Psychological Review
2001, Vol. 108, No. 3, 550–592

0033-295X/01/$5.00 DOI: 10.1037//0033-295X.108.3.550

The Time Course of Perceptual Choice: The Leaky, Competing Accumulator Model

Marius Usher
Birkbeck College, University of London

James L. McClelland
Carnegie Mellon University and the Center for the Neural Basis of Cognition

The time course of perceptual choice is discussed in a model of gradual, leaky, stochastic, and competitive information accumulation in nonlinear decision units. Special cases of the model match a classical diffusion process, but leakage and competition work together to address several challenges to existing diffusion, random walk, and accumulator models. The model accounts for data from choice tasks using both time-controlled (e.g., response signal) and standard reaction time paradigms and its adequacy compares favorably with other approaches. A new paradigm that controls the time of arrival of information supporting different choice alternatives provides further support. The model captures choice behavior regardless of the number of alternatives, accounting for the log-linear relation between reaction time and number of alternatives (Hick's law) and explains a complex pattern of visual and contextual priming in visual word identification.

When input is presented to the senses, the need often arises to determine its identity or to make some other judgment about it. In experimental paradigms, the time course of this judgment process is often studied in experiments requiring a choice among two or more response alternatives. Choice responses in these tasks tend to take more time when the judgment is difficult, and they are quite variable, both in their outcome and in their duration. Progress in understanding these findings has come through the development and testing of models that capture two crucial principles of information processing: First, they treat information processing as a gradual process, based on the accumulation of information over time. Second, they treat the process as stochastic or intrinsically variable, so that the information accumulated within each small time interval is subject to random fluctuations. Many variants of such models have been explored (e.g., Ashby, 1983; Audley & Pike, 1965; Laming, 1968; Link & Heath, 1975; Ratcliff, 1978, 1988; Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999; Stone, 1960; Vickers, 1970, 1979).

Marius Usher, Department of Psychology, Birkbeck College, University of London, London, England; James L. McClelland, Department of Psychology, Carnegie Mellon University, and the Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania.

The research reported here was supported by National Institute of Mental Health Grants MH00385 and MH47566. Special acknowledgments are due to Douglas Vickers for providing the data from his experiments. We also thank Mark Chappell, David Plaut, and Tony Marley for comments on a draft of this article, Zeev Olami for discussions, and E. Davelaar for running Experiment 2.

Correspondence concerning this article should be addressed to Marius Usher, Department of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England, or to James L. McClelland, Center for the Neural Basis of Cognition, 115 Mellon Institute, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent to m.usher@bbk.ac.uk or mcclelland+@cmu.edu.

Models of this class have had considerable success in addressing a wide range of experimental findings on the time course and outcome of perceptual choice experiments. However, we argue that there are additional principles that should also be incorporated: First, information accumulation is subject to leakage or decay, and second, representations of the alternative outcomes of the decision process compete with each other, through a process of lateral inhibition. We develop a new model called the *leaky, competing accumulator model* as an extension of the classical models in the field. Two additional principles, recurrent excitation and nonlinearity, also are considered. We propose specific formulations of the principles that allow their incorporation into stochastic information accumulation models, and we explore their utility in accounting for a broad range of phenomena on the time course of information processing in perceptual choice.

Classical Stochastic Information Accumulation Models

To set the stage for our exploration of these matters, we begin with a brief discussion of key aspects of the literature on existing stochastic information accumulation models. This discussion focuses on the reasons why it is important to elaborate the framework to take the additional principles of leakage and competition into account.

In early work, two types of models were considered: the accumulator or counter models (Audley & Pike, 1965; LaBerge, 1962; Vickers, 1970, 1979; Vickers, Caudrey, & Wilson, 1971) and the random walk models (Ashby, 1983; Laming, 1968; Link & Heath, 1975; Stone, 1960). An introduction to the basic properties of these models is available in Townsend and Ashby (1983). Both kinds of models are typically applied to experiments consisting of trials in which a stimulus is presented and the participant must decide which of two responses to assign to it. For example, the stimulus might be a colored light, and the participant may be asked to decide if it is red or green. All these models span a range of

variants, in which information is accumulated over time. Increments may be binary, multivalued, or continuous, and sampling may be assumed to occur at discrete time steps or continuously. In one variant of the accumulator model (Audley & Pike, 1965), a series of discrete, binary information samples is assumed. The process is equivalent to drawing a series of balls, which might be thought of as either red or green, from an urn, with replacement, so that the probability of drawing a red or green ball remains fixed during sampling. The actual stimulus would be viewed on this analogy as setting the proportion of red or green balls in the urn. A red stimulus would establish a high proportion of red balls, whereas a green stimulus would establish a high proportion of green balls. Before processing, accumulators for balls of each color would be initialized, and at each time step, a ball would be drawn, and a count added to the appropriate accumulator. In applications to the standard reaction time (RT) paradigm, a response criterion is assumed, so that a response is produced when the count associated with one of the accumulators reaches the criterion value.

The classical random walk and the related diffusion model are similar to the accumulator model. Particular models differ in terms of whether the sampled values are discrete or continuous and whether sampling is thought of as occurring at discrete time points or continuously. The crucial distinction between these models and the accumulator models, for our purposes, is that in all random-walk models, a single cumulative variable is used. To illustrate, we can consider a classical random walk, where this variable can be thought of as the difference between the number of red or green balls that have been drawn from an urn. Before processing, the single counter is initialized, and at each step, a count is added if the ball is red or subtracted if the ball is green. In other variants of this class of models, the counter variable is incremented by a variable amount, for example, it might be a sample from a normal distribution with mean μ and standard deviation σ. When such a process is thought to proceed in continuous time (rather than in discrete jumps), it is called a *diffusion* process. In the version that we consider, here called the *classical diffusion model,* the stimulus condition is thought to set the value of μ, and σ is left invariant across conditions. Thus, for our case, a red light might be associated with $\mu_r > 0$, and a green light might be associated with $\mu_g < 0$. Responses are generated using two decision boundaries. When the value of the diffusing variable reaches the positive boundary, the red response is initiated, and when it reaches the negative boundary, the green response is initiated.

