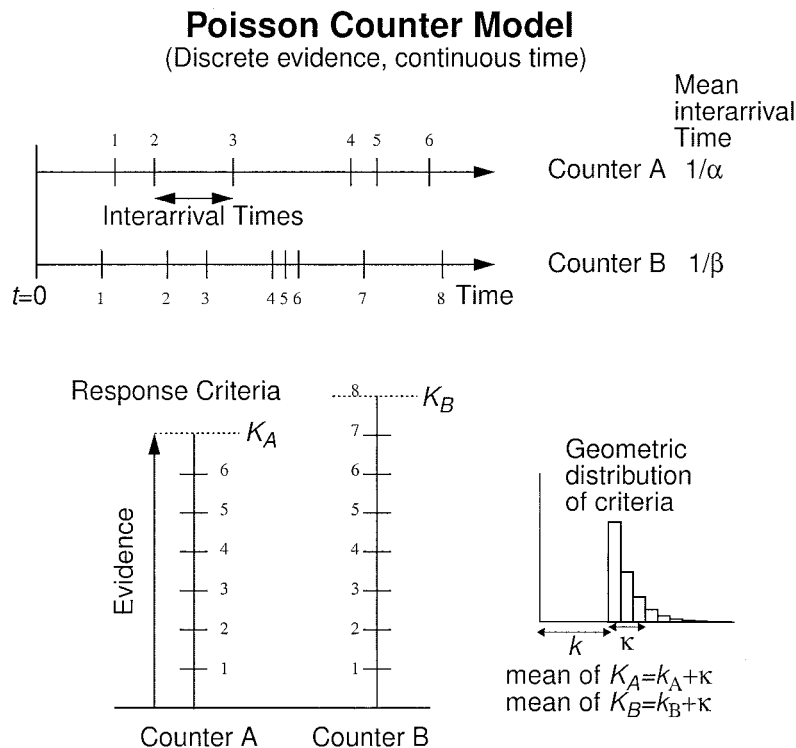

Accumulator Model
(Discrete time, continuous evidence)

Figure 3. Illustration of the accumulator model with a list of parameters. RT = response time; distrib. = distribution; S.D. = standard deviation.

Parameters of the model:
1. Response criteria, k_A and k_B
2. Criterion variability (exponential distribution, mean=$1/\kappa$)
3. T_{er} Mean of the nondecision component of RT
4. s_t Range in rectangular distrib. of the nondecision component of RT
5. λ A constant converting cycles to ms
6. S.D. in accrual rate across trials σ_μ (normally distributed)
7. Mean accrual rate μ (one for each condition)

Poisson Counter Model
(Discrete evidence, continuous time)

Figure 4. Illustration of the Poisson counter model with a list of parameters. RT = response time.

overall rate of evidence accumulation is constant, paralleling the assumption in the accumulator model that evidence is accumulated at a constant rate (λ ms per count), regardless of its magnitude, and also paralleling the assumption in the diffusion models that s^2, which determines the rate at which a process moves toward its boundaries, is constant while drift rate varies with stimulus difficulty. As each count arrives, it is accumulated in the appropriate counter. A response is initiated when one or the other counter reaches its criterion value, K_A or K_B.

As with the diffusion models and the accumulator model, there are three sources of across-trial variability. First, the nondecision component of RT has a rectangular distribution with mean T_{er} and range s_t. Second, we introduce variation in the accumulation rates across trials by allowing the relative probability of increments to the two counters to vary. The probability that a given count is added to Counter A, for example, is the ratio of its rate parameter to the sum of the rate parameters, $\pi = \alpha/(\alpha + \beta)$, and π varies from trial to trial. Because π is constrained to lie in the range zero

to one, a normal distribution would not be appropriate because its values are unbounded. Instead, the value of π is drawn from a beta distribution (see the Appendix). The beta distribution is a distribution with zero–one bounds that includes symmetric, positively and negatively skewed, uniform, and U-shaped forms as special cases (Johnson & Kotz, 1970, pp. 42–43). It provides the most general possible model for accrual rates. Third, the criterion values are geometrically distributed with a different value of the mean for each speed–accuracy condition tested. The geometric is the discrete analog of the exponential and so gives the Poisson counter model the same properties as the accumulator model, allowing it to predict more positively skewed RT distributions than it otherwise would produce.

Neither the accumulator model nor the Poisson counter model has previously been examined with all the sources of across-trial variability that we have added to them. These sources of variability give the models the same potential flexibility as the diffusion models. However, as we show later, although this flexibility helps

each model, it still does not allow the Poisson counter model to fit the patterns of empirical data that we present in the experiments.

Mimicking Between Models

One problem that occurs as models become more complex, as they have in the sequential sampling domain, is the possibility of model mimicry. Although none of the models we evaluate mimics another exactly, we show that some can mimic each other sufficiently to render them, for all practical purposes, empirically indistinguishable. In this article, we evaluate mimicry in two ways.

First, the models are fit to three comprehensive sets of data chosen to be representative of data for two-choice decision tasks in which detailed RT data have been collected and also to exhibit all of the main qualitative features that have been considered important theoretically in the RT modeling literature (cf. Luce, 1986). For each of the data sets, there was a within-block manipulation of stimulus difficulty, such that accuracy varied from near floor to near ceiling, and a between-blocks manipulation intended to affect the values of response criteria. If competing models fit the data equally well under these circumstances, then from an empirical standpoint, they can be said to mimic each other.

The second way we examine mimicking between pairs of models is to use the parameters obtained from fitting one model to a data set to generate exact predictions from that model and then attempt to fit those predictions with the other member of the pair. The second model of the pair should fit well if the two models truly mimic each other, but in cases in which one of the models does not fit the data, the other model may or may not fit the predictions.

Before applying the models to the data from the three experiments, we explain the methods we use to display the data and the methods by which the models are fit to the data. Then, details of the procedures for the experiments and their data are presented.

Quantile Probability Functions

In fits of any model to RT data, there are two dependent variables to consider, accuracy and RT. The proportion of correct and error responses and the relationship between their RTs, as well as the distributions of the RTs, must all be considered when assessing the fit of the model. Traditionally, accuracy, mean RTs, and RT distributions have all been plotted separately as a function of experimental condition. Here, instead, we display them all together in quantile probability functions. This method of displaying data has the advantage that the joint behaviors of the dependent variables can be more easily examined. The quantile probability function (QPF; Ratcliff, 2001) is a development of the latency probability function, which was used to display the joint behavior of mean RT and accuracy in early work on sequential sampling models by Audley and Mercer (1968), Audley and Pike (1965), LaBerge (1962), Pike (1973), Pike and Ryder (1973), and others.

A QPF is constructed by plotting the quantiles of the distribution of RTs for positive responses and the quantiles of the distribution of RTs for negative responses for each experimental condition on the y-axis and the probability of the response on the x-axis. For the data presented in this article, we use five quantiles, with the plotted quantile points representing the RTs below which fall .1, .3, .5, .7, and .9 of the total probability mass in the distribution.

In Figure 5, RT quantiles are plotted as they are in QPFs. The left panel shows three plots of the same distribution, one with 5 quantiles, one with 10, and one with 25. The right panel shows the associated RT density functions, along with pseudohistograms constructed from the quantiles. To obtain an approximation to the density function, we constructed equal-area rectangles corresponding to equal amounts of probability mass between the quantiles between each of the quantile RTs (as for the left most plot in the left panel) so that closely spaced quantiles are spanned by taller rectangles. For the 5-quantile plot, the distance between the .5 and .7 quantiles is identified as "Y" and the distance between the .7 and .9 quantiles is identified as "X," and these distances are shown in the top right panel as the distances between the quantiles plotted on the x-axis. As the plots suggest, the set of rectangles derived from the quantiles approaches the continuous density function as the number of quantiles increases.

For each condition in an experiment, information about the shape of the RT distribution is carried by the vertical separations of the quantile points for that condition. The fastest responses in the distribution map onto the lower quantile points, and the slow responses in the tail of the distribution map onto the higher quantile points. Because RT distributions are usually right skewed, the separation of the higher quantile points is greater than that of the lower points.

A full representation of the data from an experiment requires two QPFs, one for each response. However, if the data are symmetric for the two responses (i.e., RTs and accuracy values for the two choices are about the same), they can be averaged across responses to give a single QPF. Experiment 1 below yielded symmetric data of this kind.
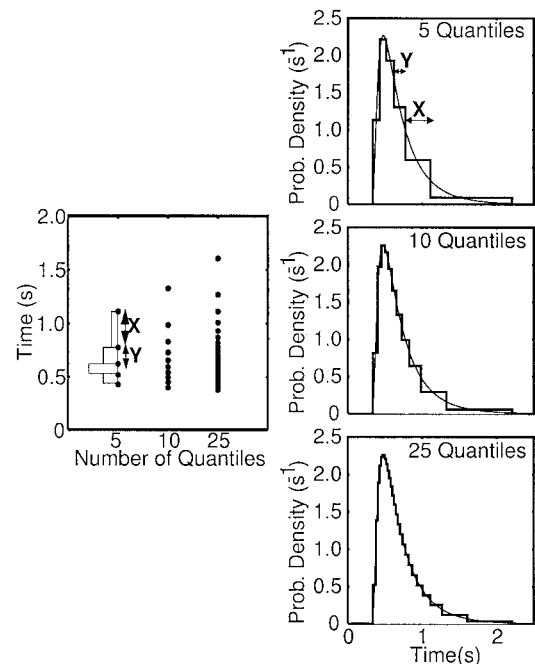

Figure 5. Mapping from quantiles to response time distributions. The distances between quantiles (e.g., X and Y in the left panel) map into width of the rectangles in the histograms on the right. Prob. = probability.

Within-Block Variables

When stimulus difficulty is varied in a within-block design, there are two important constraints on the models. First, the effects of difficulty on the QPF are determined by only one parameter, the rate of accumulation of evidence, namely, drift rate in the diffusion models and accrual rate in the accumulator and Poisson counter models. With only drift or accrual rate varying, accuracy rates plus mean RTs and RT distributions for both correct and error responses must be fit. Predictions from a model can be plotted with isoquantile lines connecting equivalent quantiles across the range of predicted probability values.

Shape of QPFs

The second constraint on the form of the QPFs is that their shape is determined by only a few parameters of the models. As an example, in Ratcliff's Wiener diffusion model with starting point equidistant from the two boundaries, the form of the QPF is determined by the three parameters a (boundary separation), η (across-trial variability in drift rate), and s_z (across-trial variability in starting point). Assuming symmetric RTs and accuracy values for the two responses, then when η and s_z are zero, RTs for correct and error responses are equal and the QPF is symmetric with an inverted U shape. When η is high and s_z is low, error responses are slower than correct responses, and the QPF has a peak to the left of the .5 probability point. When η is low and s_z is high, error responses are faster than correct responses, and the QPF has a peak to the right of the .5 probability point. Thus, the shape of the QPF allows the relative speeds of correct and error responses to be determined by visual inspection. The vertical location of the QPF is determined by the nondecision component of reaction time, T_{er}.

Fitting Methods

There are three ways predicted values of RT and accuracy can be generated from the models. First, if there are exact solutions for a model (i.e., formulas for the RT distribution and accuracy), then exact predictions can be produced. Second, numerical approximations can be used to produce predictions. Within the limits of generating predictions in some reasonable amount of time, the predictions can be as accurate as those from the first method. Third, the model can be simulated on a trial-by-trial basis. For simulations, within-trial and across-trial variability give rise to variable correct and error responses and RTs, and average accuracy and RT are determined from running many trials.

For the accumulator, Poisson counter, and two diffusion models, either exact solutions or numerical approximations, or both, are available and were used in fitting the models. Although the simulation method does not produce predicted values with the same degree of accuracy as the first two methods, it is easier to implement for the models considered here, and we used it to provide checks on the other methods. Also, simulation is the only method available for some of the neurally inspired models discussed later.

To fit each model to data, we used one of the methods just described to generate predicted data from the parameter values. Specifically, for each condition in an experiment, we generated predictions for accuracy and five RT quantiles for both correct and error responses. We also selected either a chi-square or sum-of-squares measure of how well the predictions match the data (see below). We then used the SIMPLEX minimization routine (Nelder & Mead, 1965) to adjust the parameter values until a minimum value of the chi-square or sum-of-squares measure was obtained.

We evaluated the models using data from three experimental tasks that are representative of the kinds of tasks for which sequential sampling models have previously been proposed. The first task was a perceptual judgment task, in which probabilistic feedback was used to vary difficulty. The second task was a lexical decision task that required judgments about word identity. The third task was a recognition memory task that required judgments about whether items had occurred on a previously studied list. For all three tasks, the data were typical of those reported for similar tasks in the literature.

We chose to evaluate the models against group data, obtained by averaging quantiles of the RT distributions and response probabilities across subjects (Ratcliff, 1979; Thomas & Ross, 1980). This had the advantage of reducing variability among subjects, thereby bringing out the qualitative effects of the experimental manipulations as clearly as possible. It also allowed us to keep the evaluation task to manageable proportions. It has been our experience that fits to individual subjects and fits to quantile-averaged group data exhibit very similar features and that parameter values for group fits are in good agreement with the average parameter values for fits to individual subjects (see Ratcliff, Thapar, & McKoon, 2003, and Thapar, Ratcliff, & McKoon, 2003, for concrete examples with four groups of about 40 subjects per group).

We used two different criteria to assess how well a model fit experimental data: a minimum chi-square statistic and a penalized maximum likelihood statistic called the Bayesian information criterion (BIC; Schwarz, 1978). We discuss how well the models fit data in terms of the chi-square statistic, and we additionally report the BIC statistic in tables. For all three data sets, the conclusions are the same whether the models are evaluated with the minimum chi-square or the BIC. We used a third statistic, a weighted least squares (WLS), to fit one model to the predictions of another model.

The minimum chi-square statistic we used is the Pearson statistic. For N observations grouped into bins, this statistic has the form

$$\chi^2 = \sum_i \frac{N(p_i - \pi_i)^2}{\pi_i},$$

where p_i is the proportion of the observations in the ith bin and π_i is the proportion in the bin predicted by the model. In our fits, the empirical quantiles were used to form the boundaries of the bins, giving 12 bins per pair of distributions (6 each for correct response and error distributions). The probability masses p_i and π_i in the formula are joint probabilities that sum to unity across each pair of correct and error distributions. For each of our data sets, there were 12 proportions in each experimental condition, and the total probability mass in each condition summed to 1.0, reducing the number of degrees of freedom to 11. For a total of k experimental conditions and a model with M parameters, the number of degrees of freedom in the fit was therefore $df = k(12 - 1) - M$.

The BIC statistic, for binned data, is

$$\text{BIC} = -2\left(\sum_i Np_i \ln(\pi_i) \right) + M\ln(N),$$

where p_i and π_i are the same as in the previous equation and m is the number of free parameters in the model. The term $M\ln(N)$ on the right of the equation is a penalty term that penalizes models in proportion to their number of free parameters and the logarithm of the size of the sample.

Besides providing a penalty for the number of parameters in a model, the BIC also penalizes models for the complexity of their functional form. This occurs because when the sample size is large, the BIC is asymptotically equivalent to a Bayesian model selection (BMS) method that weights the assessment of model fit according to the prior probabilities of the parameters (Pitt, Myung, & Zhang, 2002; Wasserman, 2000). The BMS method is in turn asymptotically equivalent to the *minimum description length* method recently advocated by Pitt et al. (2002). This latter method penalizes models both for their number of free parameters and the complexity of their functional form.

The BIC is also closely related to the likelihood ratio chi-square statistic, G^2. For any given set of data, G^2 and the BIC differ by a constant, so the parameters that minimize one also minimize the other. The G^2 statistic can be written

$$G^2 = 2 \sum_i Np_i\ln\left(\frac{p_i}{\pi_i}\right).$$

This statistic is equal to twice the difference between the maximum possible log likelihood and the log likelihood predicted by the model. The χ^2 and G^2 statistics approach one another as sample sizes become large (Jeffreys, 1961, p. 197); both are distributed as a chi-square random variable with degrees of freedom presented above.

Because our fits were carried out on group data, obtained by averaging quantiles across subjects, it was not appropriate to weight the observed and predicted proportions in the χ^2 statistic by the total sample size N as is done in the usual Pearson chi-square test. Instead, we calculated the statistic from the observed and predicted proportions instead of frequencies and multiplied the values by 100 for readability. We have used this statistic as a relative rather than absolute measure of fit and denote it by the symbol X^2 to emphasize that it is not a proper chi-square because it has been calculated from quantile-averaged data. Here, we use it to provide a numerical measure of fit that serves as an adjunct to the qualitative comparisons that are the main focus of this article. We also present an example of the sampling distribution of this X^2 statistic for Experiment 1. Because the penalty term in the BIC depends on sample size, the BIC and G^2 statistics were calculated with N set equal to the average number of observations per subject in each condition in the data set.

For fitting models to exact predictions from other models, we used the WLS statistic because this does not depend on the number of observations as do the other statistics and because it is robust to systematic deviations, for example, in the lower quantile RTs (Ratcliff & Tuerlinckx, 2002). The WLS statistic minimizes the sum of squared differences between the observed and predicted accuracy values plus the sum of the squared differences between the observed and predicted quantile RTs for correct responses and errors. For one experimental condition, the fit statistic, sum of squared errors (SSE), is given by the expression

$$SSE = 4(P_{th} - P_{ex})^2 + \Sigma_i w_i[Q_{th}(i) - Q_{ex}(i)]^2,$$

where P is probability, $Q(i)$ is the quantile RT in seconds, "th" stands for predicted, "ex" stands for experimental, and the w_i are quantile weights. The value of SSE is summed over experimental conditions. The quantile weights, w_i, were set equal to 2 for the .1 and .3 quantiles, 1 for the .5 and .7 quantiles, and 0.5 for the .9 quantile. This weighting scheme reflects, approximately, the relative variability of the accuracy measures and the RT quantiles.

In the computation of SSE, accuracy is explicitly represented, but in the chi-square and BIC statistics, accuracy is not explicitly represented. However, because the proportions of the probability mass in the correct and error distributions may differ for the predicted and observed distributions, the fitting method will attempt to make the probability mass for predicted and observed values as similar as possible so that these statistics take into account discrepancies in fit for both accuracy and RT.

Variability in the Nondecision Component of Processing

A critical difference between the WLS and minimum chi-square statistics in application to our RT and accuracy data is the following: The minimum chi-square method attempts to minimize the discrepancies between observed and predicted proportions between adjacent pairs of quantiles and outside the two extreme quantiles. In the X^2 statistic, the square of each discrepancy is weighted by the reciprocal of the predicted proportion in the bin. For this statistic to be well-defined, the predicted proportion below the .1 quantile must be nonzero for all distributions that contribute to the fit. This means that if T_{er} has no variability, then it is required to be smaller than the smallest .1 quantile. If the true variability in the .1 quantile across conditions is greater than that predicted by the model, this requirement can produce severe distortions in fit because the requirement results in systematic underestimates of the larger of the .1 quantiles, which in turn produces large distortions in the tails of the fitted distributions. In some of the fits we carried out, this resulted in discrepancies of several hundred milliseconds between the predicted and observed .9 quantile RTs.