In earlier research, these classical models (and other variants) were typically applied to data from what we call the *standard RT* paradigm, in which participants are instructed to respond as rapidly and accurately as possible, to indicate which of two stimulus alternatives has been presented. We consider the RT paradigm in detail later. For the present, we consider a difficulty that all these models share in another paradigm. In this paradigm, called the *time-controlled* paradigm, an attempt is made to track the time course of information accumulation by asking participants to respond at different specified times after stimulus presentation on different trials. In one common variant of the time-controlled paradigm, a response signal is presented at some point in time after stimulus onset on each trial, and the participant is required to make a response choice within a very brief interval thereafter (Corbett & Wickelgren, 1978; McElree & Dosher, 1989; Ratcliff & McKoon, 1982; Reed, 1973, 1976; Wickelgren, 1977; Wickelgren & Corbett, 1977). In this case, the natural assumption is that information continues to accumulate until the response signal is detected, at which point a decision is made based on the state of the accumulators (Ratcliff, 1978; but see also Ratcliff, 1988, for a more complex process). For simplicity, we consider the case in which there are no biases in the decision process. For accumulator models, the two counters are both initialized at 0, and the response associated with the largest count is chosen. For the random walk and diffusion models, the single countervalue is initialized to 0, and one response is chosen for all values greater than 0, with the other response chosen otherwise.

Constraints on Asymptotic Accuracy: Drift Variance and Leakage

The difficulty that the classical models face in this situation is that the information accumulation process proceeds indefinitely without loss or decay of information. The effect of this is that as time from stimulus onset to the response increases, discriminability, as measured by d', always increases without bound unless the stimuli are exactly identical. Specifically, as N (the number of samples accumulated) increases, d' continues to increase following a square-root function.[1] In all cases, accuracy increases without bound, even when p and q are very close to .5 (or, for the continuous case, μ_r and μ_g are very close to 0.0). As long as there is any difference among the stimuli, accuracy will be perfect if enough time is allowed for information accumulation.

Perfect accuracy with unlimited processing time may be a useful approximation in some cases, but there are clearly experiments in which accuracy asymptotes at a finite level. This typically occurs when stimuli that are difficult to discriminate are used without any time pressure. Swensson (1972) noted that "an accuracy ceiling is assumed implicitly in any experimental procedure that uses error rate as its sole measure of discrimination performance, without attempting either to make S respond quickly or to limit the time for which the stimulus is available to him" (p. 28). Indeed, Swensson explicitly demonstrated the accuracy ceiling in one of his experiments. He presented one participant with tilted rectangles with sides differing by less than 2% in length, with the task of indicating which pair of sides was longer. Payoffs were adjusted across sessions to manipulate RT. Accuracy increased as RT increased, but only up to a point; for RTs above 500 ms, accuracy remained fixed at about 97% correct. Such a pattern is inconsistent with the information accumulation assumptions of all the classical models. Swensson (1972) suggested that "the presence of an accuracy ceiling could reflect a limit on the number of useful observations

[1] Consider the random-walk model, with a red stimulus presented, so that $p > .5$ of the balls in the urn are red and $q = 1 - p$ are green. The expected value of the number of red and green balls drawn as a function of the number of samples N is pN and qN, respectively; thus, the expected value of the difference is $(p - q)N$. At the same time, the standard deivation of the difference increases according to a different function, proportional to the square root of N. Discriminability, as measured by d', is the expected value of the difference divided by the standard deviation of the difference, which is proportional to $N/\sqrt{N}$, or just $\sqrt{N}$. Similar arguments apply to the accumulator model and the classical diffusion model.

S can sample. If Ss observations were truncated, there would be a limit on the expected amount of evidence he could obtain before some 'critical duration' elapsed" (p. 30).

Ratcliff (1978) introduced a variant of the diffusion model that overcame this shortcoming of the classical models, but in a way that differed from Swensson's (1972) proposal. Working within the framework of the classical diffusion model, Ratcliff (1978) introduced the assumption that there may be variability in μ, the mean direction of drift, from trial to trial within the same experimental condition. This introduces a new parameter, the standard deviation of the drift, here given by σ_d. In this case, the asymptotic discriminability d' of stimuli presented in two different conditions converges to a finite value equal to $(\mu_1 - \mu_2)/\sigma_d$, where μ_1 and μ_2 are the mean directions of drift for the two different conditions. We call the diffusion model that incorporates this assumption the *diffusion-with-drift-variance* (DDV) model to distinguish it from from the classical diffusion model without drift variance.

The DDV model has considerable appeal because it directly links the outcome of processing given unlimited time to the standard picture of stimulus discrimination presented in signal-detection theory. Using the drift variance assumption, Ratcliff (1978) has been able to account quite well for a wide range of data from recognition-memory experiments obtained both with the standard RT and the time-controlled paradigms. More recently, Ratcliff and Rouder (1998) and Ratcliff et al. (1999) have applied the same model to perceptual classification experiments and have shown that the model can account quite well for individual-participant data in experiments involving the perceptual classification of ambiguous stimuli.

In the DDV model, the limitation on discriminability arises from trial-to-trial variation in the input to the information accumulation process. While not denying that such a factor is likely to play a role in many experiments (especially the memory experiments that were the main focus of the Ratcliff, 1978, analysis), we suggest that in addition, limitations on discriminability may also arise from characteristics of the information accumulation process itself. Specifically, we suggest that information that has been accumulated may be subject to leakage or decay. In such a case, we shall see that the information accumulated reaches an asymptotic level reflecting a balance between the accumulation of new information and the loss of information already accumulated. This proposal is a graded version of the suggestion made by Swensson (1972): Leakage provides a soft or graded truncation of the information accumulation process. At some point, the gains from further sampling are matched by losses of what has already been sampled, so that further information accumulation is of no benefit.

Leaky integration has previously been proposed by several investigators, including Grice (1972) and McClelland (1979). Smith (1995) has incorporated leakage in stochastic information accumulation models applied to luminance increment detection; Busemeyer and Townsend (1993) have used this assumption in their model of decision making (see also Diederich, 1997); Diederich (1995) has used it in a model of intersensory RT; and Kim and Myung (1995) have used it in a model of semantic priming. McClelland (1993) has proposed leaky stochastic information integration as part of a general framework for modeling cognitive processes. In later sections, we compare the shapes of time–accuracy curves produced by incorporating drift variance with those that arise from the effects of leakage. We see that leakage and drift variance make different predictions for the shapes of empirical time–accuracy curves, with leakage leading to curves that more closely approximate the shapes of empirical time–accuracy curves.

Extending the Framework Beyond Two Alternatives

Another issue for classical information accumulation models concerns how to extend them to choices among more than two alternatives. We consider this issue as it arises in the standard RT paradigm. In the two-alternative situation for which the models were developed, the diffusion and random walk models assume that the decision to respond is based on a single decision variable, which represents the relative evidence for one alternative compared with another. For the accumulator model, the response is made whenever either accumulator reaches its own independent criterion, regardless of the state of the other accumulator. Here a quandary arises. Although the use of an absolute criterion for each alternative generalizes directly from two to any number of alternatives, it is not immediately apparent how to generalize the relative evidence criterion. Yet, the use of a relative evidence criterion has several advantages. First, the use of relative evidence has been shown to be optimal, in the sense of being the procedure that produces the fastest possible RTs for a given desired level of accuracy (Wald & Wolfowitz, 1948). Furthermore, current diffusion models (Ratcliff & Rouder, 1998; Ratcliff et al., 1999) do an excellent job of accounting for an impressive range of aspects of RT data from two-choice experiments. Thus it would be desirable to retain the relative-evidence characteristic of the random walk–diffusion approach in extensions to situations involving more than two alternatives.

A number of ways of extending the relative-evidence characteristic of the random-walk and diffusion models have been explored. The problem we face in this case is really an excess of possibilities, where it is difficult to know in advance which approach to choose. Here, the study of neural information processing mechanisms suggests an approach based on the use of lateral inhibition. In our use of this mechanism, there is a separate evidence accumulator for each alternative. Inhibition among the accumulators allows the evidence accrued at each to influence the state of every other accumulator. As input builds up in one accumulator, it sends inhibition to all of the others. The accumulator with the strongest input tends to dominate the others, so that in the end the states of the accumulators tend to reflect the relative amount of information accumulated in comparison with the other accumulators. Given that relative evidence is already reflected in the state of activation of each accumulator, an absolute criterion can be used as in the classical accumulator models. Thus, the use of lateral inhibition allows the construction of models that emulate the relative evidence characteristic of the random walk and diffusion models while using the absolute criterion of the accumulator models, so that the model can be applied equivalently to experimental situations involving two or more than two alternatives.

One alternative to lateral inhibition that may at first seem similar is *bottom-up* or *feed-forward* inhibition. A stochastic information accumulation model consistent with this approach has in fact been proposed by Heuer (1987). On this scheme, whenever an input provides positive (excitatory) input to the accumulator for one alternative, it simultaneously provides negative (inhibitory) input to the other alternative(s). However, bottom-up inhibition can

become problematic as the number of alternatives is varied, and this makes it less attractive than lateral inhibition as a general mechanism for use with any number of alternatives. To see the problem, imagine first a situation in which some evidence that favors one alternative has an excitatory effect on that alternative and an equal inhibitory effect on all other alternatives. In this case, the alternative with the strongest excitation will also receive the weakest inhibition; thus, on balance, it will receive an excitatory input. But when there are $N > 2$ alternatives and several are partially supported, each one will receive inhibition from $N - 1$ sources of evidence and excitation from only a single source. The input to all units can be negative in this case, preventing any alternatives from achieving a positive level of activation. One could compensate for this effect by reducing the amount of inhibition as the number of choice alternatives increases. Although he did not consider more than two alternatives, Heuer's model does assume that task variables can modulate the strength of the negative coupling. However, this will mean that the amount of inhibition that can be produced by evidence for just one alternative will shrink as the number of choice alternatives grows. The problem can be avoided by using lateral instead of feed-forward inhibition. In this approach, bottom-up input to accumulators is only excitatory. Many alternatives will receive some excitation, and the one that is the most consistent with the input will receive the most excitation. These alternatives all inhibit each other, but the one with the largest excitatory input will tend eventually to dominate, even if there are many partially activated competitors. Effects of this type have been simulated in many models, including the interactive activation model (McClelland & Rumelhart, 1981). The mechanism automatically adjusts itself to the number of alternatives, so that the alternative with the strongest support tends eventually to dominate the competition.

Thus, lateral inhibition appears to be a powerful mechanism for selection. We are not, by any means, the first to emphasize this (see, e.g., Grossberg, 1976, 1978). The issues addressed by lateral inhibition are very general and apply whenever the occasion arises to select the alternative that is most consistent with one or more sources of input. Such occasions arise continually, we believe, whenever one is processing any kind of natural stimuli (such as spoken or handwritten input) in the presence of any form of degradation (e.g., line noise in telephone circuits or faintness of signals due to distance), and whenever one is attempting to exploit linguistic context, which generally supports multiple possibilities.

Additional Principles of Information Accumulation

The considerations we have reviewed suggest two useful ways in which the classical models might be extended. First, incorporation of leakage may allow these models to capture the time course of information accumulation more accurately. Second, incorporation of lateral inhibition may allow them to capture the desirable properties of the relative-evidence decision criterion in a way that extends naturally to more than two alternatives. These considerations nicely dovetail with neurophysiological evidence that is considered below, suggesting that such mechanisms are indeed at work in the neural machinery underlying performance in information-processing tasks.

Our model also explores two more principles. One of these is recurrent self-excitation. The motivation for including this principle comes not so much from the RT and time–accuracy studies that are the primary target for our present analysis, but from other domains of psychological and computational investigation. The use of recurrent excitation allows models to exhibit a tendency to settle to a fixed point or stable equilibrium. It has been suggested that this tendency may underlie a large number of phenomena in perception, language processing, and working or active memory (Anderson, Silverstein, Ritz, & Jones, 1977; Cohen & Servan-Schreiber, 1992; Grossberg, 1976, 1978; G. Hinton & Shallice, 1991; McClelland & Elman, 1986; McClelland & Rumelhart, 1981; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Usher & Cohen, 1999).[2] Given the utility of the assumption in a wide range of models and the neurophysiological evidence we consider below, it seems useful to consider what role it may play in models of the time course of information processing.

The final principle we consider is nonlinearity, and here the motivation for inclusion of this principle rests on general computational considerations. As is well known, many computations thought to be essential for perception, cognition, and action cannot be carried out without at least one layer of nonlinear computation (see, e.g., Rumelhart, Hinton, & McClelland, 1986, for discussion). Negation, detection of symmetry and disparity, and other computations that depend crucially on conjunctions of elements are among those requiring the use of some form of nonlinearity. Therefore, it is important that a type of nonlinearity that appears to characterize neural information processing can be incorporated into our model without affecting its ability to account for any of the data that we consider in this article.

Linking Psychological and Neural Levels of Analysis

The present work grows in part from a long-standing effort to understand perceptual and cognitive processes within the *parallel-distributed processing* framework, where processing arises from the interactions of ensembles of simple, neuronlike processing units (Rumelhart, McClelland, & the PDP Research Group, 1986). Much of this work has focused on understanding phenomena at the psychological level. Principles of neural information processing have been exploited where useful, but often a principle or assumption whose motivation is fundamentally computational is introduced without regard to its possible biological implementation. Thus, the introduction of the back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986) has fostered the development of a powerful computational framework. Although back-propagation has been helpful in the development of models that can capture psychological data on a number of different aspects of cognitive and perceptual processing (Cohen,

[2] The idea that recurrent excitatory influences operate in perception has not gone uncontested, and Massaro (1989) and Massaro and Cohen (1991) have pointed out that the original interactive activation model failed to capture a pattern of data seen frequently in experiments manipulating both context and stimulus information in perception and have claimed that this failure indicated that information processing is a strictly feed-forward process. However, McClelland (1991) and Movellan and McClelland (2001) have shown that with intrinsic variability incorporated, the interactive activation model accounts for the pattern of data in question, demonstrating that Massaro's claim was incorrect. Given this finding, we know of no empirical evidence against the possibility that recurrence plays a role in perception.