If there is a rectangular distribution of values of the nondecision component of processing with mean T_{er} and range s_t, as we assume above, then $T_{er} - s_t/2$ has to be smaller than the smallest .1 quantile, and the distortions are reduced or eliminated (see Ratcliff & Tuerlinckx, 2002). The improvement in the fit occurred because variability in the nondecision component of processing stretches out the leading edge of the distribution, allowing for greater variability in the location of the .1 quantile. For the WLS statistic, the estimate of T_{er} does not have to be less than the smallest .1 quantile, and it is robust to variability in the .1 quantile, although the recovered parameter values will often be biased (Ratcliff & Tuerlinckx, 2002).

The assumption of variability in the nondecisional components of RT is not new to our application. In domains like simple RT, such variability often forms an integral part of quantitative models (see Luce, 1986; Smith, 1990b) and is justified on physiological grounds. It has rarely been included in modeling two-choice RTs, although Smith (1989) used distributions of simple RTs to estimate the nondecisional components of RT in a two-choice task in which subjects made judgments about the orientation of an array of randomly oriented line segments.

All of the fits to data in this article were carried out using both WLS without variability in T_{er} and minimum chi-square with variability in T_{er} (note that we use "variability in T_{er}" as shorthand for "variability in the nondecision component of processing"). Although we report only the chi-square fits, the conclusions drawn from both fits were the same. In general, the Wiener and OU diffusion models benefited more from the introduction of variability in T_{er} than did the accumulator and Poisson counter models because the diffusion models without variability in T_{er} predict RT distributions that have sharper leading edges than are usually seen in data.

Method for Experiments 1–3

As discussed earlier, three experiments were chosen to represent reasonably common two-choice tasks with a long history in the RT domain. In the first, a signal detection-like experiment (Ratcliff et al., 2001), two vertically aligned dots were displayed on each trial, and subjects were asked to decide whether the separation between them was "large" or "small." Stimulus difficulty was varied within block via dot separation: There were 32 possible separations, labeled 1 through 32 with 1 being the smallest separation, ranging from 1.75 cm to 3.33 cm in equal intervals. After each trial, subjects were given feedback such that the response was designated as "correct" or "error." The feedback was probabilistic, and was chosen from a probability associated with each stimulus: For Stimuli 1 through 6, the "small" response was designated correct with probability .999. For Stimuli 7, 8, 9, 10, 11, 12, 13, 14, and 15, "small" was designated correct with probabilities .913, .888, .856, .819, .774, .722, .664, .601, and .534, respectively. For large separations, for Stimuli 25 through 32, the "large" response was designated correct with probability .999, and for Stimuli 24 through 16, "large" was designated correct with the same probabilities as for "small" for Stimuli 7 through 15. Subjects understood that they could not be completely accurate, that for separations in the middle of the range either response might be designated as correct, and that their task was to give their best judgment. There were 12 lists in each session, with three presentations of each of the 32 stimuli in each list. In 6 of the lists, subjects were given accuracy instructions, and in the other 6, they were given speed instructions. Speed versus accuracy instructions alternated between blocks: Subjects were asked either to respond as quickly as possible or to make as few errors as possible. In the speed blocks, responses longer than 700 ms were followed by a "TOO SLOW" message. In the accuracy blocks, "large" responses to Stimuli 1 through 6 or "small" responses to Stimuli 26 through 32 were followed by a "BAD ERROR" message. The subjects were 17 Northwestern undergraduates who each participated in two 45-min sessions.

In the second experiment, a lexical decision experiment (Wagenmakers, Ratcliff, Gómez, & McKoon, 2004), a letter string was presented on each trial, and subjects were asked to judge whether it was a word or a nonword. Trials were blocked into sets of 96, with each of 15 subjects (Northwestern undergraduates) completing 20 blocks in one session. Word frequency was varied within blocks, with equal proportions in each block of high-, low-, and very low-frequency words (mean frequency values of 325.0, 4.4, and 0.37; Kučera & Francis, 1967). Speed versus accuracy instructions alternated between blocks. On speed trials, a "TOO SLOW" message was presented if RT was greater than 750 ms. On accuracy trials, an "ERROR" message was given for error responses. There were equal proportions of words and nonwords in each block, with nonwords constructed from words in which all the vowels were randomly replaced with other vowels (see Ratcliff et al., 2004, for other details).

In the third experiment, a study–test recognition memory experiment, subjects were presented with lists of pairs of words to study (1,500 ms per pair) and then, for each list, were asked to judge for each of a series of single test words whether it had been in the list (an "old" test item) or not (a "new" test item). There were 28 pairs per study list and 56 test items per test list, and each subject was tested with a total of 30 study–test lists per session. There were 3 subjects (Northwestern undergraduates), each tested for nine sessions. Within a list, difficulty of the decision was manipulated by varying the number of times a pair was presented in the study list (one or four) and by using high-, low-, or very low-frequency words (the same pools of words as in the lexical decision experiment). Incorrect responses were followed by an "ERROR" message. The proportion of old versus new test items was varied between lists: The proportion of old to new test items was 3.5:1, 2:1, 1:1, 1:2, or 1:3.5.

For all three experiments, the constraints the data impose on the models are as follows: First, in all cases the RT distributions are positively skewed. Second, stimulus difficulty (a within-block variable in all three experiments) affects RT (mean and standard deviation) mostly by increasing the skew of the RT distributions. The between-blocks variables (speed–accuracy instructions and proportion of old–new test items) affect RT with changes in both the leading edge and the skew of the RT distributions. Third, errors are slower than correct responses in all conditions of the first and third experiments (signal detection and recognition memory), whereas in the second experiment (lexical decision), errors are slower than correct responses with accuracy instructions and faster than correct responses with speed instructions. This latter pattern of data is an especially difficult one for models to accommodate. Fourth, in moving from speed to accuracy instructions, there are large changes in RT (several hundred milliseconds), accompanied by modest changes in accuracy (about .05). This contrasts with the effect of proportion of old–new test items: large changes in RT (up to about 200 ms), accompanied by large changes in accuracy (up to about .40).

In fitting the models to the data from these experiments, we simultaneously fit accuracy rates, RTs for correct responses and errors, and their associated distributions—that is, all the data shown by a QPF. In fitting the models, we focus on which of them can capture the qualitative trends in the data. As we show later, the Wiener model and the OU model with relatively small decay mimic each other to a large degree, and both fit the experimental data reasonably well. The accumulator model fits the data from the experiments as well as the Wiener and OU models except for cases in which error responses are faster than correct responses; the accumulator model can predict only errors slower than correct responses. The Poisson counter model fits worse than the other models overall, and it too can predict only errors slower than correct responses.

Experiment 1: Signal Detection

Figure 6 displays the fits for all except the worst-fitting model, with the triangles representing the data and the lines representing the best fits of the best versions of the models. We collapsed the 32 conditions into 4 by grouping conditions with similar values of accuracy and similar values of RT. Also, because the data were symmetric for the two responses ("large" and "small"), they were collapsed both for display in the figures and for fitting the models.

The manipulation of difficulty (dot separation) produced a change in mean RT of 80 ms over the four conditions with speed instructions and 250 ms with accuracy instructions, with these changes appearing mainly in the skew of the RT distributions and only a minimal change in the .1 quantiles (leading edges) of the distributions (about 15 ms with speed instructions and 30 ms with accuracy instructions). Difficulty produced a change in accuracy from .95 to .55, about the same size for both speed and accuracy instructions, with correct RTs decreasing as accuracy increased. Error responses were always slower than correct responses, and error RTs first increased then decreased as accuracy increased, showing a nonmonotonic pattern. Over all the difficulty conditions, the speed–accuracy manipulation produced differences in
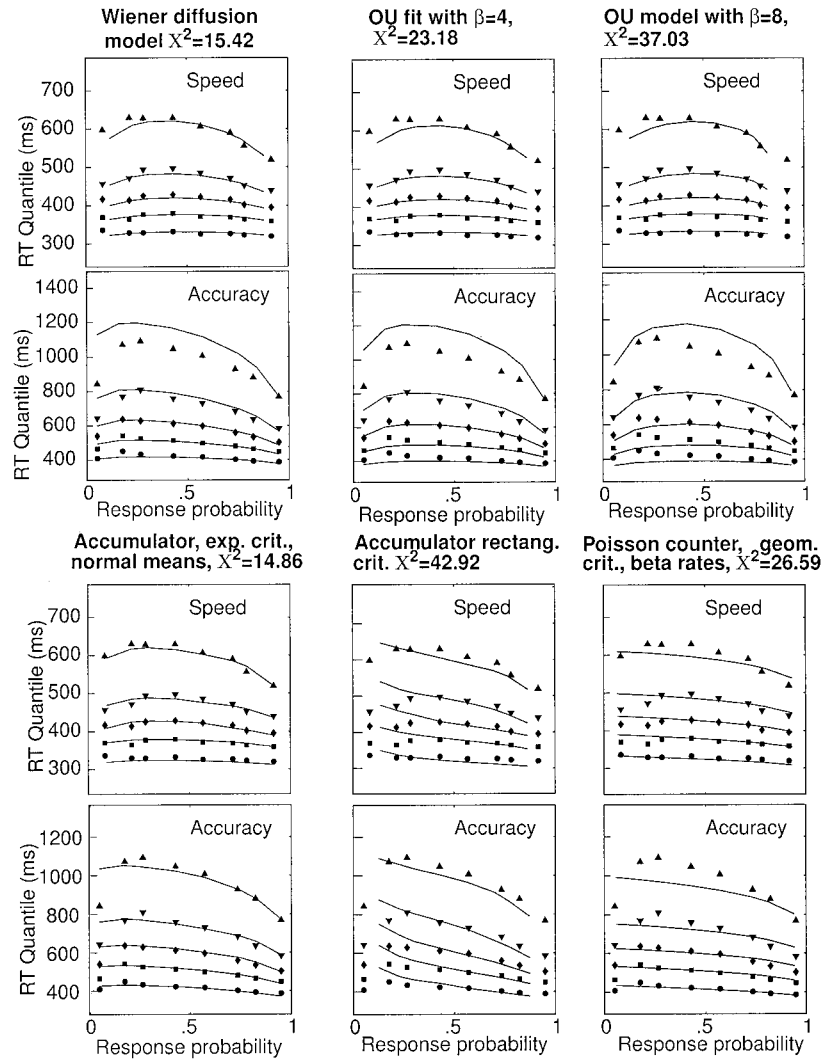
Figure 6. Fits of the Wiener and Ornstein–Uhlenbeck (OU) diffusion models and the accumulator and Poisson counter model for the data from Experiment 1. The decay parameter (β) was fixed for the two OU model fits. RT = response time; exp. crit. = exponential criteria; rectang. crit. = rectangular criteria; geom. crit. = geometric criteria; ● = .1 quantile RT; ■ = .3 quantile RT; ◆ = .5 quantile RT; ▼ = .7 quantile RT; ▲ = .9 quantile RT.

accuracy of about 4%–8% and differences in mean RTs of about 100–250 ms. The RT distributions in the accuracy conditions were more skewed, and the .1 quantiles were 50–70 ms longer than in the speed conditions, but the RT distribution shape was quite similar across conditions.

Because the data were largely symmetric for the two responses, the response criteria were equated for the Poisson counter and for the accumulator models, and the starting points for the diffusion models were set halfway between their boundaries. The values of the response criteria and boundaries were free to vary between the speed and accuracy conditions.

Wiener Diffusion Model

Qualitatively, the fit is reasonable (cf. Ratcliff et al., 2001; see Figure 6 and Tables 1 and 2). With only the drift rate varying with

difficulty and only the boundary positions varying with instructions, the model captures the shapes of the RT distributions and the changes in them as a function of difficulty and instructions. Errors are slower than correct responses because variability in starting point across trials has a smaller effect than variability in drift across trials. The main systematic discrepancy in the fit occurs with the .9 quantiles in the accuracy conditions; the model systematically overestimates the location of this quantile. For this model, $X^2 = 15.42$.

OU Model

The OU model was fit to the data with $\beta = 4$ and $\beta = 8$; the fit for $\beta = 8$ is shown in Figure 6 (see also Tables 1 and 2). As an indication of parameter size, the average sample path of an OU process with $\beta = 8$ is shown in Figure 2. We chose $\beta = 8$ as the

Table 1
Drift Rates, Accrual Rates, and Chi-Square Values in the Fits of the Models to Experiment 1

Model	v_1	v_2	v_3	v_4	X^2	df	BIC
Wiener	0.0391	0.1320	0.1944	0.3208	15.42	78	8,162.1
OU ($\beta = 4$)	0.0399	0.1349	0.1983	0.3312	23.18	77	8,176.6
OU ($\beta = 8$)	0.0342	0.1150	0.1702	0.2772	37.03	77	8,210.7
Rectangular accumulator	0.2150	0.7519	1.1029	1.6386	42.92	76	8,260.6
Exponential accumulator	0.1220	0.4140	0.6088	1.0252	14.86	76	8,172.0
Rectangular Poisson counter	0.5536	0.6793	0.7572	0.9002	65.62	76	8,285.8
Geometric Poisson counter	0.5460	0.6545	0.7230	0.8646	26.58	76	8,194.6

Note. v_1 to v_4 are generic parameters: They represent drift rates in the Wiener and Ornstein–Uhlenbeck (OU) models (also denoted v in the text), step increment distribution means in the accumulator model (denoted μ in the text), and the relative accumulation rates in the Poisson counter model (denoted ρ in the text). BIC = Bayesian information criterion.

extreme value of β to examine because with this value much of the effect of decay would occur within typical decision times and between the boundaries. We fit the model with only the drift rates varying with difficulty and only the boundary positions varying with instructions. Errors are slower than correct responses for the same reason as in the Wiener diffusion model. With $\beta = 8$, the goodness of fit is relatively poor ($X^2 = 37.03$), and the qualitative discrepancies between the model and data are easily observable (see Figure 6): The model is unable to capture the differences between the speed and accuracy conditions, and the drift values that best predict the range of response probabilities in the accuracy conditions produce systematically depressed response probabilities in the speed conditions; this appears as a restricted range of x values for the fitted QPF compared with the data. Also, the shapes of the predicted RT distributions are systematically distorted because the increased boundary position values needed to produce higher correct response probabilities and longer mean RTs in the accuracy conditions are accompanied by a sharp increase in the skewness of the predicted RT distributions.

With $\beta = 4$, the fit is better; it is qualitatively similar to that of the Wiener model, although numerically somewhat poorer ($X^2 = 23.18$). Allowing β to vary as a free parameter in fitting produces the best fits when $\beta = 0$, that is, when the OU model becomes identical to the Wiener model. The qualitative similarity between

the fits for the OU model with $\beta = 4$ and the Wiener diffusion model came as a surprise to us; we had expected the two models to produce different predictions on the basis of our notions of the effects of decay.

Accumulator Model

Initially, we examined fits of the accumulator with normal and rectangular distributions of variability for the response criteria. The results are similar; so, we report only the rectangular case here. The model was fit with only accrual rate varying with difficulty and only the response criteria varying with instructions. Because the "large" and "small" data are symmetric, k_A was set equal to k_B. The range of the rectangular distributions of criterion values were different for the speed and accuracy conditions. Quantitatively, the fit is poor ($X^2 = 42.92$), and the qualitative pattern of predictions is incorrect. One problem is that the RT distributions are too symmetric. Also, unlike the data, which exhibit bowed QPFs, the predicted QPFs decrease monotonically with accuracy, except for a sharp increase for errors in the highest accuracy condition (most apparent in the tail quantiles). Furthermore, the range of accuracy values is substantially underpredicted.

It was the finding that fits of the accumulator model with normal and rectangular distributions of criteria variability produce RT

Table 2
Parameters of the Wiener and OU Diffusion Models for Experiments 1–3

Model	Experiment	a_s	a_a	T_{er}	η	s_z	s_t	z_1	z_2	z_3	z_4	z_5
Wiener	1	0.0821	0.1440	0.3109	0.1475	0.0324	0.1000					
OU ($\beta = 4$)	1	0.0700	0.1128	0.3302	0.1417	0.0111	0.1500					
OU ($\beta = 8$)	1	0.0687	0.0975	0.3185	0.0845	0.0100	0.1000					
Wiener	2	0.0855	0.1593	0.4082	0.1135	0.0531	0.1430	0.0410	0.0774			
OU ($\beta = 4$)	2	0.0748	0.1228	0.4150	0.0771	0.0320	0.1430	0.0357	0.0586			
Wiener	3a	0.1139		0.5222	0.1658	0.0393	0.2140	0.0310	0.0373	0.0550	0.0680	0.0768
OU ($\beta = 4$)	3a	0.0996		0.5309	0.1867	0.0294	0.2140	0.0293	0.0345	0.0474	0.0571	0.0639
Wiener	3b	0.1151		0.5227	0.1758	0.0425	0.2140	0.0316	0.0384	0.0562	0.0682	0.0768
OU ($\beta = 4$)	3b	0.0958		0.5290	0.1417	0.0105	0.2140	0.0239	0.0297	0.0467	0.0587	0.0661

Note. In Experiments 1 and 2, a_s represents boundary separation in the speed condition, and a_a represents the boundary separation in the accuracy condition. In Experiment 3, a_s represents boundary separation. In Experiment 2, z_1 represents the starting point for the speed condition, and z_2 represents the starting point for the accuracy condition. In Experiment 3, z_1 to z_5 represent the starting points in the five probability conditions (22%, 33%, 50%, 67%, and 78% old, respectively). Experiment 3a refers to fits to Experiment 3 with no drift criteria, and 3b refers to fits to Experiment 3 with drift criteria. T_{er} = mean of the nondecision component of response time; η = variability in drift rate across trials; s_z = variability in starting point across trials; s_t = variability in nondecision component of response time across trials.

distributions that are too symmetric that led us to explore modeling the criteria with a distribution that has a long tail, the Weibull distribution. The best fits were obtained with the shape parameter of the Weibull equal to 1.0, which represents an exponential distribution. The criterion parameters k_A and k_B were set equal to each other, with different values for the speed and accuracy conditions. There were also different values of the exponential mean κ for the speed and accuracy conditions. With these parameters, the model fit the data well ($X^2 = 14.86$; see Figure 6, bottom left panel, and Tables 1 and 3). In terms of the X^2 goodness-of-fit measure, this version of the accumulator model fit the data better both qualitatively and numerically than either of the diffusion models. In terms of BIC, which penalizes models for their number of free parameters, the Wiener diffusion model, with two fewer parameters, is the better model.