Dunbar, & McClelland, 1990; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), there have been some drawbacks to this approach. First, it has tended to distance the work from insights from neuroscience that might inform and constrain it (Grossberg, 1987; McClelland, 2001). Second, the models have generally been underconstrained (Massaro, 1988; McCloskey, 1991). The range of different assumptions that have been considered in this work is extensive, and the decision about what principles should be considered and about their exact implementation has sometimes seemed arbitrary. In addition to these limitations, the relationship between models developed in this framework and classical models of information processing has not always been explicitly considered.

The present work attempts to address these limitations. First, the work draws extensively on developments in experimental and computational neurophysiology, where the goal is to understand the functional properties of actual neural mechanisms (Abbott, 1991; Ahmed, Anderson, Douglas, Martin, & Nelson, 1995; Ahmed, Anderson, Douglas, Martin, & Whitteridge, 1998; Amit, 1989; Amit, Brunel, & Tsodyks, 1994; Amit & Tsodyks, 1991; Chelazzi, Miller, Duncan, & Desimone, 1993; Douglas, Koch, Mahowald, Martin, & Suarez, 1995; Ermentrout, 1994; Hertz, Krogh, & Palmer, 1991; Hess, Negishi, & Creutzfeld, 1975; Usher & Niebur, 1996). Second, the principles included in the modeling effort have neurobiological as well as computational or psychological motivation, and the specific instantiations of the principles are informed by additional neurophysiological observations. Third, we direct our efforts, not toward multilayer network models focused on specific tasks, as in much of the parallel-distributed processing work, but toward a simplified but very general model that can be seen as a continuation of the tradition of stochastic information accumulation models. How the principles explored here would play out in more complex multilayer architectures is left for subsequent investigations.

Note that the implications of findings from neurophysiology for models of cognition are not transparent. There are many aspects of the physiology that we do not incorporate that may turn out to be relevant, and there are many ways in which the aspects that we do incorporate might be rendered irrelevant to understanding perceptual processes at an information-processing level. In this light, our exploration of some possible implications of neurophysiology for models of cognition must be seen as a contribution to an ongoing exploration and not as a definitive statement of how cognitive processes arise from the neural substrate.

Next, we review work on neural information processing that has informed the development of our model. This review both supports the general principles on which the model is based and provides further constraints that have been used in determining the specific quantitative formulation.

Unit Activations and Cell Populations

When considering relations between cognitive–perceptual processes and the underlying neurophysiology, it is essential to relate the two at an appropriate level of description of neural processes. As already indicated, the choice of level is itself a scientific question. We adopt the dominant approach taken in computational neuroscience today, which follows the Hebbian perspective (Hebb, 1949). In this approach, each cognitive unit is represented by a pattern of activation over a group of neurons, or *cell population,* and the activation of the cognitive unit is represented by the mean firing rate of the neurons in the population (Abbott, 1991; Amit, 1989; Amit et al., 1994; Georgopoulos, Kettner, & Schwartz, 1986; Wilson & Cowan, 1972). This approach is consistent with the idea that a representation may be distributed over many neurons at the physiological level, even while it is treated as a single variable for modeling the time course of information processing in choice-response situations (for discussion, see Smolensky, 1986).

At a physiological level, it seems likely that individual neurons participate in the representation of particular inputs to varying degrees. Neurons generally appear to have fairly broad tuning curves, so that when a particular stimulus is shown, many neurons will be activated to some extent (Georgopoulos et al., 1986; Miyashita, 1988; Tanaka & Saito, 1991). Furthermore, it seems likely that different response alternatives may in some cases be represented by potentially overlapping populations. Our present model allows overlap in the patterns of activation produced by different inputs but treats the representation of distinct response alternatives as nonoverlapping. This seems a useful first approximation for most two- or even *N*-alternative-choice RT tasks.

Input Currents and Firing Rates

In modeling neurons, it is often useful to distinguish between two local variables: the local input current and the neuron's firing rate. The firing rate corresponds to the probability that the individual neuron will or will not fire (emit an action potential) in any given small increment of time, whereas the input current is the total input to the neuron, which determines what the firing rate will be. When averaged across a cell population, those quantities are called the *population input current* and the *population firing rate.*

At the single-neuron level, the input current captures the net effect of the external input, including input arising from lateral connections from other neurons. The neural input current and the firing rate are related by a nonlinear activation function. One such function that is often used in cognitive models is the logistic function. However, response functions in cortical and motor neurons typically approximate a threshold-linear input–output relation in the range of firing rates observed physiologically (Jagadeesh, Gray, & Ferster, 1992; see also Ahmed et al., 1995, 1998; Granit, Kernell, & Shortess, 1963; Mason & Larkman, 1990). Thus, the threshold–linear function appears to provide a good approximation over the range of firing rates likely to occur in the relevant neural populations during tasks such as stimulus identification in choice-RT experiments. The threshold–linear function is also quite tractable mathematically. For these reasons, we have adopted it for use in our model.

The threshold-linear function has similarities to the brain-state-in-a-box (BSB) activation function used by Anderson et al. (1977). Like the BSB function, the threshold–linear function is linear over part of the domain of its input, and under certain conditions, some aspects of its behavior can be captured with a simple linear approximation. Even so, the imposition of a threshold makes it sufficient to perform nonlinear computations such as the XOR problem. Note that an upper bound on activation is sometimes needed to prevent runaway activation in recurrent networks. This is provided explicitly in the BSB model. Such an explicit upper

bound is not necessary when recurrent activation is overmatched by leakage, as we assume for the problems we consider here. There are circumstances, however, in which some form of upper bound may be necessary, and many models do incorporate them (Grossberg, 1976; McClelland & Rumelhart, 1981; Usher & Cohen, 1999). Thus, we assume that such an upper bound is present, but it is not necessary to incorporate it into our models to understand the phenomena of interest here.

Intrinsic Variability

Both human performance and single neurons show highly variable (probabilistic) behavior even under identical external stimulation. In the neuroscience literature, the issue of variability in the discharge patterns of nerve cells has recently been the focus of intensive research, both at the empirical level (Softky & Koch, 1993) and at the theoretical level (Shadlen & Newsome, 1994; Usher, Stemmler, Koch, & Olami, 1994; Usher, Stemmler, & Olami, 1995). As emphasized by Shadlen and Newsome, the variability in single-cell discharge suggests that the code used by nerve cells is a graded rate code. At the cell population level, the rate can be defined as the number of spikes emitted by all members of the population within a short time period. Accordingly, one can think of the fluctuations in neural discharge as a source of noise or variability in this population-rate variable. The central limit theorem shows that these fluctuations in the population activity are well characterized by Gaussian distributions. Correspondingly, the fluctuations in the input to a neuron can be thought of as the sum of the Gaussian fluctuations from all of the different populations that project to it. For this reason, we model intrinsic variability by adding a random sample of zero-mean Gaussian noise into the input to the decision units in our model. This aspect of our model is consistent with standard practice in many existing stochastic information accumulation models.