Poisson Counter Model

With rectangularly distributed response criterion values and beta distributed accrual rates, the fit of the model is poor ($X^2 = 65.62$). The model was fit with only accrual rate varying with difficulty, k_A equal to k_B, and different values of k_A, k_B, and the range of the distribution of criteria for the speed and accuracy conditions. The QPFs are almost monotonic (see Figure 6), the RT distributions are too symmetric, and the .9 quantiles for errors when accuracy is in the .8 to .9 range underestimate the empirical values by about 200 ms.

With geometric distributions for the response criteria (different values of κ for speed and accuracy conditions), the fit of the model is appreciably better (see Figure 6 and Tables 1 and 4; $X^2 = 26.58$) and numerically comparable to that of the diffusion models. However, the QPFs show monotonically increasing error RTs, and the RT distributions are still more symmetric than the data—some of the .9 quantiles for errors underestimate the data by around 100 ms.

Sampling Distribution for the Chi-Square Statistic for the Wiener Diffusion Model

We examined the sampling distribution of the chi-square statistic for the Wiener diffusion model using Monte Carlo simulations

using the method presented in Ratcliff and Tuerlinckx (2002). We generated simulated data from 22 experiments with the same number of subjects per experiment (17) and the same number of observations per condition as in Experiment 1. The parameter values were randomly selected from normal distributions with means and standard deviations obtained from fits to individual subjects from Experiment 1 here (presented in Tables 2 and 3 from Ratcliff et al., 2001). The data from the simulated individual subjects were averaged as for the real data, that is, each accuracy value and each quantile RT was averaged. We then fit the Wiener diffusion model to the 22 data sets and obtained chi-square values. The mean chi-square value was 12.8 and the standard deviation was 3.9, with the upper .05 confidence limit 17.5.

The best-fitting models have chi-square values near the .05 confidence limit, which suggests that the models fit the data quite well. However, the chi-square statistic has some limitations. It is well-known that small systematic deviations between the model and data can lead to highly significant values of chi-square as the number of observations increases. Along with differences among subjects, other factors can inflate chi-square values: For example, systematic changes in performance within sessions or across sessions, if multiple sessions are tested, such as practice effects, fatigue, or adoption of different criteria can all inflate chi-square. In fact, the latter may be responsible for the well-known long-range sequential effects in sequences of RTs (see Gilden, 2001; Wagenmakers, Farrell, & Ratcliff, in press).

Discussion of the Models' Fits for Experiment 1

Three of the models, the Wiener diffusion model, the OU diffusion model with moderate decay ($\beta = 4$), and the accumulator model, give good accounts of the data. Both the overall shapes of the predicted RT distributions and the ways the shapes are predicted to change with stimulus difficulty and speed versus accuracy instructions agree with the data, as do the associated values of accuracy. Although the chi-square for the Poisson counter model with geometric criteria is similar to that for the OU model with moderate decay, the Poisson counter model's fit is qualitatively inferior, especially in its inability to produce nonmonotonic QPFs, that is, QPFs in which RTs for errors increase and then decrease as accuracy increases.

Table 3
Parameters of the Accumulator Models

Model	Experiment	T_{er}	σ_μ	κ_s	κ_a	s_t	λ	$k_a + k_b$	k_1	k_2	k_3	k_4	k_5
Rectang accum	1	0.2853	1.436	0.603	1.574	0.100	0.0768		1.229		3.197		
Expon accum	1	0.2875	0.519	0.773	1.813	0.120	0.0666		0.509		1.532		
Expon accum	2	0.3533	0.209	0.851	1.796	0.100	0.0602		0.959	0.502	2.098	1.835	
Expon accum	3a	0.4810	0.522	2.286		0.200	0.0423	1.606	0.951	0.874	0.574	0.367	0.144
Expon accum	3b	0.4742	0.378	1.973		0.200	0.0423	1.713	0.950	0.840	0.504	0.350	0.092

Note. Experiments 1 and 2 have speed–accuracy manipulations; Experiment 3 manipulates relative stimulus probability. k_1 to k_5 are used as generic decision criteria for these experiments. For Experiments 1 and 2, k_1 and k_2 represent speed criteria, and k_3 and k_4 represent accuracy criteria. For the rectangular accumulator, k_1 and k_3 are the means of rectangular distributions for the speed and accuracy conditions, with ranges of κ_s and κ_a, respectively. For the exponential accumulator, k_1 to k_4 are offsets to exponential distributions of criteria. The means of the exponential portion of the distributions are κ_s and κ_a for the speed and accuracy conditions, respectively. For Experiment 3, $k_a + k_b$, the sum of the offset parameters for the two distributions of criteria, was held constant as the relative offset was varied. The parameters k_1 to k_5 are the offsets for one distribution of criteria; $k_a + k_b$ minus the values of k_1 through k_5 are the offsets for the other. Experiment 3a refers to fits to Experiment 3 with no drift criteria, and 3b refers to fits to Experiment 3 with drift criteria. T_{er} = mean of the nondecision component of response time; σ_μ = standard deviation in accrual rate across trials; s_t = variability in nondecision component of response time across trials; λ = a constant converting cycles to ms; Rectang = rectangular; accum = accumulator; Expon = exponential.

Table 4
Parameters of the Poisson Counter Models

Model	Experiment	T_{er}	$u + v$	$\alpha + \beta$	κ_s	κ_a	s_t	$k_a + k_b$	k_1	k_2	k_3	k_4	k_5
Rectang Poiss	1	0.2219	1.66	30.33	4.00	9.00	0.0425		4		6		
Geom Poiss	1	0.2559	3.32	37.82	2.43	5.95	0.1000		4		6		
Geom Poiss	2	0.3500	19.15	39.66	2.56	5.29	0.1000		4	3	6	6	
Geom Poiss	3a	0.4105	5.55	32.95	3.76		0.0750	10	7	7	6	5	4
Geom Poiss	3b	0.4100	6.13	32.83	3.70		0.0750	10	7	7	6	5	4

Note. Experiments 1 and 2 have speed–accuracy manipulations; Experiment 3 manipulates relative stimulus probability. k_1 to k_5 are used as generic decision criteria for these experiments. For Experiments 1 and 2, k_1 and k_2 are speed criteria, and k_3 and k_4 are accuracy criteria. For the rectangular Poisson counter model, k_1 to k_4 are means of rectangular distributions with ranges κ_s and κ_a for the speed and accuracy conditions, respectively. For the geometric Poisson counter model, k_1 to k_4 are offsets to geometric distributions of criteria; means of the geometric portion of the distributions are κ_s and κ_a for the speed and accuracy conditions, respectively. For Experiment 3, $k_a + k_b$, the sum of the offset parameters for the two distributions of criteria, was held constant as the relative offset was varied. The parameters k_1 to k_5 are the offsets for one distribution of criteria; $k_a + k_b$ minus the value of k_1 through k_5 are the offsets for the other. $\alpha + \beta$ is the sum of the Poisson rates in the two counters. This was kept fixed across all conditions as the relative accrual rate $\pi = \alpha/(\alpha + \beta)$ varied randomly according to a beta distribution with shape parameters u and v and mean $\rho = u/(u + v)$. The sum $u + v$, which determines the distribution's shape, was held constant across conditions. T_{er} = mean of the nondecision component of response time; s_t = variability in nondecision component of response time across trials; Rectang = rectangular; Poiss = Poisson counter model; Geom = geometric.

There were moderate misses in the .9 quantile RTs for the accuracy condition for some of the models. Better fits can be obtained if the nondecision component of RT is allowed to be different for the two conditions; for example, the fit of the Wiener diffusion model improves with chi-square being reduced to under 10 with two values of T_{er} differing by 25 ms (see Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, in press, for data that may speak to this). Before such assumptions can be made, systematic studies need to be conducted.

From a modeling perspective, there are two salient features of Experiment 1: The data provide RT distributions for correct responses and errors over a range of accuracy values from near chance to near perfect, and for all conditions of the experiment, correct responses are faster than error responses. The latter finding contrasts with the results obtained in many other experimental paradigms, in which errors are typically faster than correct responses. It was for this reason that the second experiment to which we applied the models was a lexical decision experiment. In this experiment, errors were faster than correct responses with speed instructions and slower than correct responses with accuracy instructions. As noted previously, this crossover is a particularly difficult pattern for a model to produce, and so, it offers a stringent test.

Experiment 2: Lexical Decision

The main result of interest is that the relationship between correct and error RTs is altered by instructions. In the speed conditions, mean error RTs were shorter than mean RTs for correct responses, whereas in the accuracy conditions, mean error RTs were longer than mean correct RTs. This crossover pattern is not unusual (e.g., Luce, 1986; Ratcliff & Rouder, 2000; Ratcliff et al., 1999; Smith & Vickers, 1988; Swensson, 1972), and it can also be obtained by a post hoc classification of subjects according to their overall speed, fast or slow (Ratcliff et al., 2004).

The experimental data are not symmetric across the two responses as they were for Experiment 1. For one reason, there were three word conditions (corresponding to different frequency values) and only one nonword condition, and also the data for word and nonword RTs and accuracy values were different. For the

Wiener and OU diffusion models, this means that the starting point is not equidistant between the two boundaries (i.e., z is not equal to $a/2$) and that the starting point, like boundary separation, varies between the speed and accuracy conditions. For the accumulator and Poisson counter models, the criterion value for word responses was different from the criterion value for nonword responses. Both the criterion values and the parameter representing the spread in the distribution of criteria had different values for the speed and accuracy conditions. For all the models, drift or accrual rate varied across the types of stimuli.

Wiener Diffusion and OU Models

The Wiener diffusion model, which was the best fitting of the models ($X^2 = 28.96$), captures the main qualitative features of the data (see Figure 7). Errors for words are faster than correct responses in the speed condition and slower in the accuracy condition. The OU model's fit with $\beta = 4$ was poorer ($X^2 = 44.87$). When β was allowed to vary freely, it converged to a value of zero.
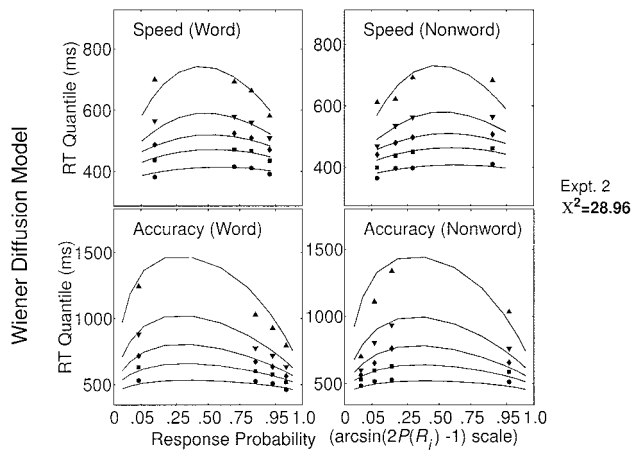

Figure 7. Fits of the Wiener diffusion model to the data from Experiment (Expt.) 2. RT = response time; ● = .1 quantile RT; ■ = .3 quantile RT; ◆ = .5 quantile RT; ▼ = .7 quantile RT; ▲ = .9 quantile RT.

That is, the best-fitting OU model was identical to the Wiener diffusion model.

For both models, the pattern of error RTs relative to correct response RTs is accommodated by the amount of across-trial variability in drift rates, the amount of across-trial variability in starting points, and the boundary positions. Starting point variability is responsible for fast errors: When boundary separation is small, the range of starting points is a greater proportion of the total amount of boundary separation, which leads to fast errors. When boundary separation is large, the range of starting points is a smaller proportion of the total amount of boundary separation, allowing variability in drift across trials to dominate, and errors slower than correct responses result.

Accumulator and Poisson Counter Models

The accumulator model's fit is intermediate between the Wiener and OU diffusion models' fits ($X^2 = 36.77$; see Tables 2, 3, 4, and 5), but it is unable to predict RTs accurately for errors on words in the speed conditions. The predicted error RTs are shorter than the RTs for correct responses, but only by 2–3 ms, considerably underestimating the 20–30 ms effects in the data. The accumulator model's inability to produce fast errors, which are often found empirically, limits its applicability as a general model of two-choice RT tasks. The Poisson counter model is also unable to produce fast errors ($X^2 = 67.02$); it was the poorest fitting of the four models.

Summary

The Wiener diffusion model and the OU model with moderate decay give good accounts of the data. Both the overall shapes of the predicted RT distributions and the ways the shapes change with stimulus difficulty and speed versus accuracy instructions agree with the data, as do the associated values of accuracy. In particular, the models are capable of producing fast errors relative to correct responses in the speed condition and slow errors in the accuracy condition. They do this with only the values of the boundary positions and the starting point varying between the speed and accuracy conditions. The accumulator and Poisson counter models cannot produce errors faster than correct responses in the speed condition, and so, they fail on qualitative grounds.

Experiment 3: Recognition Memory

In Experiment 3 (Ratcliff, 2004; see also Murdock & Anderson, 1975; Ratcliff & Murdock, 1976), subjects studied lists of pairs of words and then were tested for recognition. Test-word difficulty was varied within lists via the number of times a word was presented in the study list (one or four) and Kučera–Francis frequency. Between blocks, the proportion of old to new test words in the test lists was varied across five levels. Subjects biased their responses toward the most likely response, manifested as a shift and skewing of the whole RT distribution including the leading edge. The .1 quantiles for the favored response decreased by about 100 ms, and the .9 quantiles decreased by about 250 ms. The leading edges of the RT distributions did not shift across experimental conditions in Experiments 1 and 2; so, this experiment provides a different test of the models. Also, the between-blocks manipulation was accompanied not only by large changes in RT but also by large changes in accuracy: Responses to the favored alternative were highly accurate (95% correct for the most accurate condition), but responses to the disfavored alternate were not (70% correct for the most accurate condition).

For the within-list variables, the only parameter that could vary in the models was the rate of accumulation of evidence, drift rate in the diffusion models and accrual rate in the accumulator and Poisson counter models. There were six within-block conditions for old test words; they were studied either once or four times, and there were three values of word frequency. For new test words, there were only the three values of word frequency. For the between-lists variable, proportion of old to new test words, two parameters varied in the Wiener and OU models, starting point and drift criterion (Ratcliff, 1985; Ratcliff et al., 1999). Varying the starting point toward the more favored alternative is equivalent to moving the favored response boundary nearer the starting point while moving the other boundary farther away by the same amount.

As mentioned when the diffusion models were introduced, the drift criterion operates in the same way as the criterion in signal detection theory. To illustrate its application to recognition memory, suppose one of the types of old items has drift rate .2 and one of the types of new items has drift rate −.2. Then, when old items are favored, the drift criterion can be adjusted so that the type of old item that previously had drift .2 now has drift rate .3 and the new item previously with drift −.2 now has drift −.1. The difference between the two types of items is constant; the drift criterion adjustment has simply added a value of .1 to both (see Ratcliff et al., 1999, Figure 32).

Changes in starting point and drift criterion have different effects. Changes in starting point produce changes in both the leading edge and skew of the RT distribution, whereas changes in

Table 5

Drift Rates, Accrual Rates, and Chi-Square Values in the Fits of the Models to Experiment 2

Model	v_1	v_2	v_3	v_4	X^2	df	BIC
Wiener	0.4316	0.2749	0.1818	−0.2566	28.96	76	7,889.9
OU ($\beta = 4$)	0.4190	0.2740	0.1840	−0.2371	44.87	75	7,930.8
Exponential accumulator	1.0417	0.6621	0.4306	−0.5296	36.77	74	7,925.6
Geometric Poisson counter	0.8364	0.7365	0.6468	0.3128	67.02	74	7,987.6

Note. v_1 to v_4 are generic parameters. They represent drift rates in the Wiener and Ornstein–Uhlenbeck (OU) models (also denoted v in the text), step increment distribution means in the accumulator model (denoted μ in the text), and the relative accumulation rates in the Poisson counter model (denoted ρ in the text). BIC = Bayesian information criteria.

drift criterion affect mainly the skew of the RT distribution with only small changes in leading edge.

In the accumulator and Poisson counter models, the same types of criteria were allowed to vary as for the diffusion models. The criterion for the favored response was set lower than the criterion for the other response, but the sum of the two was kept constant. Variability in the criterion values, the value of κ, was assumed to be constant across all conditions. For the accumulator model, the equivalent of varying the drift criterion in the diffusion models was to vary the zero point in the distributions of amounts of evidence. This was done by adding a constant to the means of all the word frequency and repetition conditions for both responses. For the Poisson counter model, the relative accrual rate parameter, ρ, was biased by a constant amount, while the sum of the rates in the two counters was held constant (see the Appendix).

Wiener and OU Models

The Wiener diffusion model fits the data well with only starting point varying and the drift criterion fixed at zero ($X^2 = 241.77$). When the drift criterion varied across proportion conditions, the fit is only slightly better ($X^2 = 240.44$). The OU model with moderate decay ($\beta = 4$) fits worse than the Wiener diffusion model when only starting point was free to vary ($X^2 = 277.53$), but almost as well when drift criterion was free to vary ($X^2 = 247.62$). The adjustment in drift criterion was small, no larger than 8% of the average drift rate. The fit of the Wiener diffusion model with a fixed drift criterion is slightly superior numerically to the OU model, even when drift criterion is free to vary for the OU model. For the OU model, when only one value of drift criterion was allowed and the decay parameter was allowed to vary freely, the best-fitting value of decay approaches $\beta = 0$; that is, the OU model approaches identity with the Wiener model. When drift criterion was also free to vary, the estimate of β shows little tendency to vary from its starting value of 4.

The main features of the data from Experiment 3 that differentiate it from the first two experiments are the shift in the leading edge of the RT distribution to faster responses for the favored response alternative and the large decrease in accuracy for the disfavored response alternative. In the diffusion models, the movement of the starting point nearer to the favored response boundary is responsible for the change in the leading edge of the RT distribution and the change in accuracy.

Figure 8 shows an example of fitted values for the Wiener model with drift criterion and starting position free to vary. The model captures the major trends: Error responses are generally slower than correct responses and tend to become faster at the extremes. The RT distributions shift as a function of probability condition, with decreases in the .1 quantile for the more probable response. Also, when the probability of old items is high, the QPF for "old" responses is shifted to the right, indicating a bias toward these responses. When the probability of old items is low, the QPF for "old" responses is shifted to the left, indicating a bias away from these responses. Qualitatively, the OU model with $\beta = 4$ produces fits of about the same quality as shown in Figure 8.

Accumulator and Poisson Counter Models

Neither model fit the data as well as the diffusion models (see Tables 2, 3, 4, 6, and 7). For the accumulator model, the goodness of
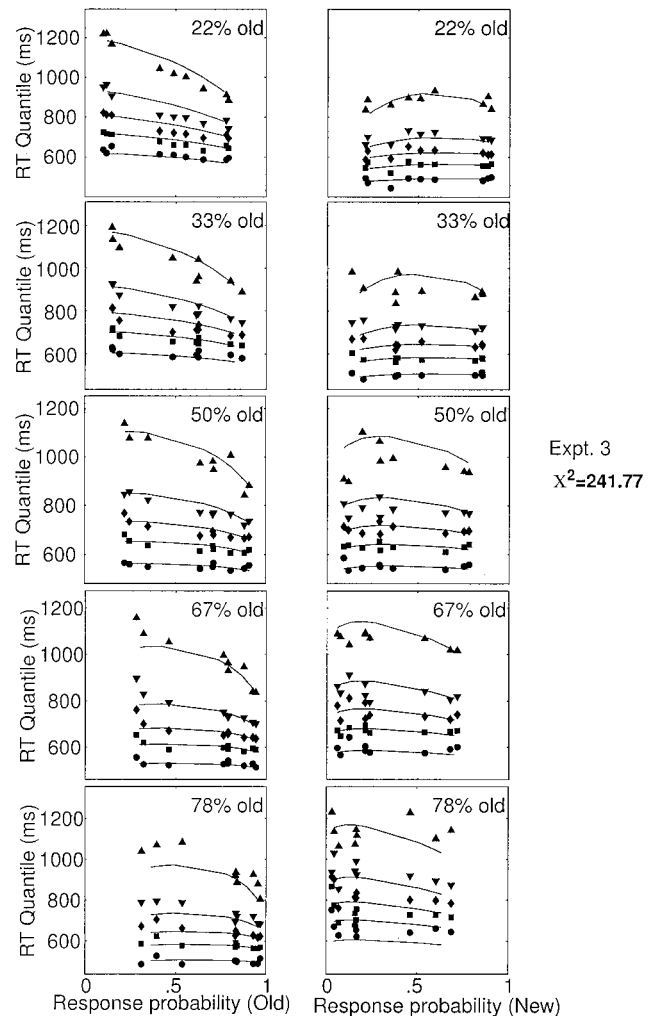

Figure 8. Fits of the Wiener diffusion model to the data from Experiment (Expt.) 3. RT = response time; ● = .1 quantile RT; ■ = .3 quantile RT; ◆ = .5 quantile RT; ▼ = .7 quantile RT; ▲ = .9 quantile RT.

fit without the zero point in the evidence distributions varying was $X^2 = 284.62$; this improved a little with the zero point varying to $X^2 = 280.91$. The Poisson counter model chi-square value was $X^2 = 384.92$, without the accrual rate bias parameter varying, and $X^2 = 365.34$, with it varying. As for Experiments 1 and 2, the accumulator was better than the Poisson counter model at describing the shapes of the RT distributions because its predicted distributions are somewhat more skewed, consistent with the empirical data.

Model Freedom and Model Selection

The BIC provides a basis for model selection that depends on parsimony, goodness of fit, and model complexity. Models with more free parameters are penalized more heavily, and the penalty increases with the size of the sample. When modeling data from speed and accuracy conditions, the accumulator and Poisson counter models need two more parameters than the diffusion models. They therefore incur larger penalties than do the diffusion models. As a result, the Wiener diffusion model is selected by the

Table 6
Drift Rates, Accrual Rates, and Chi-Square Values in the Fits of the Models to Experiment 3

Model	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	X^2	df	BIC (10^5)
Wiener	0.2132	0.3015	0.3228	0.1009	0.1430	0.1509	−0.0890	−0.1484	−0.1675	241.77	476	1.422
OU ($\beta = 4$)	0.2408	0.3372	0.3578	0.1212	0.1661	0.1737	−0.0923	−0.1583	−0.1791	277.53	475	1.424
Exp accum	0.6210	0.8283	0.8967	0.3324	0.4375	0.4315	−0.1923	−0.3781	−0.4753	284.62	475	1.424
Geom Poiss	0.7282	0.8165	0.8258	0.6385	0.6725	0.6824	0.4605	0.3940	0.3753	384.93	475	1.438
Wiener	0.2154	0.3058	0.3272	0.0997	0.1426	0.1506	−0.0973	−0.1587	−0.1781	240.44	472	1.423
OU ($\beta = 4$)	0.2832	0.3666	0.3865	0.1785	0.2185	0.2250	−0.0018	−0.0582	−0.0751	247.62	471	1.423
Exp accum	0.6052	0.8133	0.8219	0.3440	0.4147	0.4508	−0.0839	−0.1992	−0.2656	280.91	471	1.425
Geom Poiss	0.7049	0.7922	0.8009	0.6158	0.6500	0.6593	0.4423	0.3772	0.3587	365.34	471	1.433

Note. v_1 to v_9 are generic parameters. They represent drift rates in the Wiener and Ornstein–Uhlenbeck (OU) models (also denoted v in the text), step increment distribution means in the accumulator model (denoted μ in the text), and the relative accumulation rates in the Poisson counter model (denoted ρ in the text). The top four rows present parameters for fits to Experiment 3 without drift criteria, and the bottom four rows present parameters for fits to Experiment 3 with drift criteria. BIC = Bayesian information criteria; Exp accum = exponential accumulator; Geom Poiss = geometric Poisson counter.

BIC statistic as the best model for Experiment 1 by a small margin, although the accumulator with exponential criteria had the smallest chi-square. Similarly, for Experiment 2, the OU model with $\beta = 4$ is deemed to be a better model by BIC than the accumulator model, despite the latter's smaller chi-square. Apart from these cases, the use of statistics that take into account model freedom compared with a more traditional chi-square statistic appears to have had little influence on the relative quantitative goodness of fit of the models in these experiments.

The four models all have approximately similar numbers of free parameters for Experiments 1 and 2, which each have 88 degrees of freedom in the data. For the signal detection data in Experiment 1, the Wiener diffusion model fit the data well with 10 parameters: 4 drift rate parameters, 1 for each of the dot separation conditions; a pair of boundary separation parameters (1 for speed conditions and 1 for accuracy conditions); and 1 parameter each for T_{er}, variability in T_{er} (s_t), variability in drift across trials (η), and variability in starting point across trials (s_z). The OU model adds a decay term, β, to these parameters, making its total number of parameters $M = 11$. It also fit the data well so long as the value of β was small. For the lexical decision in Experiment 2, the Wiener diffusion model gave a good fit for the data with $M = 12$: 4 drift rate parameters (1 for each of the word frequency conditions plus 1 for nonwords) plus a pair of starting point parameters, z, necessary because the data were not symmetric for the two responses, plus the same other parameters as for Experiment 1. The OU

model also fit well with the same parameters as the Wiener diffusion model plus $\beta = 0$.

The accumulator model accounts accurately for the data from Experiment 1 with $M = 12$. It has two more parameters than the Wiener diffusion model because it requires the parameter λ, which maps discrete to continuous time, and it requires different values of criterion variability (the mean of the exponential) for the speed and accuracy conditions. In contrast, the diffusion models use the same value of starting point variability for both the speed and accuracy conditions. The model has two additional free parameters for Experiment 2 because there are different decision criteria for word and nonword responses for both the speed and accuracy conditions, leading to $M = 14$ parameters. However, the model cannot fit the data because it cannot produce error responses faster than correct responses.

The Poisson counter model also has two more free parameters than the Wiener diffusion model for Experiments 1 and 2. One is the overall accumulation rate parameter, $\alpha + \beta$, which corresponds to the parameter λ, which maps counts onto time in the accumulator model. The other is an additional criterion variability parameter that is required because, like the accumulator model, the model requires two criterion variability parameters to represent speed versus accuracy conditions. The Poisson counter model is not successful with the data from either Experiment 1 or 2 because it does not produce distributions as skewed as the data and it produces predictions for error RTs that are too long.

Table 7
Drift Criteria (Diffusion Models), Step Increment Distribution Criteria (Accumulator Model), and the Accumulation Rate Criteria (Poisson Counter Model) for the Probability Conditions in Experiment 3

Model	67% old	50% old	33% old	22% old
Wiener	0.0048	0.0025	−0.0106	−0.0163
OU ($\beta = 4$)	0.0257	0.0850	0.1174	0.1388
Exponential accumulator	0.0550	0.1237	0.1308	0.2331
Geometric Poisson counter	−0.2054	−0.2538	−0.1980	0.0114

Note. These criteria are relative to zero criterion for the 78% old (3.5 old:1 new) condition. The parameters represent drift rates in the Wiener and Ornstein–Uhlenbeck (OU) models (denoted v in the text), step increment distribution means in the accumulator model (denoted μ in the text), and the relative accumulation rates in the Poisson counter model (denoted ρ in the text).

The ability of the diffusion models to capture data for both speed and accuracy conditions while holding across-trial variability constant in all model components (starting point, drift rate, and the nondecision components of processing) is an appreciable parsimony advantage. It means that the relative speed of correct and error RTs (which can change from speed to accuracy instructions) and the shapes of RT distributions for speed and accuracy instructions are all accounted for with no change in parameters other than changes in the decision criteria. In contrast, with two levels of speed versus accuracy instructions, the accumulator and Poisson models need two additional criterion variability parameters in addition to the different decision criteria. Additional levels of speed versus accuracy instructions would each require still another variability parameter.

For Experiment 3, there were also similar numbers of free parameters for the four models. The Wiener diffusion model accounted well for the data with nine drift rate parameters (three levels of word frequency for both old and new test words and old test words studied once or four times): a single boundary separation parameter; five starting point parameters (one for each level of the proportion of old to new test words); and the same set of drift variability, starting point variability, mean T_{er}, and variability in T_{er} parameters as were used in Experiments 1 and 2. We let the drift criterion parameter vary freely, but good fits were obtained with only one value of this parameter for all the levels of proportion of old–new test words; so, the total number of free parameters is $M = 19$. The OU model also fit the data well with its additional parameter β set to 4 but with four additional drift criteria parameters ($M = 23$).

The accumulator model produced fits a little worse than the Wiener diffusion model but with $M = 24$ parameters. The fits were qualitatively comparable with those presented in Figure 8. The Poisson counter model produced fits that were a little poorer than those shown in Figure 8, with the main discrepancy in the shape of the RT distributions: The Poisson counter model produced predictions that were more symmetric than the data. This model too had $M = 24$ parameters.

Across the three experiments, when there was no variability in T_{er}, the two counter models performed better than the diffusion models in capturing the changes in the leading edge of the RT distribution, whereas the diffusion models were better at capturing distribution shape, especially in the extreme tails. The diffusion models benefited more than the two counter models by the introduction of variability in T_{er}, mainly through an improved ability to capture changes in the leading edge of the distribution. As discussed previously, the initial rise of the leading edge, which is typically steeper for diffusion models than for counter models, is reduced by the introduction of variability in T_{er}. For diffusion models, this reduction in the initial rise usually brings the models into better alignment with the data. However, from a theoretical perspective, some variability in T_{er} is necessary, and the diffusion models benefit more than the counter models.

Model Mimicry

In the preceding sections, we evaluated the models in terms of their abilities to explain experimental data. The converse of this is to examine how constrained the models are; that is, can they fit any pattern of data at all? Of course, we cannot examine the models' fits to an infinite number of possible configurations of data. In-

stead, to provide a modest examination of flexibility, we generated predictions from one model, using the best-fitting parameter values from the fits to the data from the experiments, and then attempted to fit the predicted data with a second model. For these fits, we set variability in T_{er} to zero and used the WLS fitting method because it is robust to variability in the leading edge of the RT distribution.

As might be expected, we found that the OU with moderate decay and the Wiener diffusion model mimicked each other, that the Wiener diffusion model was unable to mimic the Poisson counter model or the accumulator model for some patterns of predictions, and that the accumulator model was more flexible in mimicking the Poisson counter model than vice versa.

Mimicking Between the OU and Wiener Diffusion Models

Advocates of the OU model have sometimes argued for its superiority over the Wiener model on two grounds: first, that the OU model's decay term imposes an upper bound on the accumulation of evidence and second, that the decay term makes the OU model more neurally plausible (e.g., Usher & McClelland, 2001). However, Ratcliff's Wiener model also has an upper bound on the accumulation of evidence because of across-trial variability in drift rate, as described above. With respect to neural plausibility, single-cell recordings show that neural firing rates saturate as a function of stimulus intensity and that the rates decrease when the stimulus is removed. Both of these properties have been thought better represented by the dynamics of the OU process.

Our results provide an alternative, empirical perspective on the decay issue. The issues with which we were concerned are whether, and to what extent, decay has a measurable influence in tasks requiring simple judgments about suprathreshold stimuli. The fits already presented and those to follow show that it is difficult to distinguish between the Wiener and OU models in typical data. When the decay term is large ($\beta = 8$ in our parameterization), so that the asymptote of the average sample paths falls inside the response boundaries, the two models are distinguishable, and the OU model's predictions are qualitatively inconsistent with the data. But when decay is moderate ($\beta = 4$), so that the asymptote of the average sample paths falls outside the correct response boundary, the quantitative and qualitative properties of the models are similar. When the decay term is allowed to vary freely in fitting the model to data, the best fits are obtained when it is 0. These findings lead us to conclude that the effects of decay, to the extent that it is present, are at most moderate. Further, small to moderate amounts of decay are virtually indistinguishable empirically from the absence of decay.

To amplify this point, Figure 9 (top left panel) shows fits of the Wiener model to predictions from the OU model with $\beta = 4$ for the data from Experiment 1. The two sets of QPFs coincide almost exactly ($SSE = 14.2$). Overall, the two models exhibit the same qualitative features, and small differences in chi-square values notwithstanding, a decisive conclusion that one model fits a set of experimental data successfully and the other model fails is unlikely.

Mimicking Between the Wiener Diffusion and the Accumulator and Poisson Counter Models

The question addressed in this section is whether the Wiener diffusion model is so flexible that it can fit predictions from other
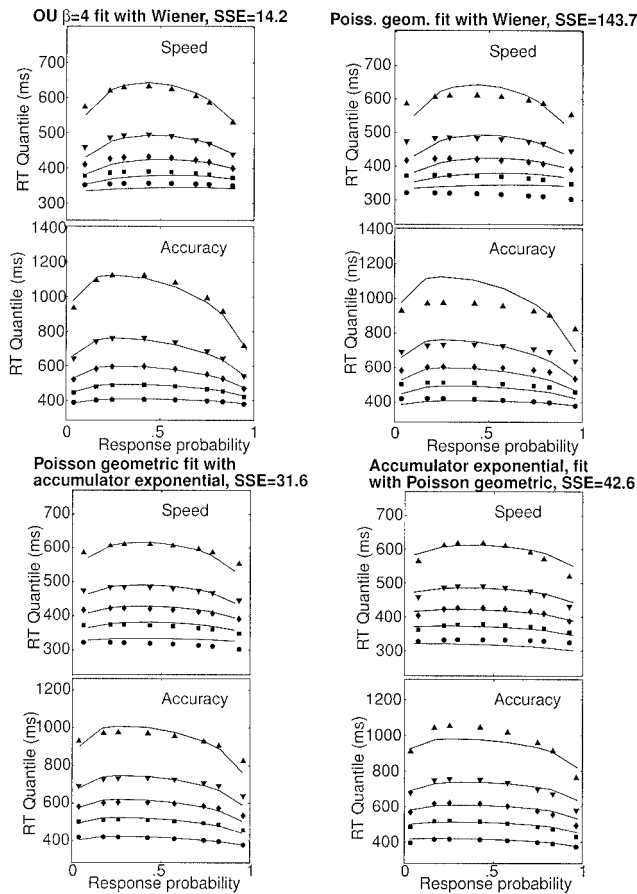
Figure 9. Fits of the Wiener diffusion model to predictions from the Ornstein–Uhlenbeck (OU; $\beta = 4$) and the Poisson counter models with geometrically distributed criteria, fits of the Poisson counter model with geometrically distributed criteria to the accumulator model with exponentially distributed criteria, and fits of the accumulator model with exponentially distributed criteria to the Poisson counter model with geometrically distributed criteria. Poiss. geom. = Poisson geometric; RT = response time; ● = .1 quantile RT; ■ = .3 quantile RT; ◆ = .5 quantile RT; ▼ = .7 quantile RT; ▲ = .9 quantile RT.

models, even when those predictions do not closely resemble empirical data. Ratcliff (2002) showed that the Wiener diffusion model could not fit a number of artificially constructed data sets, data with distributions that were more skewed or less skewed than typical empirical distributions, and data for which the leading edges of the RT distributions changed much more across conditions than is found in empirical data. He concluded that the model is well constrained.

Figure 9 (top right panel) shows that the Wiener model fails to fit predictions generated from the Poisson counter model with geometrically distributed criteria ($SSE = 143.7$) using the best-fitting parameters for the data from Experiment 1. The discrepancies were large, including 30 ms in the .1 quantiles and 150 ms in the .9 quantiles, enough to discriminate between the two models. Although not presented here, the discrepancies between the Wiener model and the Poisson counter model with rectangularly distributed criteria were even greater because the predicted distributions with rectangularly distributed criteria are much more symmetric than with geometrically distributed criteria.

The Wiener model fit the predictions of the accumulator with exponential criteria well ($SSE = 35.3$). This is unsurprising because both models fit the data from Experiment 1 reasonably well. There are discrepancies of up to 70 ms in the .9 quantile RTs in the accuracy condition, with the Wiener model predicting higher values. Apart from these few extremes, the two models cannot be distinguished using the data from Experiment 1. The variability in the data is greater than the differences between the predictions from the two models.

The more interesting case is the one in which the accumulator model with rectangularly distributed criteria did not fit the data from Experiment 1 well. The fit of the Wiener model to the accumulator was not as good as for the exponential case ($SSE = 59.8$ vs. 35.3). There were discrepancies of up to 100 ms in the .9 quantiles and systematic discrepancies in the remaining quantiles for error responses in the accuracy conditions. These differences are large enough to allow discrimination on the basis of relative goodness of fit.

Generally, the Wiener model is unable to produce the flat QPFs that are characteristic of the Poisson counter model or the somewhat less flat QPFs that are obtained for the accumulator with normal or rectangular distributions of criteria. This lack of flexibility in the Wiener model is consistent with the results of the simulations by Ratcliff (2002). It shows that the success of diffusion models in fitting experimental data cannot be attributed simply to their flexibility. Rather, their range of predictions, like that of counter models, has identifiable bounds, and only data that fall within them can be fit successfully.

Mimicking Between the Accumulator and Poisson Counter Models

The fit of the accumulator with exponentially distributed criterion values to the predictions of the Poisson counter model with geometrically distributed criterion values was better than was the converse ($SSE = 31.6$ vs. 42.6; see Figure 9). In the latter case, there were discrepancies of up to 70 ms in the .9 quantile for the accuracy condition and up to 20 ms in the .1 quantile for the speed condition. With rectangular distributions of criteria, the accumulator also fit the predictions of the Poisson counter model fairly well ($SSE = 25.5$). In contrast, the Poisson counter model with rectangular criteria fit the predictions of the accumulator with rectangular criteria poorly ($SSE = 121.6$), with discrepancies in the .1 quantile of up to 30 ms and in the .9 quantile of up to 120 ms. The accumulator with exponential criteria fit the predictions of the Poisson counter model with rectangular criteria well ($SSE = 40.7$).