Leakage and Recurrent Self-Excitation

An important characteristic of the neural input current is its passive exponential decay (Abbott, 1991; Amit & Tsodyks, 1991). This decay results in the leakage, or dissipation, of information over time, preventing the neural current from providing a lossless summation of incoming signals. The leakage results in an exponential approach to an asymptotic neural current in response to a fixed sensory input, or an exponential decay of the current to zero when stimulation is terminated. For an isolated cell, the time constant for decay of excitatory currents is very short (5–10 ms). With such a short time constant, the neural signal would decay very quickly in the brain, and actual accumulation of information would be very strictly curtailed. A mechanism that can counteract and indeed reverse this tendency is recurrent excitation. From the early days of neural modeling to the present, modelers working in a neuroscience framework have suggested that recurrent excitation may play a prominent role in maintaining activity in neural populations (Amit, 1989; Anderson et al., 1977; Edelman, 1982; Grossberg, 1976, 1978; Hopfield, 1982).

From the neuroscience perspective, an abundance of interactivity is evident. For example, in the visual cortex, 80% of the excitatory input to cortical cells is provided by lateral connections from other cells within the same layer or from connections from other layers further up in the processing stream (Douglas et al., 1995; Douglas & Martin, 1990). The neurophysiological data (e.g., Miller, Erickson, & Desimone, 1996) are consistent with the idea that the strength of the recurrent excitation (and thus the effective time constant) increases along the ascending stream of processing. This provides a constraint for the rates of the units that in the cascade model (McClelland, 1979) were arbitrarily chosen. From the computational point of view, the recurrent excitation can achieve two functions: the amplification of the incoming sensory input (Douglas et al., 1995; Douglas & Martin, 1990) and (when the recurrent effect is very strong) the maintenance of the neural activity during delay periods without a stimulus (Amit et al., 1994), obtaining thus a form of short-term memory (see also Anderson et al., 1977; Usher & Cohen, 1999). In this work, we limit the discussion to the weaker effect of recurrent excitation, which as we will see in our mathematical treatment of the model, decreases the effective rate of leakage or dissipation of the population activity as it amplifies the overall magnitude of the response to sensory input.

Lateral Rather Than Feed-Forward Inhibition

As previously noted, lateral inhibition appears to be a powerful mechanism for selection and has significant advantages compared with feed-forward inhibition. This computational advantage of lateral inhibition may explain in part why the cortex appears to adhere to the same general constraint that we impose on our models. It is very well known that axonal projections from one brain area to another have purely excitatory effects, while within each local brain region there are both excitatory and inhibitory interactions. Although it would be simplistic to identify conceptual levels of processing in information-processing models with separate regions of the brain, it may be reasonable to construe each brain region, at least in the visual system, as deriving a particular type of representation of the input, based on inputs from many other regions. If so, the use of excitatory connections between areas and the use of inhibition to mediate competitive interactions within areas may allow each region to select the representation that is most consistent with the possibilities suggested by information represented in every other area.

There is considerable neurophysiological evidence consistent with the idea that inhibitory influences are predominantly of the lateral or recurrent type. Figure 1 shows what the dynamics of neural activity would look like in this case, based on a simulation of two competing accumulators in a simplified version of the model we introduce below. The trajectory of the unit receiving the weaker input is nonmonotonic, showing transient activation followed by suppression. Such transients are a characteristic property of systems with lateral inhibition and arise because the amount of inhibition depends on the activity of the other units involved in the competition. Initially, activity is low, and therefore the external input dominates, but as the activity increases, a point when the inhibition from the stronger alternative outweighs the excitatory input to the weaker unit is reached. Another interesting feature is apparent: The activation of the unit receiving the stronger input shows a transient deceleration while the competition is being resolved; the magnitude of this deceleration is parameter dependent.

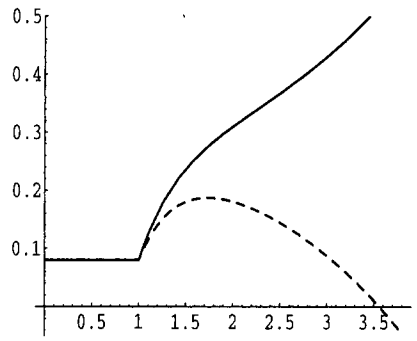

Figure 1. Dynamics of lateral inhibition. Activity of Unit 1 (solid lines) and Unit 2 (dashed lines) in response to a slightly stronger input to Unit 1 than Unit 2. Activations are based on the equations $dx_1 = (I_1 - x_1 - \beta f_2)dt/\tau$ and $dx_2 = (I_2 - x_2 - \beta f_1)dt/\tau$, where dx_1 and dx_2 represent the change in activation of each accumulator in a time step of size dt; x_1 and x_2 represent the activations of the two accumulators; I_1 and I_2 represent the excitatory input; βf_2 and βf_1 represent the inhibitory input of each accumulator to the other; and τ is a time scale. Here f_i (i = 1 or 2) is equal to x_i if x_i is greater than 0, or to 0 otherwise, on the assumption that activations less than 0 are not propagated from one neuron to another. The stimulus is applied at t = 1. Before this, both units receive a background input I_0 = 0.2. After stimulation, the input to Unit 1 increases to 0.7, and the input to Unit 2 increases to 0.6.

Corresponding characteristics of the dynamics of neural activation have been reported in several experiments. The clearest examples come from studies in which the input provides explicit support for a competitor of the designated target. In one such study (Chelazzi et al., 1993), spike trains were recorded from cells in the inferotemporal cortex of monkeys during the performance of a visual search task. Monkeys were first shown a target shape. After a delay, a visual display containing the target and a distractor was presented, and the monkey had to respond by moving its eyes toward the target. As can be seen in Figure 2, almost immediately after the presentation of the display with the choice stimuli, the activity in cells that respond to either of the two stimuli increased. However, after rising at about the same rate for about 100 ms, the activity in cells that responded to the cued stimulus showed a brief deceleration and then continued to rise while the activity of cells that responded to the distractor was suppressed. This suppression takes place before the execution of the eye movement; and therefore cannot be due to a change in the visual input. This pattern of results is highly consistent with the use of lateral inhibition for selection (Desimone, 1998; Reynolds, Chelazzi, & Desimone, 1999). A more detailed model based on lateral inhibition (which was mediated by a separate inhibitory pool) was shown to explain fine details of the activation trajectories, including the exact timing of deceleration seen in the activation of the cells that responded to the cued stimuli in this experiment (Usher & Niebur, 1996). Similar neural signatures have been reported in an evoked response potential (ERP) study (Gratton, Coles, Sirevag, Eriksen, & Donchin, 1988) in which participants must identify a target (S or H) flanked by presentations of the competing alternative (e.g., S in the center of a row of four Hs).