The conclusion we draw from these comparisons is that the accumulator and Poisson counter models mimic each other closely within the regions of the parameter spaces that are used to account for data like those from Experiment 1. However, the accumulator appears to be more flexible than the Poisson counter because it can accommodate a wider range of patterns of data. The data from Experiment 1 are representative of data from a large number of paradigms, and so, the conclusion about model mimicry has wide generality.

Neurally Inspired Accumulator Models

We use the term *neurally inspired* to refer to a class of recent models that combine the attributes of counter models and diffusion

models. They are neurally inspired in that they have features, such as a limit on evidence or activation in the accumulators and inhibition between accumulators, that have been argued to be aspects of neural processing. They potentially offer the capability of relating behavioral data and neural data (Gold & Shadlen, 2001; Ratcliff, Cherian, & Segraves, 2003; Roitman & Shadlen, 2002; Smith & Ratcliff, in press).

All three of the models we discuss are like counter models in that evidence is accumulated in separate counters for the two responses, but the accumulation processes themselves are modeled as diffusion processes. In the first of the models, developed by Usher and McClelland (2001) and termed by them the *leaky competing accumulator model* (see Figure 10 and the Appendix), evidence is continuously distributed and accumulates in continu-
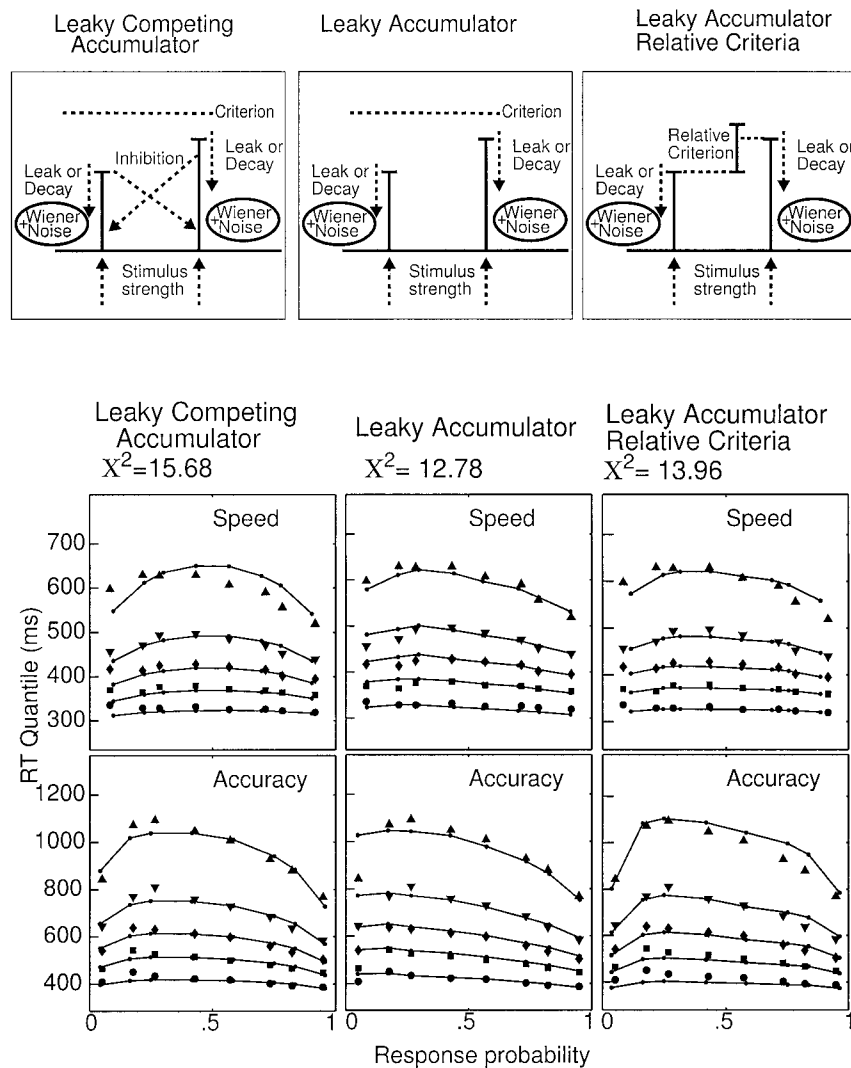
Figure 10. Illustrations of the Usher and McClelland (2001) leaky competing accumulator, the leaky accumulator, and the leaky accumulator with a relative criterion (top) and fits of these models to data from Experiment 1 (bottom). Parameters of the fit for the leaky competing accumulator were as follows: speed criterion = 1.30; accuracy criterion = 1.94; criterion range = 1.02; T_{er} = 276 ms; s_t = 7.3 ms; four accrual rates = .556, .676, .759, and .965; inhibition (β) = 3.49; decay constant (k) = 0.077; and standard deviation in within-trial noise (σ) = 0.675. The parameters for the leaky accumulator were as follows: speed criterion = 1.30; accuracy criterion = 1.91; criterion range = 0.372; T_{er} = 240 ms; s_t = 114 ms; four accrual rates = .555, .685, .780, and .994; decay constant (k) = 0.309; standard deviation in within-trial noise (s) = 0.503; and variability in drift rate across trials = 0.240. The parameters for the leaky accumulator with a relative criterion were as follows: speed criterion = 0.801; accuracy criterion = 1.34; criterion range = 0.203; T_{er} = 233 ms; s_t = 104 ms; four accrual rates = .555, .678, .755, and .954; decay constant (k) = 0.308; standard deviation in within-trial noise (s) = 0.627; and variability in drift rate across trials = 0.178. RT = response time; T_{er} = mean of the nondecision component of RT; s_t = range in rectangular distribution of nondecision component of RT; ● = .1 quantile RT; ■ = .3 quantile RT; ◆ = .5 quantile RT; ▼ = .7 quantile RT; ▲ = .9 quantile RT.

ous time, just as in other diffusion process models. The rate at which evidence accumulates in each accumulator, that is, the equivalent of drift rate of the diffusion process, is a combination of the quality of the information from the stimulus and two other components: One is decay in the amount of accumulated evidence, with decay growing as the amount of evidence in the counter grows, and the other is inhibition from the other counter, with the amount of inhibition growing as the amount of evidence in the other counter grows. If inhibition is large, the model exhibits features similar to the random walk and diffusion models because an increase in evidence for one alternative produces a decrease in evidence for the other alternative. In its assumption of cross-coupling between counters, the model also resembles an earlier, discrete-time model proposed by Heuer (1987). Variability in the diffusion processes is represented by Gaussian noise as in the diffusion models, and decay in the amounts of evidence in the counters corresponds to decay in the OU model.

The second model, the leaky accumulator, is the same as the first except that there is no inhibition between the counters (Smith, 2000). The third model also does not have inhibition, and it has a different decision rule. Instead of a decision being made when one of the counters attains a criterial amount of evidence, a decision is made when the evidence in one counter exceeds that in the other by a criterial amount (i.e., a relative rather than absolute stopping rule). For this reason, we termed it the *leaky accumulator with relative criteria.* For all three models, if the amount of evidence in a counter would become negative because of negative samples of noise or high inhibition from the other counter, it is instead set to zero (a form of rectifying nonlinearity; cf. Smith, 2000).

We fit all three models to the data from Experiments 1 and 2. As we show later, all three fit the data from Experiment 1 quite well. For the data from Experiment 2, two of the models fit reasonably well, but the leaky accumulator model (the one with an absolute stopping rule and no competition) did not produce error responses faster than correct responses.

In Usher and McClelland's (2001) article, the initial presentation of the leaky competing accumulator model had no across-trial variability in any of its components. Later in the article, they added across-trial variability in the starting points of the counters to accommodate fast errors, and we included this in all three models. In the standard accumulator model, variability in the starting points of the counters would be equivalent to variability in the response criterion values. However, in the leaky competing accumulator model, they are not the same because an increased starting point in one counter produces inhibition in the other counter whereas a reduced response criterion does not. Here, we present only data from models with across-trial variability in starting points.

Another source of across-trial variability for the other models considered in this article is variability in drift or accrual rate. Usher and McClelland (2001) showed that the leaky competing accumulator model does not need this source of variability to produce error RTs slower than correct RTs. They acknowledged that for some paradigms, variability in drift should be included because it is implausible to assume that each stimulus in an experimental condition provides exactly the same information to the decision process. For the leaky competing accumulator model, we followed Usher and McClelland and simulated the model without across-trial variability in drift.

For the other two models, the leaky accumulator and the leaky accumulator with relative criteria, we did include variability in drift rate across trials. We modeled variability by selecting a drift rate value from a normal distribution for which the standard deviation was a parameter of the model; the value was truncated to 0 if the value selected was below 0 and to 1 if the value selected was above 1. In addition to the components of the decision process, across-trial variability in the nondecision components of processing was assumed. Just as for the other models, the distribution was uniform with mean T_{er} and range s_t.

The traditional models have been amenable to exact numerical solution as detailed in the Appendix, but no exact solutions are available for the Usher and McClelland (2001) model or the leaky accumulator with relative criteria, and they must be evaluated by simulation. We also evaluated the leaky accumulator by simulation because it involved a change in the fitting program of only two lines of code.

Usher and McClelland (2001) used a metropolis algorithm for fitting simulations of the model to data. In essence, this method generates random sets of parameter values and evaluates the function at each of the sets of parameter values. Parameters that yield the better values of the fit statistic are retained; those that yield poor values are discarded. The range of the set of randomly chosen candidate parameter values is reduced on each iteration of the algorithm until a stable set of values is obtained. For all three accumulator models, we used the SIMPLEX algorithm, which can be set up to operate similarly to metropolis. Both methods are known to be robust with poorly behaved objective functions like those generated by simulations of a model with components that randomly vary across trials. In fitting each of the models to data, all the parameters were free to vary, but as for the traditional models evaluated above, only drift rate can vary between experimental conditions and only the decision criteria can vary between speed and accuracy instructions. However, the resulting fits, regardless of the method used, should be viewed as approximations rather than exact fits because they are produced with simulations and can get only as close to the values of the parameters that provide the best fits as error in the simulated predictions allows.[3]

For Experiment 1, all three models fit the data about as well as did the Wiener diffusion model, the OU model with decay $\beta = 4$, and the accumulator model with exponentially distributed criteria (see Figure 10), with chi-square values of 15.68, 12.78, and 13.96 (and BIC values of 8,210, 8,190, and 8,210) for the leaky competing accumulator, the leaky accumulator, and the leaky accumulator with relative criteria, respectively ($M = 12$). The RT distributions have the appropriate right skew, and the QPFs show error responses slower than correct responses with both speed and accuracy instructions just as the data do. The degree of asymmetry in the QPF is controlled by the amount of inhibition in the leaky competing accumulator model and by the amount of variability in

---

[3] In principle, a model with coupled processes may be approximated using a finite-state Markov chain model, using an approach similar to that pioneered by Pike (1966) and used more recently by Busemeyer and Townsend (1993) and Diederich (1995, 1997). However, as there have not yet been any studies published applying these methods to multivariate diffusion models of RT, we chose to follow Usher and McClelland (2001) and investigate this model by simulation.

drift across trials in the other two models, with larger values leading to slower errors. The size of inhibition in the leaky competing accumulator was 3.49, which is quite large and leads to a large amount of suppression (the counter that starts higher usually wins). The effects of speed versus accuracy instructions are the result of changes in the values of the response criteria.

For Experiment 2, the leaky competing accumulator and the leaky accumulator with a relative criteria were able to produce errors faster than correct responses with speed instructions and slower than correct responses with accuracy instructions. However, the models could produce changes in the leading edges of the RT distributions (i.e., the .1 quantiles) for word responses across the three word frequency conditions that were only about half those observed in the data, making the models' fit poorer than the fits of the Wiener diffusion model and the accumulator model with exponentially distributed criterion values (X^2 for the leaky competing accumulator and the leaky accumulator with relative criteria were 36.11 and 38.49, respectively, and the BIC values were 7,923 and 7,910, respectively). The leaky accumulator did not produce errors faster than correct responses, although the chi-square value was in the same range as for the other two models ($X^2 = 37.61$ and BIC = 7,928).[4]

Summary

Although the leaky competing accumulator model fits the data from our experiments reasonably well, there are several aspects of that model that deserve further consideration. One is the use of inhibition between the counters as the mechanism by which error responses that are slower than correct responses are produced. The problem with explaining slow errors with inhibition between counters is that the inhibition would ordinarily be thought to be an architectural feature of the cognitive system carrying out the task, rather than a feature that depends on stimulus materials or instructional set. However, as the data presented here show, the relative speeds of correct and error responses are not constant across experimental settings but instead vary with stimuli and instructions. For a purely inhibition-based account to be plausible, it must explain why inhibition is high for some stimuli and instruction sets but not others and show how this explanation leads to the relative speeds of correct and error responses that are obtained in experimental data. If the model is augmented by variability in drift rate across trials (Usher & McClelland, 2001), inhibition and variability in drift rate across trials will covary. Therefore, it will be difficult to identify what proportion of a slow down in error RTs relative to correct response RTs is due to each factor.

A second issue for the leaky competing accumulator model also concerns inhibition. Inhibition between the two counters serves to make the behavior of the Usher and McClelland (2001) model sensitive to initial conditions, especially when inhibition is large, as it is in the fits to the data from Experiments 1 and 2. Also, when inhibition is large, the distribution of the difference in amount of evidence between the two counters becomes bimodal (e.g., Usher & McClelland, 2001, Figure 5, bottom right panel). Underlying this bimodality in the difference are bimodal distributions of evidence in each counter individually: If one counter has a lot of evidence, the second counter has little evidence, and inhibition is large, the evidence in the second counter is suppressed to zero. Thus, for some proportion of trials, one counter is active and the other is not, and vice versa for the other trials.

For the parameter values we used to produce the fits for Experiment 1, inhibition was large, and the behavior of the model was strongly dependent on its initial conditions. On the majority of simulated trials, the counter that had the largest amount of evidence initially was the counter that ultimately won. If one counter has moderately higher evidence than the other (because of the initial few values of random noise, ξ_1 and ξ_2 in the equations in the Appendix), then the inhibitory coupling between the counters suppresses evidence in the counter with the lower amount of evidence, thus amplifying the effects of initial noise. This feature of the model's dynamics is somewhat at odds with the presumed biological function of sequential sampling, which is to improve reliability of the decision process by averaging out the effects of processing noise.

Growth of Accuracy

The leaky competing accumulator model, like the OU model without across-trial variability in drift, is able to explain data from response signal experiments. Both models assume that there are no response boundaries for the response signal task, that is, processing is time limited rather than information limited, and they predict a roughly exponential growth of accuracy to an asymptote that is typical of response signal data (see Busemeyer & Townsend, 1992; Usher & McClelland, 2001, Figure 5). Usher and McClelland (2001) argued that the shape of the time–accuracy function in response signal experiments supports an OU model as an approximation to their leaky competing accumulator model, and they argued against the Wiener diffusion model with variability in drift across trials (applied to response signal data with the assumption of time- not information-limited processing; Ratcliff, 1988) because the OU model fits their data better. However, the leaky competing accumulator model and the OU model without across-trial variability in drift, unlike the Wiener diffusion model with across-trial variability in drift, have the problem that they allow accuracy to grow without bound in standard RT tasks (see also Busemeyer & Townsend, 1993).

We illustrate the problem with an extremely simplified version of the OU model (without across-trial variability in drift): The accumulation of evidence is assumed to take place at discrete, very widely spaced, time steps. The model produces distributions of amounts of accumulated evidence across trials that are normal at each step. The top panel of Figure 11 shows two distributions from the OU model at each of three time steps, one distribution for processes for which the top response is correct and one for processes for which the bottom response is correct. For illustrative purposes, the full normal distributions are shown as they would be without response boundaries. We have drawn in boundaries, however, because they are necessary to represent a standard RT paradigm. In reality, processes that hit boundaries terminate, thus reducing the number of decision processes remaining in the distribution for the next time step (see Ratcliff, 1988, Figures 2 and 3). In the figure (top panel), the curved lines represent the means of the processes for which the top boundary is the correct response

---

[4] We recently found that the leaky accumulator can produce errors faster than correct responses if starting point variability is large. We obtained better fits by assuming starting point variability with negatively correlated starting points ($X^2 = 26.53$).
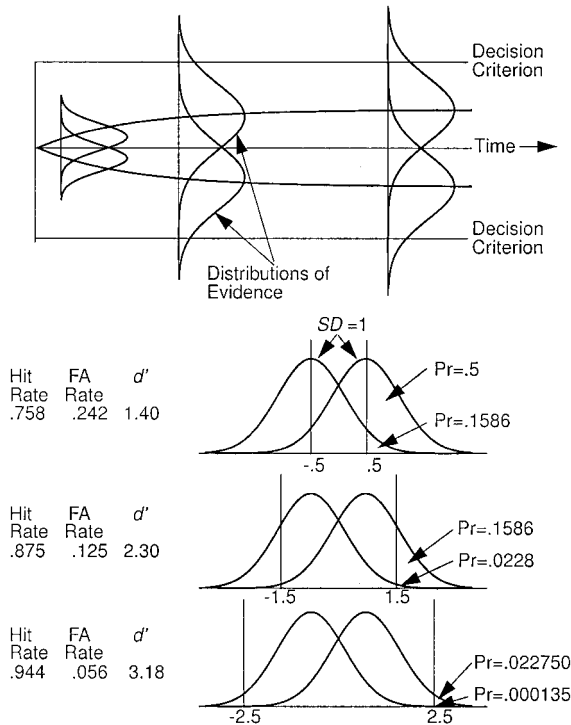
Figure 11. Illustration of the growth of evidence in the Ornstein–Uhlenbeck and Usher and McClelland (2001) models (top). The Usher and McClelland model would have low inhibition between counters. The bottom three panels show the hit rates and false alarm (FA) rates for repeated sampling from the asymptotic distributions of evidence. The values labeled Pr are the hit and FA rates for single samples, and the overall hit and FA rates are computed from repeated samples using the single-sample probabilities (e.g., .758 = .5/[.5 + .1586]).

and the processes for which the bottom boundary is the correct response. The processes reach asymptote after the second time step and before the third.

To demonstrate how accuracy can grow without bound, suppose the first time step occurs after the decision process has asymptoted, that is, at a point at which the two distributions are no longer changing. The bottom three panels of Figure 11 show three cases, each with the same normal distributions (with $SD = 1$) but with increasingly wider boundary positions. The boundary positions are at -0.5 and 0.5 in the first case, -1.5 and 1.5 in the second, and -2.5 and 2.5 in the third (these positions are arbitrarily chosen; the conclusion would be the same with any increasing range of boundary positions).

Consider the first case with boundary positions -0.5 and 0.5. At the first time step, if the accumulated amount of evidence is above the upper criterion or below the lower criterion, a response is made; the proportions of hits and false alarms are .5 and .1586, respectively. The total proportion of processes terminating is .6586, leaving .3413 of the processes to proceed to the next time step. At the second time step, the proportions of hits and false alarms are $.5 \times .3413$ and $.1586 \times .3413$, and only .1164 of the processes proceed to the next step. The sequence of time steps gives a geometric series, with the overall hit rate $.5/(.5 + .1586) = .758$ and the overall false alarm rate .242, resulting in $d' = 1.40$.

The mean number of steps to termination is $1/(1 - .3413) = 1.5$. For the next two panels, d' is 2.30 and 3.18 with criteria at -1.5 and 1.5 and -2.5 and 2.5, with number of steps to termination 5.5 and 43.7, respectively.

To show that the OU model behaves as shown in this illustration, we generated predictions from an OU model with $s = 1$, $\beta = 4$, and $v = .5$, with boundary separation increasing from 0.5 to 2.0 in steps of .5. Accuracy values were .643, .835, .943, and .980; that is, they grew in the way illustrated above.

This illustration shows how accuracy increases without bound as subjects widen their response criteria. This is accompanied by a dramatic increase in RT, as indicated by the increase in the mean number of steps to response. The simplifying assumptions that were made for the purposes of the illustration do not qualify the conclusion, for either the OU model without variability in drift across trials or the Usher and McClelland (2001) model: For both models, in a standard RT paradigm, subjects can be conservative enough in their criterion settings to achieve arbitrarily high values of accuracy. For example, if accurate performance is strongly stressed in instructions, subjects should be able to make their probability of an accurate response approach 1 in any condition in any experiment.

In contrast, in the Wiener diffusion model, slow errors and an asymptotic limit on accuracy come from across-trial variability in drift rate. The assumption of across-trial variability in drift rate arises naturally from the assumption that nominally identical members of a stimulus set are not always processed identically. This is a widely accepted assumption, forming the cornerstone of such classical methods as Thurstonian (Thurstone, 1927) scaling and signal detection theory.

Mathematically, the Usher and McClelland (2001) model is a member of the class of diffusion models, although because of its assumption of coupled processes (inhibition between accumulators), it differs from existing models in important respects. The leaky accumulator with a relative criterion is more consistent with standard models, but it does require communication between the processes to determine whether evidence in one counter exceeds evidence in the other by a criterial amount. We found that both these models can mimic the Wiener and OU diffusion models, just as they mimic each other, but at this point, the Usher and McClelland model has not been applied widely enough for us to be able to determine whether the models have the ability to fit all of the data that the Wiener diffusion model can and at the same time fail to fit data that never appear empirically. In fact, preliminary investigations have shown that the Usher and McClelland model can mimic predictions from the Poisson counter model that are not obtained empirically, namely, symmetric RT distributions and slow errors, but not the other patterns in Ratcliff (2002). However, further systematic and comprehensive studies are needed to fully explore these new models.

Van Zandt et al. (2000) Comparison of the Poisson Counter and Wiener Diffusion Models

Van Zandt et al. (2000) conducted three experiments in which subjects made same–different judgments about pairs of letters. In two of the experiments, subjects were asked to respond before one of three deadlines, and the deadline time was varied between blocks of trials. The third experiment was a standard RT experi-

ment, and the proportion of same versus different pairs was varied between blocks. For all three experiments, Van Zandt et al. concluded that the Poisson counter model provided a better description of the decision process than did the Wiener diffusion model. Here, we argue that the assumptions they made in fitting the deadline experiments were not the most appropriate assumptions for fitting deadline data, and we present new fits of the Wiener diffusion model for their third experiment to show that the model can account for the data satisfactorily.

The procedure Van Zandt et al. (2000) used was similar to a deadline procedure used in a two-choice letter identification experiment by Ratcliff and Rouder (2000, Experiment 2). Ratcliff and Rouder applied the Wiener diffusion model, assuming that the decision process is time controlled, not information controlled. With time control, responses are made when a criterion time is reached, unlike the usual information-controlled process for which responses are made when the accumulated evidence reaches a response boundary. Van Zandt et al. assumed the usual information-controlled process. With time control, when the time criterion is reached, one response is made if the total amount of accumulated evidence is above the starting point, and the other response is made if it is below the starting point (Ratcliff & Rouder, 2000). It would be possible to have both types of control in the Wiener diffusion model for application to deadline data, both a time and information control, but Ratcliff (1988) showed that a model with only time control can mimic a model with both types in situations in which there is a progressive growth of accuracy as a function of time.

One consequence of the use of deadlines in Ratcliff and Rouder's (2000) experiment was that the RT distributions were relatively symmetric, except at the longest deadline, with the RT distributions centering on times slightly longer than the deadline times. Also, as would be expected, the distributions shifted as a function of deadline, with little or no change in skewness or standard deviation. When diffusion models are combined with information-controlled processing, they do not predict symmetric RT distributions; such distributions are more consistent with accumulator and Poisson counter models, as pointed out earlier in this article. Thus, it seems likely that the good fits Van Zandt et al. (2000) found for the Poisson counter model came about because RT distributions were relatively symmetric with the deadline procedure, and it is also likely that the Wiener diffusion model would fit well with time-controlled processing.

Van Zandt et al.'s (2000) Experiment 2 used a standard RT procedure. The proportion of trials for which "same" was the correct response was manipulated between blocks of trials: It was .20, .50, or .80. We fit the Wiener diffusion model to the data from the 3 subjects in this experiment. Van Zandt et al.'s results suggested larger than typical variability in the data; so, we used the WLS method, which is more robust than the chi-square method (see Ratcliff & Tuerlinckx, 2002). Van Zandt et al. fit the Wiener diffusion model under the assumption that drift rates, boundary separation, and starting point were free to vary across probability conditions. In our fits, we made slightly different assumptions. Following Ratcliff (1985), Ratcliff et al. (1999), and Experiment 3 here, we kept boundary separation and the difference between drift rates for same and different stimuli constant across proportion conditions and allowed drift criterion and starting point to vary. Allowing drift criterion to vary meant that the difference between

the drift rates for same and different stimuli remained constant across conditions, but the zero point varied. In addition, unlike Van Zandt et al., we assumed variability in starting point across trials, s_z.

The results are presented in Figure 12, plotted in a form designed to facilitate comparison with the results of Van Zandt et al. (2000). The fits of the model are reasonably good, especially in comparison to those shown in Van Zandt et al.'s Figures 9 and 11. Generally, the observed and predicted accuracy values and mean RTs are close, although the differences in mean error RTs are somewhat variable. More important, the data and predictions fall around a line of unity slope, which means that there are no systematic deviations between theory and data. This contrasts with Van Zandt et al.'s Figure 11, which showed extremely poor fits for accuracy, good fits for correct RTs, and poor fits for error RTs. The parameter values are shown in Table 8.
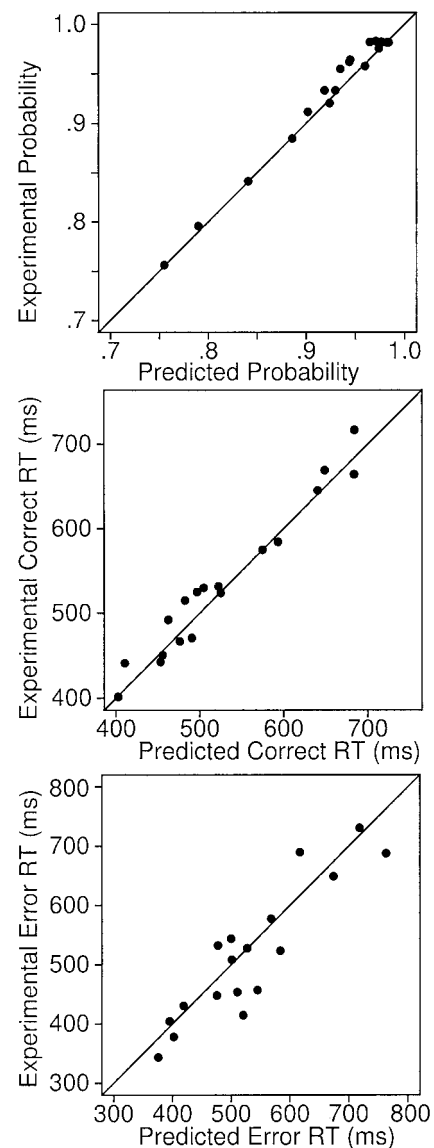

Figure 12. Fits of the Wiener diffusion model to data from Experiment 2 of Van Zandt et al. (2000). RT = response time.

Table 8

Average Parameter Values Across Subjects for the Fits of the Wiener Diffusion Model to Experiment 2 of Van Zandt et al. (2000)

Bias	a	T_{er}	η	s_z	v_s	v_d	z	v_c
80	0.1017	0.3646	0.0252	0.0168	0.2161	−0.3659	0.0795	0
50							0.0589	0.0335
20							0.0374	0.0905

Note. The drift criterion (v_c) is added to both the same and different drift rates. a = boundary separation; T_{er} = mean of the nondecision component of response time; η = variability in drift rate across trials; s_z = variability in starting point across trials; v_s = drift rate for *same* condition; v_d = drift rate for *different* condition; z = starting point.

There are several possible reasons why we obtained good fits of the Wiener diffusion model to Van Zandt et al.'s (2000) Experiment 2 when they did not. One concerns estimates of T_{er} and choice of fitting method. Their estimates of T_{er} from the chi-square method were between 210 and 270 ms, whereas our estimates from the WLS method are between 330 and 400 ms. As noted above, the chi-square method is insensitive to a few short (outlier) RTs, but if there are enough short RTs to affect the lowest quantile RT used in fitting (the .1 quantile in our fits), the chi-square method is extremely sensitive and can produce distortions in the fit and parameter values. For chi-square to be well-defined, T_{er} must be estimated to be smaller than the smallest empirical quantile across all experimental conditions. If there is extraneous variability in the leading edge of the RT distribution, due either to psychological processes not represented in the model or to short outlier RTs, then the chi-square method will set T_{er} too low, and in doing so, it will distort the fit of the model to the rest of the data. In contrast, the WLS fit statistic is more robust to the presence of outliers. Some of the RTs in the .8 probability conditions in Van Zandt et al.'s data are between 200 and 300 ms, which suggests that at least some responses are fast outliers and some proportion comes from variability in T_{er}. The presence of a small proportion of fast outliers and/or extraneous variability in Van Zandt et al.'s data is likely responsible for their poor fit of the Wiener diffusion model to the data from their Experiment 2.

General Discussion

Sequential sampling models are attractive models for simple two-choice decisions because they predict the behavior of accuracy as well as correct and error RTs and their distributions. They provide a way to understand both the speed and the accuracy of performance in a common theoretical framework. In this article, we described comparative fits of sequential sampling models to three sets of experimental data. We also examined whether the models can mimic each other, which of the models have the most flexibility, and whether the models can fit patterns of data that are not obtained experimentally.

The four traditional models we investigated in detail were the Wiener diffusion model, the OU diffusion model, the accumulator model, and the Poisson counter model. For each model, across-trial variability in the drift or accrual rate, in the boundary positions or the starting point of the decision process, and in the nondecision component of RT were assumed. Three neurally inspired models were also examined, the Usher and McClelland (2001) leaky

competing accumulator model and two variants, the leaky accumulator and the leaky accumulator with relative criteria. Across-trial variability in some components of processing was also assumed for these models.

For the Wiener and OU diffusion models, results show that the models mimic each other when decay in the OU model is moderate ($\beta = 4$), and the best fits of the OU model are obtained when there is no decay at all, that is, when the OU model becomes the Wiener model. The models give a good account of the data from all three of the experiments presented here. When stimulus difficulty is manipulated in such a way that subjects cannot change their response criteria or drift criterion according to the type of item being tested, the models have only one free parameter, the rate of accumulation of evidence. Changes in this parameter account for the effects of difficulty on accuracy, the leading edges and skews of the RT distributions, and the relationship between RTs for correct and error responses. The models also give a good account of the effects of between-blocks manipulations, manipulations across which the response criteria and the drift criterion can vary. Subjects can change their criteria to adapt to speed or accuracy instructions or to changing proportions of one kind of test item versus the other.

The OU model has sometimes been argued to be more plausible than the Wiener diffusion model because the decay component of the OU model has been thought to have neural plausibility. However, the best fits of the OU model to the data from the experiments presented here were obtained when the decay parameter was zero, which suggests that the neural plausibility argument should be revisited. It may be that aggregating over populations of neurons averages out decay in these paradigms, or it may be that the standard OU model is not a plausible model.

Usher and McClelland's (2001) leaky competing accumulator model can also explain the data from all three experiments. It has more flexibility than the Wiener and OU diffusion models, in part because the rate of accumulation of evidence is determined by three factors (information from the stimulus, decay, and inhibition from the competing counter). The best fits of the model are obtained when decay is moderate and inhibition is large, both of which lead the model to behave in ways similar to the Wiener model and the OU model with moderate decay. The leaky accumulator with relative criteria fit the data about as well as the leaky competing accumulator, but the leaky accumulator did not produce fast errors (but see Footnote 4).

Both the leaky competing accumulator model and the OU diffusion model without across-trial variability in drift rate have the problem that subjects can make accuracy arbitrarily high by appropriately setting their decision criteria. The models have this problem because the distribution of sample paths asymptotes in a normal or approximately normal distribution. Because the tails of the normal distribution fall off quickly, by setting the response criteria farther and farther from the starting point, a smaller and smaller proportion of errors is made relative to correct responses, and so, accuracy becomes higher and higher. The Wiener diffusion model with variability in drift across trials avoids this problem because there is always some proportion of processes in the lower tail of the distribution of drift rates that has negative drift rates. These processes more consistently hit the error boundary as the boundaries are moved farther apart.

The accumulator model with exponentially distributed criteria fails on only one major aspect of the data, and that is that it cannot predict error responses faster than correct responses. This pattern was found in Experiment 2 and also is typically observed in experiments from the choice RT paradigm. In the diffusion models, the slow errors that come from across-trial variability in accumulation rate can be offset by fast errors that come from across-trial variability in the starting point. However, in the accumulator model, slow errors arise from the underlying structure of the model, and they cannot be offset by across-trial variability in any component of the model.

To fit RT distributions, the accumulator model requires the assumption that the values of the response criteria vary exponentially across trials. Under the assumption that the distributions of criteria become more spread as decision criteria are increased, the model's predicted RT distributions become more skewed instead of becoming more symmetric as they would without this assumption of increased spread in the criteria. The distributions skew because the tails of the RT distributions tend to follow the tails of the exponentially distributed criteria. In contrast, predictions from the Wiener diffusion model and the OU diffusion model are insensitive to the shape of the distribution of variability in starting points.

None of the variants of the Poisson counter model that we considered are able to produce RT distributions that match empirical data, nor are they able to produce error responses faster than correct responses. It may be possible to obtain better fits for this model by relaxing the constraint that overall accrual rate ($\alpha + \beta$) stays constant while the relative rates for the two counters vary. However, we could see no principled way to do this and still retain the assumption that the only model parameters that should vary within a block of trials are those that reflect the difficulty of the stimulus.

We attempted to ensure that the conclusions outlined here generalize beyond the data sets we investigated by choosing the data sets that are widely representative of two-choice paradigms. The data from Experiment 1 (dot separation) are similar to those obtained with other signal detection tasks (e.g., brightness discrimination, numerosity judgments, red–green discrimination, and same–different judgments of brightness; Ratcliff & Rouder, 1998; Ratcliff et al., 1999). The data are also similar to those obtained with a letter discrimination task in which subjects identify backward masked letters. We fit the same four models as in this article to data from this latter task (data from the young subjects in Thapar

et al., 2003) and came to the same conclusions as presented here. Experiment 2 provided data representative of lexical decision experiments, and the finding of errors faster than correct responses with speed instructions is representative of data from many choice RT experiments. Also, for both Experiments 1 and 2, there were large changes in the data as a function of speed versus accuracy instructions, again representative of the effects of such instructions in many experiments. Finally, Experiment 3 provided recognition memory data and a manipulation of bias toward one or the other of the two responses.

The work reported here cannot rule out all versions of models of the unsuccessful types. We have not considered all possible assumptions about how processing components vary, either between or within conditions, and we have restricted our evaluations to stationary models, that is, models in which both the rate of accumulation of information from the stimulus and the response criteria are constant over time. There may be principled ways to relax these constraints, but new assumptions would need to be motivated theoretically.

A major theoretical assumption that underlies all of the work reported here is the assumption of variability across trials in components of processing. Variability of this kind is plausible theoretically and supported experimentally. Indeed, in our view, it is a necessary part of any complete theory of simple decisions. All existing sequential sampling models require some variability of this kind if they are to fit all of the relevant features of experimental data. With across-trial variability in model components, the Wiener and OU diffusion models capture a large number of degrees of freedom in the experimental data with relatively few parameters and with considerable invariance of parameters across experimental manipulations.

A particular strength of the diffusion models is that they predict RT distributions that change in shape with manipulations of difficulty and speed–accuracy instructions in the same way as the data do. For example, the data show that the difficulty of the decisions has only a small effect on the location of the leading edge of the distribution, with most of the change in mean RT across experimental conditions being due to spread in the tail. Conversely, manipulation of speed–accuracy instructions has a large effect on the leading edge of the distribution (about half the size of the total change in mean RT). These behaviors are captured by the diffusion models with considerable economy of parameters. The effects of speed–accuracy instructions are modeled by changes in boundary separation alone; changes in difficulty are modeled by changes in drift rate alone. It is important to note that the Wiener and OU diffusion models cannot be modified to produce RT distributions that behave differently from those presented here, unlike the Poisson counter and accumulator models. For example, the models could not handle data in which stimulus difficulty produced a shift in the leading edge of the RT distribution without an accompanying change in skewness (e.g., Ratcliff, 2002).

In all of the successful models, there is interaction between the accumulating amounts of evidence for the two-choice alternatives. In the Wiener and OU diffusion models and the leaky accumulator with relative criteria, evidence for one alternative is evidence against the other alternative. In the leaky competing accumulator, the more evidence there is for one alternative, the more it inhibits evidence for the other alternative. In terms of neural populations,

these behavioral models suggest that competing populations either inhibit each other or communicate so that relative activity can be monitored (but see Footnote 4).

It is clear that current versions of the Wiener and OU diffusion models, the leaky competing accumulator model, and the leaky accumulator with relative criteria can fit experimental data that are rich and systematic. In fact, it is remarkable that all these models do so well. However, we have also shown that various models of the sequential sampling class can be discriminated from each other on qualitative grounds and that it is possible to understand why particular models fail and under what conditions. The research we have reported here provides a summary of the current state of modeling simple two-choice decisions and also provides a starting point for further evaluation of sequential sampling models.

References

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

Audley, R. J., & Mercer, A. (1968). The relation between decision time and the relative response frequency in a blue-green discrimination. *British Journal of Mathematical and Statistical Psychology, 21,* 183–192.

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

Buonocore, A., Giorno, V., Nobile, A. G., & Ricciardi, L. (1990). On the two-boundary first-crossing-time problem for diffusion processes. *Journal of Applied Probability, 27,* 102–114.