Simplifying Assumptions

As the foregoing discussion suggests, information processing in the brain may involve quite subtle dynamics, dependent on processing distributed over multiple brain regions. It is evident that large numbers of neurons participate in recurrent amplification and competition among alternative representations and that representations of response alternatives may be partially overlapping. Rather than incorporate all of this complexity directly in our model, we have adopted several simplifications. We examine models in which all of the action occurs in a single layer of units,

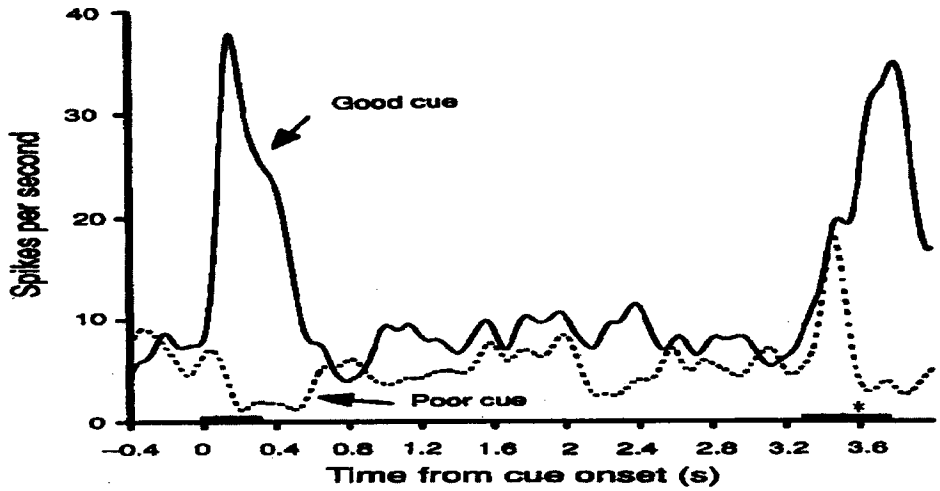

Figure 2. Cell response in monkey infero-temporal cortex, as reported by Chelazzi, Miller, Duncan, and Desimone (1993). Shown is the average firing rate of a cell when it responds to the cued shape (solid line) or to the distractor shape (dotted line). Black horizontal bars on the abscissa indicate presentation of cue and choice stimuli. Average time of saccade onset (297 ms) is indicated by an asterisk. From "A Neural Basis for Visual Search in Inferior Temporal Cortex," by L. Chelazzi, E. Miller, J. Duncan, and R. Desimone, 1993, *Nature, 363,* p. 345. Copyright 1993 by Macmillan Magazines Ltd. Reprinted with permission.

each corresponding to a choice alternative. These units function as accumulators, with one assigned to each of the N possible response alternatives in the experiment being simulated. We assume a set of inputs to these units (and connection weights modulating the effects of this input on the unit corresponding to each alternative), but we treat activations of these input units as occurring after a fixed delay from stimulus onset, neglecting the dynamics of the input. We also ignore the details of transmission delays along axons from the neurons representing the input and among the neurons participating in the same population, and indeed we ignore the additional lags that would be imposed by the interneurons that mediate competition among the different populations. Finally, the generation of the response is also simplified, adding an additional fixed delay. Again, this is merely a simplifying assumption. In reality, response generation is sure to be a complex process with intrinsic variability that can be modeled.

As we have stated previously, it is not self-evident that simplifications such as those we have adopted will lead to a useful model; this is essentially an empirical question, one that we explore extensively in what follows. To the extent that a simplified model provides a perspicuous account of experimental data, the simplifications allow for mathematical analysis and deeper understanding of the essential functional implications of the underlying mechanisms. We have adopted these simplifications for two principal reasons. First, they allow a fairly transparent mathematical analysis. Second, they facilitate direct comparison with existing stochastic information accumulation models and facilitate comparison with aspects of the experimental data. To the extent that the model provides a good fit to the data, it suggests that the simplifications it adopts are useful for characterizing the emergent properties of systems with far greater underlying complexity.

Related Models

We turn in the next section to the leaky, competing accumulator model itself. The model draws extensively on the whole body of prior work in stochastic information integration discussed above and incorporates additional principles, all of which have been widely used in other settings. We do not repeat all these citations here. However, we do wish to note the similarities between our current model and two other models. Smith (1995) has incorporated leakage into his stochastic accumulator model of luminance increment detection, so that the dynamics of the decision stage in his model are the same as the dynamics of an individual accumulator in our model in the absence of lateral inhibition. Busemeyer and Townsend (1993) have presented a model of the dynamics of deliberative choice among actions with uncertain outcomes in which leakage combines with differential gradients of approach to desirable outcomes and avoidance of undesirable outcomes leads to an expression for the time evolution of the decision variable that is equivalent to the expression we arrive at by combining leakage with lateral inhibition and self-excitation. Our model (early versions of which were reported in Usher & McClelland, 1994, 1995) arose independently from separate motivations (McClelland, 1993; Usher & Niebur, 1996). We address a set of issues not previously considered by models of this type, most notably the shapes of time–accuracy curves that arise in perceptual identification experiments, characteristics of latency discriminability functions, and effects of number of alternatives on RT. The fact that related models have been highly successful in accounting for data in other domains attests to the importance and generality of the underlying principles.

The Leaky, Competing Accumulator Model

We call our model the *leaky, competing accumulator* model because leakage and competition (via lateral inhibition) are the key features that differentiate it from the classical stochastic information accumulation models reviewed previously in this article. As we have stressed, our model shares the principles of gradual and stochastic accumulation of information with the classical models. Nonlinearity and recurrent excitation are encompassed by the model as well, but the main role of recurrent excitation is to balance leakage, and indeed we will mostly be concerned with the net leakage resulting from the combined effects of leakage and recurrent excitation. Similarly, as we see, nonlinearities also play a minor role.

In line with the principles and simplifying assumptions outlined above, the model that we study in the remainder of this article takes the form of a simple two-layer network, consisting of a set of input units indexed by j, over which the external input to the network is represented, and a set of accumulator units indexed by i, one for each of the response alternatives (Figure 3). Accumulator units are thought of as corresponding to populations of neurons, and they are characterized by two variables: Their activation, x_i, which corresponds to the population current, and their output, f_i, which corresponds to the population firing rate, as discussed above. We use a simple nonlinear function called the *threshold–linear* function in which $f_i = x_i$ for $x_i \geq 0$ and $f_i = 0$ for $x_i < 0$. Activations of accumulators determine responses: Under time-controlled conditions, the response is triggered by the unit that is most active at the time the response choice is made. In standard RT conditions, the unit that reaches a preset criterion first becomes the overt response, and the RT is determined by the time it takes to reach the criterion.