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics, 32,* 117–133.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences, 23,* 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,* 432–459.

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes.* London and New York: Chapman & Hall.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology, 41,* 260–274.

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

Dosher, B. A. (1976). The retrieval of sentences from memory: A speed–accuracy study. *Cognitive Psychology, 8,* 291–310.

Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology, 16,* 519–555.

Falmagne, J. C. (1965). Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology, 12,* 77–124.

Falmagne, J. C. (1968). Note on a simple fixed-point property of binary mixtures. *British Journal of Mathematical and Statistical Psychology, 21,* 131–132.

Feller, W. (1968). *An introduction to probability theory and its applications.* New York: Wiley.

Fortet, R. (1943). Les fonctions aléatoires du type Markoff associées à certaines équations linéares aux dérivées partielles du type parabolique [Random Markov functions associated with certain linear, parabolic partial differential equations]. *Journal de Mathématiques Pures et Appliquées, 22,* 177–243.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review, 108,* 33–56.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Science, 5,* 10–16.

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

Heath, R. A. (1992). A general nonstationary diffusion model for two choice decision making. *Mathematical Social Sciences, 23,* 283–309.

Heuer, H. (1987). Visual discrimination and response programming. *Psychological Research, 49,* 91–98.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, England: Oxford University Press.

Johnson, N. L., & Kotz, S. (1970). *Continuous univariate distributions 1.* New York: Wiley.

Kirby, N. (1980). Sequential effects in choice reaction time. In A. T. Welford (Ed.), *Reaction times* (pp. 129–172). London: Academic Press.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LaBerge, D. A. (1962). A recruitment theory of simple behavior. *Psychometrika, 27,* 375–396.

LaBerge, D. A. (1994). Quantitative models of attention and response processes in shape identification tasks. *Journal of Mathematical Psychology, 38,* 198–243.

Laming, D. R. J. (1968). *Information theory of choice reaction time.* New York: Wiley.

Laming, D. R. J. (1969). Subjective probability in choice-reaction experiments. *Journal of Mathematical Psychology, 6,* 81–120.

Link, S. W. (1975). The relative judgement theory of two choice response time. *Journal of Mathematical Psychology, 12,* 114–135.

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.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 145–194). Hillsdale, NJ: Erlbaum.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308–313.

Ollman, R. T. (1966). Fast guesses in choice-reaction time. *Psychonomic Science, 6,* 155–156.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). New York: Halstead Press.

Pike, A. R. (1966). Stochastic models of choice behaviour: Response probabilities and latencies of finite Markov chain systems. *British Journal of Mathematical and Statistical Psychology, 21,* 161–182.

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

Pike, R., & Ryder, P. (1973). Response latencies in the yes/no detection task: An assessment of two basic models. *Perception & Psychophysics, 13,* 224–232.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491.

Rabbitt, P. M. A. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology, 70,* 305–311.

Rabbitt, P. M. A., & Rogers, B. (1977). What does a man do after he makes

an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29,* 727–743.

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

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Ratcliff, R. (1980). A note on modelling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology, 21,* 178–184.

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

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

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

Ratcliff, R. (2001). Diffusion and random walk models. In *International encyclopedia of the social and behavioral sciences* (Vol. 6, pp. 3668–3673). Oxford, England: Elsevier.

Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9,* 278–291.

Ratcliff, R. (2004). *Fitting reaction time and ROC functions with the diffusion model.* Manuscript in preparation.

Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two choice decisions. *Journal of Neurophysiology, 20,* 1392–1407.

Ratcliff, R., Gómez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 142–165.

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., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214.

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

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

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16,* 323–341.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics, 65,* 523–535.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9,* 438–481.

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

Reed, A. U. (1973). Speed-accuracy trade-off in recognition memory. *Science, 181,* 574–576.

Remington, R. J. (1969). Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology, 82,* 250–257.

Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (in press). On the locus of speed–accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General.*

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 interparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22,* 9475–9489.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Smith, P. L. (1989). A deconvolutional approach to modelling response time distributions. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models* (pp. 267–289). Amsterdam: Elsevier.

Smith, P. L. (1990a). 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. (1990b). Obtaining meaningful results from Fourier deconvolution of reaction time data. *Psychological Bulletin, 108,* 533–550.

Smith, P. L. (1994). Fechner's legacy and challenge [Review of *The wave theory of difference and similarity*]. *Journal of Mathematical Psychology, 38,* 407–420.

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

Smith, P. L. (1998). Bloch's law predictions from diffusion process models of detection. *Australian Journal of Psychology, 50,* 139–147.

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. (2001). Stochastic dynamic models. In *International encyclopedia of the social and behavioral sciences* (Vol. 22, pp. 15115–15121). Oxford, England: Elsevier.

Smith, P. L., & Ratcliff, R. (in press). Psychology and neurobiology of simple decisions. *Trends in Neuroscience.*

Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology, 53,* 293–315.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32,* 135–168.

Smith, P. L., & Vickers, D. (1989). Modeling evidence accumulation with partial loss in expanded judgment. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 797–815.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North Holland.

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

Swensson, R. G. (1972). The elusive tradeoff: Speed versus accuracy in visual discrimination tasks. *Perception & Psychophysics, 12,* 16–32.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18,* 415–429.

Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology, 21,* 136–152.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–287.

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

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

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7,* 424–465.

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

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time distributions: Mixtures and parameter variability. *Psychonomic Bulletin & Review, 2,* 20–54.

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

Vickers, D. (1978). An adaptive module of simple judgements. In J. Requin (Ed.), *Attention and performance, VII* (pp. 598–618). Hillsdale, NJ: Erlbaum.

Vickers, D. (1979). *Decision processes in visual perception.* New York: Academic Press.

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

Wagenmakers, E. M., Farrell, S., & Ratcliff, R. (in press). Estimation and interpretation of 1/f noise in human cognition. *Psychonomic Bulletin & Review.*

Wagenmakers, E. M., Ratcliff, R., Gómez, P., & McKoon, G. (2004). *Strategic control in the lexical decision task.* Manuscript submitted for publication.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44,* 92–107.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85.

Appendix

The Mathematical Models

This appendix describes the mathematical models that were evaluated in the text. The notation used to present the models is as follows. At some time after presentation of the stimulus, one of two responses, R_a or R_b, is made with probability $P(a)$ or $P(b)$, respectively, and response time T. In sequential sampling models, T is identified with the time at which the accumulated stimulus information first exceeds a response criterion or absorbing barrier. This time is referred to as the *first passage time* of the accumulation process. The predicted distribution of decision times in the models is the distribution of first passage times. First passage time probability density functions are denoted by $g(t)$, and (cumulative) distribution functions by $G(t)$. All of the expressions for density functions given here are for joint density functions. That is, they are functions of the form $g_i(t)$, $i = a, b$, where

$$g_i(t)h \approx P[R_i \ \& \ t \leq T < t + h],$$

for small h. (This expression becomes exact in the limit, as h goes to zero.) Conditional density functions are obtained by dividing the functions $g_i(t)$ by their associated response probabilities, $P(i)$, to make the mass in each density function equal to unity. The marginal density function, $g(t)$, which describes the distribution of T, irrespective of the response, is just the sum of the joint densities:

$$g(t) = g_a(t) + g_b(t).$$

Expressed in terms of conditional densities, this expression becomes

$$g(t) = P(a)g(t|R_a) + P(b)g(t|R_b),$$

where $g(t|R_i)$, $i = a, b$ is the conditional density of response R_i. Because a response is always made in finite time in all of the models considered here, it is always the case that $P(a) + P(b) = 1$ (Cox & Miller, 1965).

Diffusion Process Models

A diffusion process is continuous-time Markov process, $X(t)$, whose sample paths are also continuous. Diffusion processes may arise either as the solutions of stochastic differential equations (e.g., Smith, 1995, 2000) or, classically, as the solutions of a pair of partial differential equations: the so-called Kolmogorov backward and forward equations (e.g., Ratcliff, 1978, 1988). Let $f(x, t|z, \tau)$ denote the transition density for the unconstrained diffusion process, that is, the process in the absence of absorbing barriers:

$$f(x, t|z, \tau)h \approx P[x \leq X(t) < x + h|X(\tau) = z],$$

for small h. The transition density satisfies the backward equation

$$-\frac{\partial f(x, t|z, \tau)}{\partial \tau} = \frac{1}{2}\sigma^2(z, x)\frac{\partial^2 f(x, t|z, \tau)}{\partial^2 z} + \mu(z, x)\frac{\partial f(x, t|z, \tau)}{\partial z}, \quad \text{(A1)}$$

and

and its adjoint, the forward (or Fokker–Planck) equation (Cox & Miller, 1965). The latter equation is similar to the backward equation, but partial derivatives are taken with respect to the "forward," or current, state variable, x, rather than the "backward," or initial, state variable, z. Typically, the forward equation is used to characterize the transition density as a function of its current state when its starting point is fixed; the backward equation is used to characterize the density as a function of its starting point when the final state is fixed, as occurs when the process is constrained by absorbing barriers or response criteria.

The first passage time densities, $g_i(t)$, $i = a, b$, also satisfy the backward and forward equations subject to the initial condition $X(0) = z$ and to appropriate boundary conditions (Cox & Miller, 1965, p. 231). The particular diffusion process described by these equations is determined by the functions $\mu(z, x)$ and $\sigma^2(z, x)$ in Equation A1. These functions, which are known as the *drift* and *diffusion coefficient* of the process, respectively, describe the change in $X(t)$ per unit of time as a function of its initial and final state.

The two diffusion process models considered in this article are the Wiener diffusion process and the OU process. For both these models, the diffusion coefficient is constant, $\sigma^2(z, x) = s^2$. The Wiener model also has constant drift, $\mu(z, x) = \xi$, whereas in the OU model the drift depends on $(x - z)$, the difference between the current state of the process and its starting point: $\mu(z, x) = \xi - \beta(x - z)$. The state-dependent part of the drift in the OU model can be interpreted as a restoring force that pulls the process back toward its starting point, the strength of which depends on the distance from the starting point and on the magnitude of the decay constant β. The initial and final states both appear in the notation for the drift and diffusion coefficients in Equation A1 to emphasize that in some diffusions, like the OU model, either the drift or the diffusion coefficient, or both, may depend jointly on these two variables. This explicit representation is useful for the models we consider in this article, in which starting point is allowed to vary across trials.

Spectral Representation of First Passage Time Densities

For a Wiener process with drift ξ, starting position $X(0) = z$, and absorbing barriers at a and b, such that $b < z < a$, the first passage time densities $g_a(t)$ and $g_b(t)$ may be shown to be

$$g_a(t) = \frac{\pi s^2}{(a-b)^2}\exp\left[\frac{\xi(a-z)}{s^2} - \frac{\xi^2 t}{2s^2}\right]$$

$$\times \sum_{k=1}^{\infty} k \exp\left[-\frac{k^2\pi^2 s^2 t}{2(a-b)^2}\right]\sin\left[\frac{k\pi(a-z)}{a-b}\right] \quad \text{(A2a)}$$

(Appendix continues)

$$g_b(t) = \frac{\pi s^2}{(a-b)^2} \exp\left[\frac{\xi(z-b)}{s^2} - \frac{\xi^2 t}{2s^2}\right]$$

$$\times \sum_{k=1}^{\infty} k \exp\left[-\frac{k^2\pi^2 s^2 t}{2(a-b)^2}\right] \sin\left[\frac{k\pi(z-b)}{a-b}\right] \quad \text{(A2b)}$$

(cf. Feller, 1968; Ratcliff, 1978, Equation A9; Smith, 1990a, Equations 16a and 16b).

The probabilities of responding at the upper and lower barriers, $P(a)$ and $P(b)$, may similarly be shown to be

$$P(a) = \frac{\exp(-2\xi z/s^2) - \exp(-2\xi b/s^2)}{\exp(-2\xi a/s^2) - \exp(-2\xi b/s^2)} \quad \text{(A3a)}$$

and

$$P(b) = \frac{\exp(-2\xi d/s^2) - \exp(-2\xi z/s^2)}{\exp(-2\xi a/s^2) - \exp(-2\xi b/s^2)}. \quad \text{(A3b)}$$

It is often more convenient to work with the first passage time distribution functions, $G_a(t)$ and $G_b(t)$, than with the associated density functions. These functions also satisfy the backward and forward equations with appropriate boundary conditions. The distribution function corresponding to Equation A2b may be found in Ratcliff (1978, Equation A12) and Ratcliff et al. (1999, Appendix). Its complement may be obtained via a symmetry argument. The parameterization of diffusion models in this article follows the conventions used in previous publications of Ratcliff, namely, $b = 0$, with z and a being free to vary. The diffusion coefficient is treated as a fundamental scaling parameter of the model whose value is set to $s = 0.1$. When $b = 0$, Equation A2b reduces to Equation A9 in Ratcliff (1978).

Integral Equation Representation of First Passage Time Densities

The first passage time densities for the OU model were computed using an integral equation method developed by Buonocore, Giorno, Nobile, and Ricciardi (1990). This method is described in detail in Smith (2000, 2001). Let $g_a(a, t|z, 0)$ and $g_b(b, t|z, 0)$ be the first passage time density functions for a diffusion process $X(t)$, with initial condition $X(0) = z$, through absorbing barriers a and b, respectively, and let $f(x, t|y, \tau)$ be the transition density of the unconstrained process. The first passage time density functions satisfy the Fortet (1943) equations

$$f(a, t|z, 0) = \int_0^t g_a(a, \tau|z, 0)f(a, t|, a, \tau)d\tau$$

$$+ \int_0^t g_b(b, \tau|z, 0)f(a, t|b, \tau)d\tau \quad \text{(A4a)}$$

and

$$f(b, t|z, 0) = \int_0^t g_a(a, \tau|z, 0)f(b, t|, a, \tau)d\tau$$

$$+ \int_0^t g_b(b, \tau|z, 0)f(b, t|b, \tau)d\tau. \quad \text{(A4b)}$$

Equations A4a and A4b express the unknown first passage time densities as functions of the free transition density of $X(t)$. They are obtained by decomposing the sample paths of $X(t)$ that pass through a pair of open intervals, one above the upper barrier and one below the lower barrier, at time t:

$$f(a, t|z, 0)h \approx P[a < X(t) < a + h]$$

$$f(b, t|z, 0)h \approx P[b - h < X(t) < b].$$

To avoid awkwardness when dealing with limiting cases in which the transition density becomes singular, the barriers are excluded from the intervals—an interpretation that is justified by the continuity of the transition distribution.

Because these intervals lie outside the absorbing boundaries of the process, any sample path that passes through one of them at time t must, of necessity, have made at least one boundary crossing at some time prior to t. For example, Equation A4a describes sample paths that pass through the interval $(a, a + h)$ at time t. As all such paths have made at least one boundary crossing, there must have been a first such crossing, either at a or at b, at some time τ, $\tau < t$. For the path to pass through $(a, a + h)$ at t, the process must make a further transition from a or b into the interval $(a, a + h)$ during the period (τ, t), possibly making further, unspecified, boundary crossings while doing so.

Equation A4a provides an exhaustive and mutually exclusive decomposition of sample paths of this kind. The initial segment of the path, up to the first boundary crossing at a or b, is described by the first passage time densities, $g_a(a, \tau|z, 0)$ and $g_b(b, \tau|z, 0)$, respectively. The subsequent transition into the interval $(a, a + h)$, irrespective of the number of intervening boundary crossings, is described by the densities $f(a, t|a, \tau)$ and $f(a, t|b, \tau)$, respectively, depending on whether the first boundary crossing was at a or at b. Because $X(t)$ is a Markov process, the probability density function for the entire path is given by the product of the two densities, and the integral over τ sums over all possible times at which the first boundary crossing can occur. Equation A4b provides a similar decomposition of sample paths passing through an interval $(b - h, b)$ on the lower barrier at time t.

Mathematically, Equations A4a and A4b are Volterra equations of the first kind, which can be solved analytically only in special cases. Such equations can in principle be solved numerically, by approximating the integrals with sums. However, any attempt to approximate them directly will be numerically unstable, because the transition density $f(x, t|y, t - \Delta)$ approaches a Dirac delta function as the interval of approximation, Δ, becomes small. In the terminology of integral equation theory, the kernel of the equation is singular. Buonocore et al. (1990) showed that these equations could be transformed into Volterra integral equations of the second kind, in which the unknown first passage time densities at time t are expressed as functions of their values at all preceding times, $\tau < t$, and of a kernel function, $\Psi(x, t|y, \tau)$, that goes to zero as τ approaches t:

$$g_a(a, t|z, 0) = -2\Psi(a, t|z, 0) + 2\int_0^t g_a(a, \tau|z, 0)\Psi(a, t|a, \tau)d\tau$$

$$+ 2\int_0^t g_b(b, \tau|z, 0)\Psi(a, t|b, \tau)d\tau \quad \text{(A5a)}$$

and

$$g_b(b, t|z, 0) = 2\Psi(b, t|z, 0) - 2\int_0^t g_a(a, \tau|z, 0)\Psi(b, t|a, \tau)d\tau$$

$$- 2\int_0^t g_b(b, \tau|z, 0)\Psi(b, t|b, \tau)d\tau. \quad \text{(A5b)}$$

In these equations, the kernel function $\Psi(x, t|y, \tau)$ depends on the characteristics of the diffusion process in question. For an OU process with drift $\xi - \beta(x - z)$ and diffusion coefficient s^2 the kernel has the form

$$\Psi(x, t|y, \tau) = \frac{f(x, t|y, \tau)}{2}\left(\beta(x - z) - \xi - \frac{2\exp[-\beta(t - \tau)]}{1 - \exp[-2\beta(t - \tau)]}\right.$$

$$\times \left.\{\exp[\beta(t - \tau)][\beta(x - z) - \xi] - \beta(y - z) - \xi\}\right). \quad (A6)$$

The transition density is the Gaussian density

$$f(x, t|y, \tau) = \sqrt{\frac{\beta}{\pi s^2\{1 - \exp[-2\beta(t - \tau)]\}}}$$

$$\times \exp\left(\frac{-\beta\{(x - z) - \xi/\beta - \exp[-\beta(t - \tau)][(y - z) - (\xi/\beta)]\}^2}{s^2\{1 - \exp[-2\beta(t - \tau)]\}}\right).$$

$$(A7)$$

The preceding equations are special cases of the more general formulas given by Smith (2000, Equations 28 and 58) for an OU process with time-varying drift and absorbing barriers, modified to allow for a nonzero starting point. The kernel function and transition density for the Wiener diffusion process, which are obtained as limiting cases of Equations A6 and A7 when β goes to zero, may also be found in Smith (2000).