The focus of attention in the model is on the dynamic behavior of the activations of the units x_i, which can be seen as playing a similar role as the countervariables in the classical accumulator model. The starting point of our formulation of this dynamic behavior is a stochastic version (see Cox & Miller, 1965) of the cascade equation (McClelland, 1979), which has frequently been used in neural models to characterize neural activation processes (see Abbott, 1991; Amit & Tsodyks, 1991; Ermentrout, 1994). In psychological research, it has been used by Smith (1995) in his

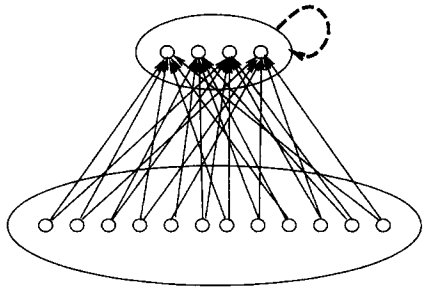

Figure 3. Network architecture for an N-alternative-choice task.

model of simple visual RT and by Busemeyer and Townsend (1993) in their model of stochastic decision making:

$$dx_i = [I_i - \lambda x_i]\frac{dt}{\tau} + \xi_i\sqrt{\frac{dt}{\tau}}. \quad (1)$$

Here, τ is a time scale chosen for convenience,[3] and ξ_i is a Gaussian noise term with zero mean and variance σ^2. This equation implies that within a time interval dt/τ, the change in the activation of an accumulator unit, dx_i, is driven by input from other units, I_i, with a characteristic decay rate λ, which reflects leakage of the activation. The noise term scales with the square root of dt/τ, because the variance of uncorrelated stochastic random variables is additive, leading to the square-root behavior for the standard deviation.

Equation 1 characterizes a leaky, stochastic accumulator. For tractability, we consider the case in which I_i suddenly changes from 0 to a fixed positive value. In this case, the equation produces an exponential approach (like a charging capacitor) to a steady state in which $dx_i/dt = 0$. In the absence of noise, the value of x_i at steady state, x_i^{asy}, is equal to I_i/λ. In the presence of noise, x_i is a random variable whose distribution evolves over time so that it comes to fluctuate around this same value, which is in this case the expected value of the mean of x_i for long times after the onset of the fixed input I_i.

The input I_i can be decomposed into three distinct components: an external source, I_i^{ext}, a recurrent excitatory source, I_i^{rec}, coming from the unit back to itself, and lateral inhibition between accumulator units. I_i^{rec} is simply a scalar multiple of the output of the unit itself. That is, $I_i^{rec} = \alpha f_i$ where α is the scaling factor for the recurrent excitation.

The external input, I_i^{ext}, is thought of as a weighted sum of the output of the input units, so that $I_i^{ext} = \Sigma_j W_{ij} f_j$. For simplicity, we treat the input units as having activations of 0 before the presentation of the stimulus. Likewise the accumulator units are initialized at 0 at stimulus onset except where otherwise stated. Output of the input units is set to values $f_j(\Phi)$ after the presentation of the stimulus Φ. Because of the threshold–linear function mapping activations to outputs, the f_j will always be ≥ 0. We also impose the restrictions that the connection weights W_{ij} are always ≥ 0. As a result, $I_i^{ext}(\Phi)$ will also be ≥ 0 for all accumulator units. To simplify the notation, we use the symbol ρ_i to represent this feed-forward input to accumulator i.

Finally, we introduce the inhibitory influences exerted by each accumulator unit on each other accumulator unit. As with the recurrent and external input, this input is a weighted sum of the output of the relevant units. For simplicity, the inhibitory influences of each unit on each of the other units all have the same weight β. Thus, the total inhibitory input to a unit becomes $\beta\Sigma_{i' \neq i} f_{i'}$. The equation that incorporates all of the separate influences in our model is therefore

$$dx_i = [\rho_i - \lambda x_i + \alpha f_i - \beta\Sigma_{i' \neq i} f_{i'}]\frac{dt}{\tau} + \xi_i\sqrt{\frac{dt}{\tau}}. \quad (2)$$

The equation above contains several parameters, each of which introduces a degree of freedom in fitting the model to data, and it is nonlinear, thereby making analysis quite complex. However, when our model operates within certain constraints on the parameters that we discuss below, the selection of one alternative over others occurs through a competition that reaches its climax when the activations of units that might win the competition are close to each other and greater than 0. Once the activation of a unit goes below 0, it has essentially lost out in the competitive dynamics and has no appreciable chance of becoming the most active unit later. In the light of this, note that the above equation is completely linear as long as all of the x_i are greater than 0. In this case, we can replace the f terms with their x counterparts:

$$dx_i = [\rho_i - (\lambda - \alpha)x_i - \beta\Sigma_{i' \neq i} x_{i'}]\frac{dt}{\tau} + \xi_i\sqrt{\frac{dt}{\tau}}. \quad (3)$$

Note that whenever $x_i > 0$ for all i, this equation gives exact results. In some other important cases, as we shall see, it provides an extremely useful approximation. In this equation, we see that λ and α function together, so that their difference, $\lambda - \alpha$, becomes a net leakage, which can be replaced with the single parameter, k. When k is greater than 0, the net effect is a decay toward 0 that produces stability; when it is less than 0, the activation tends to self-amplify and is not stable. We include some examination of the self-amplifying case in our analysis below, but for the simulations of data from perceptual classification experiments, we assume that decay dominates and so will constrain k to be greater than 0.

A limitation of Equation 3, which may be especially relevant when the number of competing units becomes large or when the external input to one or more of the competitors is very weak ($\rho_i \approx 0$), is that the activations x_i of many of the units may go negative well before the time of responding. In that case, whenever a unit's activation becomes negative, use of Equation 3 leads to an inappropriate and unphysiological boosting of other units through the $-\beta\Sigma_{i' \neq i} x_{i'}$ term (subtracting the effect of the negative activation leads to a positive influence). The nonlinear version of the model in Equation 2 avoids this by using f_i and $f_{i'}$ rather than x_i and $x_{i'}$ to mediate both the self-excitation and the recurrent inhibition, but this then forces us to separate λ and α. We can avoid this difficulty by adopting a slightly different formulation, in which we use Equation 3, with the proviso that activations that go below 0 are immediately truncated to 0, so that negative activation never propagates. We show in Appendix A that this formulation closely approximates Equation 2 when the parameters are constrained within the range conforming to our conception of the characteristics of the underlying physiology.