Unlike the previous integral equations, Equations A5a and A5b are numerically stable and may be solved recursively. To do so, one must replace the integrals by sum; the values of $g_a(a, t|z, 0)$ and $g_b(b, t|z, 0)$ at each of a sequence of time steps, $k\Delta$, $k = 1, 2, \ldots$, are then computed as functions jointly of their values and of the values of the kernel function at all preceding time steps, $j\Delta$, $j = 0, 1, \ldots, k - 1$. The computational formulas may be found in Smith (2000, Equations 47a and 47b) and in the original article of Buonocore et al. (2000). For the computations reported in this article, the size of the approximating step was set to $\Delta = 10$ ms. When β is small (e.g., $\beta = .01$), the OU model closely approximates the Wiener model. Under these circumstances, statistics for the model computed using the spectral method (Equations A2a and A2b) and the integral equation method agreed to better than 1 ms, thereby providing a useful check on the accuracy of both methods.

The Accumulator Model

General expressions for the first passage time probabilities for an accumulator model with arbitrary increment distributions were derived by Smith and Vickers (1988). The algorithm for the model with normal increments, used here, was described by Smith and Vickers (1989).

The accumulator model assumes that evidence is accrued as a pair of positive, real-valued evidence totals, T_a and T_b, which are initially set to zero. A sequence of normally distributed sensory samples, Z_n, $n = 1, 2, \ldots$, each distributed as $N(z; \mu, \sigma)$, is sampled at equally spaced time points, $t(n) = n\lambda$. The interval between samples, λ, is known as the *inspection time* parameter of the model. Each value of Z_n is classified by comparing it with a sensory referent c. If $Z_n > c$, the quantity $Z^+ = Z_n - c$ is added to the total T_a; if $Z_n < c$, the quantity $Z^- = -(Z_n - c)$ is added to the total T_b. Accumulation continues until $T_a \geq K_a$ or $T_b \geq K_b$, at which point the response associated with the winning total is emitted.

To write expressions for the first passage time probabilities for this model, let $f(z)$ and $g(z)$ be the conditional density functions for Z^+ and Z^-, respectively, and let $p = P(Z_n > c)$ be the probability that a given sensory sample exceeds the referent. The conditional densities are:

$$f(z - c) = \begin{cases} N(z; \mu, \sigma)/p & z > c \\ 0 & z \leq c \end{cases}$$

$$g(c - z) = \begin{cases} N(z; \mu, \sigma)/(1 - p) & z < c \\ 0 & z \geq c, \end{cases}$$

where $N(z; \mu, \sigma)$ is a normal density function, with mean μ and standard deviation σ. Let $f_k(z)$ denote the *k*-fold convolution of $f(z)$, that is, the result of convolving k copies of $f(z)$ with itself,

$$f_k(z) = \int_0^z f_{k-1}(z - x)f(x)dx, \quad (A8)$$

and define $g_k(z)$ as the *k*-fold convolution of $g(z)$ in a similar way. Then $P_a(n)$ and $P_b(n)$, the probabilities of responding R_a or R_b, respectively, after exactly n sample steps, are

$$P_a(n) = \sum_{i=0}^{n-1} \binom{n - 1}{i} p^i(1 - p)^{n-i-1} \int_0^{K_b} g_{n-i-1}(y)dy$$

$$\times p\int_0^{K_a} f_i(x)\int_{K_a-x}^{\infty} f(z)dzdx \quad (A9a)$$

and

$$P_b(n) = \sum_{i=0}^{n-1} \binom{n - 1}{i} p^i(1 - p)^{n-i-1} \int_0^{K_a} f_i(x)dx$$

$$\times (1 - p)\int_0^{K_b} g_{n-i-1}(y)\int_{K_b-y}^{\infty} g(z)dzdy. \quad (A9b)$$

In general, the probability of terminating with a particular response, say R_a, at step n is the product of the probability that at step $n - 1$, both evidence totals are less than their associated criteria (i.e., $T_a(n - 1) < K_a$ and $T_b(n - 1) < K_b$) and the probability that at step n, an increment to T_a is sampled that exceeds $K_a - T_a(n - 1)$. The integrals in Equations A9a and A9b sum the products of these probabilities over all possible nonterminating states at step $n - 1$ and over all values of the terminating increment at step n. In these equations, the joint density of the set of nonterminating states, $[T_a(n - 1), T_b(n - 1)]$, at step $n - 1$, conditional on a sequence of N_a observations favoring R_a and N_b observations favoring R_b, is

$$P[x \leq T_a(n - 1) < x + h_1 \& y \leq T_b(n - 1)$$

$$< y + h_2|N_a = i, N_b = n - i - 1] \approx f_i(x)g_{n-i-1}(y)h_1h_2.$$

The summation over i in Equations A9a and A9b computes these probabilities for all possible sequences of increments to T_a and T_b in the first $n - 1$ steps. (A formal induction proof of this relationship was given by Smith and Vickers, 1988.) The probability of response R_a is then the probability of sampling an increment to T_a at step n that is greater than $K_a - x$, the difference between the R_a criterion and the current T_a total. Similarly, the probability of response R_b is the probability of sampling an increment to T_b that is greater than $K_b - y$. The integrals of f and g over z, in Equations A9a and A9b, respectively, sum over all possible values of z that satisfy this requirement. The double integrals of the joint densities over x and y sum over all points in the rectangle $0 \leq T_a(n - 1) < K_a$, $0 \leq T_b(n - 1) < K_b$ of nonterminating states.

Because the convolution of truncated normal distributions in Equation A8 has no closed form expression, the densities $f_i(x)$ and $g_{n-i-1}(y)$ in Equations A9a and A9b must be approximated numerically. In the algorithm described by Smith and Vickers (1989), the probability density of the increment variable Z is divided into 100 equal steps on the range $-K_b$ to K_a, and the integrals are approximated by sums. Response probabilities and other RT statistics are obtained by appropriate summation of terms (see Smith & Vickers, 1988, 1989).

The Poisson Counter Model

Explicit expressions for the first passage time density function for the Poisson counter model were given by Townsend and Ashby (1983; see also

(Appendix continues)

Van Zandt et al., 2000). The model associates a positive, integer-valued evidence counter, T_a and T_b, with each of the responses R_a and R_b, respectively. Evidence for the two responses accumulates in continuous time in unit increments, independently and in parallel, until $T_a \geq K_a$ or $T_b \geq K_b$. The evidence stream is modeled as a pair of Poisson processes, one with rate α, which represents evidence for R_a, and one with rate β, which represents evidence for R_b. The probability that R_a is made during a small interval $(t, t + h)$ is the product of the probability that the last of a sequence of K_a observations favoring R_a arrives during the interval $(t, t + h)$ and the probability that the time taken to accumulate K_b observations favoring R_b is greater than $t + h$. Because the intervals between counts of either kind are distributed exponentially, the probability density associated with response R_a at time t is a K_a-stage gamma density with rate α. Similarly, the probability that K_b observations favoring R_b have not accrued by t is the survivor function of a K_b-stage gamma distribution with rate β. The same considerations apply to response R_b made at time t, with the roles of the two counters reversed. The first passage time density functions may thus be written

$$g_a(t) = \frac{(\alpha t)^{K_a - 1} \alpha e^{-\alpha t}}{(K_a - 1)!} \left\{ \sum_{j=0}^{K_b - 1} \frac{(\beta t)^j}{j!} e^{-\beta t} \right\} \quad \text{(A10a)}$$

and

$$g_b(t) = \frac{(\beta t)^{K_b - 1} \beta e^{-\beta t}}{(K_b - 1)!} \left\{ \sum_{j=0}^{K_a - 1} \frac{(\alpha t)^j}{j!} e^{-\alpha t} \right\}. \quad \text{(A10b)}$$

By an argument based on the superposition property of Poisson processes—that is, a pair of independent Poisson processes with rates α and β behave like a single Poisson process with rate $\alpha + \beta$—it may be shown that the probability that a given observation in the evidence stream favors R_a is $\alpha/(\alpha + \beta)$; the probability that it favors R_b is $\beta/(\alpha + \beta)$ (Townsend & Ashby, 1983). The probability that the final response is R_a is therefore just the probability that K_a observations favoring R_a are accrued before K_b observations favoring R_b. Any sequence consisting of between K_a and $K_a + K_b - 1$ observations, K_a of which favor R_a and j of which favor R_b, where $0 \leq j \leq K_b - 1$, will result in an R_a response. The probability of this response, $P(a)$, is just the sum of the negative binomial probabilities of such sequences. The same analysis applies to $P(b)$ with the roles of the two counters reversed:

$$P(a) = \sum_{j=0}^{K_b - 1} \binom{K_a - j - 1}{j} \left(\frac{\beta}{\alpha + \beta} \right)^j \left(\frac{\alpha}{\alpha + \beta} \right)^{K_a} \quad \text{(A11a)}$$

and

$$P(b) = \sum_{j=0}^{K_a - 1} \binom{K_b - j - 1}{j} \left(\frac{\alpha}{\alpha + \beta} \right)^j \left(\frac{\beta}{\alpha + \beta} \right)^{K_b}. \quad \text{(A11b)}$$

The index j in these equations runs over all possible numbers of observations in the nonresponse counter.

Neurally Inspired Models

These models assume that evidence for competing responses is accumulated in parallel, as occurs in counter and accumulator models, but that the evidence is continuously distributed and is accumulated in continuous time, as occurs in diffusion models. The three models were the leaky competing accumulator model of Usher and McClelland (2001), the leaky accumulator model of Smith (2000), and a leaky accumulator model with relative criterion. For each of these models, the growth of evidence as a function of time is described by a pair of stochastic, dynamic equations. These equations describe the growth of evidence for the two responses as

a function jointly of the stimulus and of the evidence already obtained. In Usher and McClelland's (2001) notation, the two evidence totals are denoted x_1 and x_2. The growth of evidence as a function of time is described by the following stochastic equations:

$$dx_1 = (\rho_1 - kx_1 - \beta x_2) \frac{dt}{\tau} + \xi_1 \sqrt{\frac{dt}{\tau}} \quad \text{(A12a)}$$

and

$$dx_2 = (\rho_2 - kx_2 - \beta x_1) \frac{dt}{\tau} + \xi_2 \sqrt{\frac{dt}{\tau}}. \quad \text{(A12b)}$$

These equations give dx_i, the change in the amount of evidence in counter i, $i = 1, 2$, during a small time step dt. This change is the sum of three terms: the information derived from the stimulus, ρ_i; a decay term, $-kx_i$, which is proportional to the evidence in the counter; and an inhibition term, $-\beta x_i$, which is proportional to the evidence in the other counter. Decay in these models operates in the same way as in the OU model: The more evidence that has accumulated, the greater the decay. The inhibition term induces competition between counters such that the more evidence there is in one counter, the more the accrual rate in the other counter is reduced. Moment-by-moment variability in the accrual rate comes from the quantities ξ_1 and ξ_2, which are independent, normally distributed random variables with a mean of zero and standard deviation of 1.0. The parameter τ functions like the inspection-time parameter, λ, in the accumulator model to fix the time scale of the model. Like the accumulator and Poisson counter models, evidence is accrued in the two counters in parallel until one or other counter exceeds its criterion, that is, until $x_1 \geq a_1$ or $x_2 \geq a_2$, and the associated response is then made.

In the leaky competing accumulator model, the accrual rates in the two counters depend on the combined effects of the stimulus, decay, and inhibition. In the leaky accumulator model of Smith (2000), there is no inhibition between counters (i.e., β_1 and β_2 in Equations A12a and A12b are both zero) and the accrual rates in the two counters depend on the stimulus and decay only. The leaky accumulator with relative criterion is identical, except it uses a relative rather than an absolute stopping rule. In this model, response R_a is made if

$$x_1 - x_2 \geq a_1;$$

response R_b is made if

$$x_2 - x_1 \geq a_2.$$

As in the other models, RT is the smallest value of t for which the stopping condition is satisfied. Following Usher and McClelland (2001), we assumed the accrual rates in the two counters were constrained so that $\rho_1 + \rho_2 = 1$. Like them, we used Monte Carlo methods to derive predictions for these models.

Distributions of Parameters

The parameters of the models investigated here were assumed to be subject to trial-by-trial variation. For all of the models, three sources of independent variability were assumed: variation in the quality of the information contained in the stimulus, variation in the amount of information required for a response, and variation in the nondecisional component of processing. For the Wiener and OU diffusion processes, variation in information quality was represented by variation in the drift parameter, ξ; variation in the amount of information needed for a response was represented by variation in starting point, z. Let $\bar{g}_i(t)$, $i = a, b$ denote the predicted first passage time distributions when variation in parameters is assumed. These functions are related to the original first passage time functions (i.e., Equations A2a and A2b for the Wiener model; Equations A5a and A5b for the OU model) in the following way:

$$\bar{g}_i(t; z, v) = \int_{T_{er}-s_t/2}^{T_{er}+s_t/2} \int_{-\infty}^{\infty} \int_{z-s_z/2}^{z+s_z/2} g_i(t-\tau; \zeta, \xi)$$

$$\times u_z(\zeta)N(\xi; v, \eta)u_t(\tau) \, d\zeta \, d\xi \, d\tau. \quad (A13)$$

In this equation, $N(\xi; v, \eta)$ is a normal density of drift with mean v and standard deviation η; $u_z(\zeta) = 1/s_z$, $z - s_z/2 \leq \zeta \leq z + s_z/2$, is a uniform density of starting points with mean z and range s_z; and $u_t(\tau) = 1/s_t$, $T_{er} - s_t/2 \leq \tau \leq T_{er} + s_t/2$, is a uniform distribution of nondecisional times with mean T_{er} and range s_t. The triple integral in Equation A13 was evaluated numerically, by approximating the continuous distributions of parameters with discrete distributions and by evaluating the integrals as sums.

The first passage time probabilities for the accumulator model with variation in parameters are related to those in Equations A9a and A9b by an equation similar to Equation A13, but with an additional inspection time parameter, λ, that maps the model from discrete to continuous time. As in Equation A13, let $\bar{g}_i(t)$, $i = a, b$, denote the first passage time density function for response i at time t. These functions are obtained from the mass functions in Equations A9a and A9b by integrating over parameters in a similar way:

$$\bar{g}_i(t; k_a + \kappa, k_b + \kappa, \mu) = \int_{T_{er}-s_t/2}^{T_{er}+s_t/2} \int_{-\infty}^{\infty} \int_{0}^{\infty} \sum_{n=1}^{\infty} \delta(t - n\lambda - \tau)$$

$$\times P_i(n; k_a + k, k_b + k, \xi)w_K(k; \kappa)N(\xi; \mu, \sigma)u_t(\tau) \, dk \, d\xi \, d\tau. \quad (A14)$$

In this equation, $N(\xi; \mu, \sigma)$ is a normal distribution of increment distribution means with mean μ and standard deviation σ, which describes trial-by-trial variation in the average information content of the stimulus. The sum of Dirac delta functions, $\sum \delta(t - n\lambda)$, represents a sequence of unit masses concentrated at $t = n\lambda$, $n = 1, 2, \ldots$, which maps the model to continuous time. The function $u_t(\tau)$ is a uniform distribution of nondecisional times, which is defined in the same way as for the diffusion models.

The function $w_K(k; \kappa)$ in Equation A14 is the probability density function for the distribution of response criteria. A number of candidate distributions were considered for this role, including normal, uniform, and Weibull distributions. The distribution function for the latter is

$$W_K(k) = 1 - \exp\left[-\left(\frac{k - k_i}{\kappa}\right)^{\alpha}\right],$$

$k \geq k_i$, $i = a, b$. In this equation, k_i is an offset parameter, κ is a scale parameter, and α is a shape parameter. Variation in α yields a wide variety of distributional shapes, including positively skewed, symmetric, and negatively skewed distributions. When α was free to vary, the best fits were obtained with a value of α around 1.0. This special case of the Weibull is an offset exponential distribution with offsets k_a and k_b and means $k_a + 1/\kappa$ and $k_b + 1/\kappa$, respectively. As for Equation A13, the triple integral in Equation A14 was evaluated by summing over values of the approximating discrete distributions.

The analogous expression for the Poisson counter model is

$$g_i(t; k_a + \kappa, k_b + \kappa, \alpha + \beta, \rho) = \int_{T_{er}-s_t/2}^{T_{er}+s_t/2} \int_{0}^{1} \sum_{k=0}^{\infty} g_i[t - \tau; k_a + k,$$

$$k_b + k, \pi(\alpha + \beta), (1 - \pi)(\alpha + \beta)]h(\pi; \rho)w_K(k; \kappa)u_t(\tau)d\pi d\tau. \quad (A15)$$

In this equation, the function $h(\pi; \rho)$ describes the distribution of relative accrual rates in the two counters, and $u_t(\tau)$ is a uniform distribution of nondecisional times that was parameterized in the same way as for the other models.

To model between-trials variation in accrual rates, the sum of the Poisson rates in the two counters, $\alpha + \beta$, was held constant while the relative accrual rate, $\pi = \alpha/(\alpha + \beta)$, varied. The rate on any trial was treated as a random sample from a beta distribution $h(\pi, \rho)$, with mean ρ. This distribution provides a general model for a binomial success probability π, $0 < \pi < 1$, when the success probability is a random variable. The probability density function for the beta distribution is

$$h(\pi) = \frac{\pi^{u-1}(1 - \pi)^{v-1}}{B(u, v)},$$

where $B(u, v)$ is the beta function and u and v are parameters that jointly determine the mean, variance, and shape of the distribution (Johnson & Kotz, 1970). The mean and variance are

$$\rho = \frac{u}{u + v}$$

and

$$\text{var}(\pi) = \frac{uv}{(u + v)^2(u + v + 1)},$$

respectively. By varying the shape parameters u and v, the beta distribution can exhibit a wide variety of shapes including uniform, symmetric, and skewed distributions. For example, with u and v equal to 2 and 3, the beta distribution produces skewed distributions of rates that are similar to those proposed by Van Zandt (2000) to model recognition memory (cf. Johnson & Kotz, 1970, pp. 42–43). For the fits reported here, the shape of the beta distribution, $u + v$, was assumed to be constant across stimulus conditions, while the mean, ρ, was free to vary. To model biases in the accrual rates in Experiment 3, a bias parameter, r, was introduced, and the mean rate in each condition was set to

$$\rho = \frac{u - r}{u + v}.$$

In this equation, the ratio $u/(u + v)$ is a within-block parameter that depends on the stimulus, whereas r is a between-blocks parameter that depends on the relative frequencies of old and new items.

The function $w_K(k; \kappa)$ in Equation A15 is a discrete distribution of criterion values. Rectangular and offset geometric distributions of criteria were both considered as candidate distributions, with the latter providing a better account of the data. The mass function for a geometric distribution with mean κ is

$$w_K(k; \kappa) = (1/\kappa)(1 - 1/\kappa)^{(k-1)},$$

where $1/\kappa$, $0 < 1/\kappa < 1$, is the Bernoulli success probability. Equation A15 was again evaluated by summing over approximating discrete distributions.

For the neurally inspired models, variability in nondecisional times was included in the same way as for the other models. In addition, variability in starting point was introduced to allow the models to predict fast errors (cf. Usher & McClelland, 2001, p. 570).