For time-controlled tasks with two response alternatives, where the response simply depends on which unit is most active at the moment of the choice response, simulations reported below show that it does not matter whether the truncation at 0 is imposed or not, because this does not alter which unit is most active. Given this, it is possible to derive useful analytical results by ignoring the truncation and treating the system as completely linear. In standard

[3] The time scale τ can be ignored if one uses a natural time scale such as seconds or milliseconds. However, when integrating the equations numerically, it is more convenient to use a time scale that is on the order of magnitude with the temporal process observed (say, 100 ms, for time–accuracy curves, as shown in Tables 1 and 2) and to ensure that the time step, dt, is smaller than τ, as required for Euler integration. A value of $dt/\tau = 0.1$ is used in all of our numerical integrations, corresponding to time steps of about 10 ms.

RT tasks, however, omitting the truncation can spuriously speed the time it takes the unit that eventually wins to reach the response criterion due to the unphysiological boosting effect mentioned above. In that case, we are forced to rely on simulations using the truncation.

In summary, then, the evolution of activations of units in the model is governed by

$$dx_i = [\rho_i - kx_i - \beta\Sigma_{i' \neq i} x_{i'}] \frac{dt}{\tau} + \xi_i \sqrt{\frac{dt}{\tau}}$$

$$x_i \rightarrow \max(x_i, 0). \quad (4)$$

This is the simplest formulation we have been able to develop that incorporates the essential stochastic and nonlinear elements into a system of leaky, competing accumulators.

As previously noted, we assume that there are additional processing stages between the actual sensory surface and the input layer of the model, as well as additional stages between the initiation of a response based on the state of the accumulator units and the muscle contractions affecting the overt response, which introduce delays that we treat as fixed in the propagation of information from input to response. As in other models, these delays are lumped into a single delay parameter T_0. Collectively, they give rise to the initial flat portion of the empirical curves relating time since stimulus onset and activation or accuracy. For completeness, some trial-to-trial variability could be included in the value of T_0. Such variability would produce a less abrupt transition of time–accuracy functions from their initial flat portion into the subsequent curvilinear approach to asymptote, as is seen in empirical data (McClelland, 1979). For simplicity, we have not incorporated this assumption in our simulations.

Applying the Model to Tasks Involving Only Two Alternatives

Because most of the classical literature revolves around experiments in which there are only two alternatives, we consider this case first. For this case, we are concerned with only two decision variables, x_1 and x_2, corresponding to the activations of accumulators for each of the two alternatives. Equation 4 and its linear approximation (Equation 3) still characterize the system's dynamic behavior in this case, with the simplification that each accumulator only receives inhibition from one other accumulator.

In many of the two-alternative experiments considered in this article, a stimulus input is varied along a continuum, such as the difference in lengths of two lines, and the task is to indicate which is longer. In such cases, as one moves along the continuum, it is natural to suppose that as the input supporting one alternative increases, the input supporting the other decreases. For this reason, we have adopted the restriction that the sum of ρ_i remains constant at all points along the continuum. This is similar to an assumption of Townsend's (1981) bounded performance model, which included an upper bound on transmitted information. Without loss of generality, we set the value of the sum to 1, so that ρ_i represents the fraction of the support provided by the input to each of the two alternatives. This choice serves to reduce the number of free parameters in the model. Thus, with $\Sigma_i \rho_i = 1$, it follows for the two-alternative case that $\rho_2 = 1 - \rho_1$.

When the task is set up as a standard RT task, it is assumed that the participant does not respond until some criterial state of activation (i.e., a threshold) is reached. Many types of response criteria have been proposed in the literature (Ashby, 1983; Audley & Pike, 1965; LaBerge, 1962; Laming, 1968; Link & Heath, 1975; Ratcliff, 1978), and as we already noted in the introduction, the choice of criterion can affect the predicted RT and accuracy data. As also noted, the analysis is simplified when the amount of time allowed for processing is controlled (e.g., by means of a signal that indicates when to respond). We consider this case first before turning to the greater complexity of the standard RT paradigm.

Time–Accuracy Functions Under Time-Controlled Conditions

Under time-controlled conditions, a general formulation would assume that participants compare the difference in activation of the two detectors, $x_1 - x_2$, with some criterion B, choosing Alternative 1 whenever $x_1 - x_2 > B$ and Alternative 2 otherwise. For simplicity, we consider cases in which B is zero (equivalent to choosing the most active alternative). Note that the role of the criterion in this formulation is only to bias the choice between alternatives and not to determine its timing. The experimental results in the time–accuracy literature use the bias-free measure d' as the dependent variable, and choosing $B = 0$ does not affect the sensitivity, as measured with d'.

Ratcliff (1988) has suggested that participants may not always wait until the response signal is presented to select a response in time–accuracy experiments but may select a response based on reaching a decision boundary, if this occurs before the response signal is presented. In our analysis, we do not consider the effects of including these decision boundaries, focusing our attention on the information accumulation process itself. Effects of introducing decision boundaries are discussed below, after the presentation of this analysis.

We explore the adequacy of the leaky, competing accumulator model to account for data from time-controlled experiments through a sequence of steps. First, we study the approximate, linear version of the model (Equation 3), showing how it reduces, in the time-controlled case with two alternatives, to a very simple expression for the relationship between time and accuracy, eliminating two of the free parameters of the full nonlinear version of the model. We then compare the linear approximation with the full model, to demonstrate that the full model still exhibits the same form of time–accuracy dependency that is seen in the linear version and to consider what role if any the extra parameters in the full model actually play in the underlying process. Given the adequacy of the approximation, we then compare the simple expression derived from it for the shapes of time–accuracy curves with a comparable expression that results from assuming a nonleaky accumulation process with drift variance and consider how well each conforms (a) to the shifted exponential function often used to summarize the data from time-controlled experiments, (b) to a focused analysis of the shape of the early part of the time–accuracy curve, and (c) to the actual data obtained in a previously unreported experiment.

Analysis of the Linear Approximation of the Leaky, Competing Accumulator Model

Our analysis of the linear approximation of the model begins by defining a new variable, $x = x_1 - x_2$, which simply represents the difference in activation between the two accumulator units. To obtain this variable, we begin by instantiating Equation 3 for x_1 and x_2 in the two-alternative case:

$$dx_1 = [\rho_1 - kx_1 - \beta x_2]\frac{dt}{\tau} + \xi_1\sqrt{\frac{dt}{\tau}},$$

$$dx_2 = [\rho_2 - kx_2 - \beta x_1]\frac{dt}{\tau} + \xi_2\sqrt{\frac{dt}{\tau}}. \quad (5)$$

We can now obtain an equation in x by subtracting the second equation from the first and making use of the fact that $\rho_2 = 1 - \rho_1$:

$$dx = [(2\rho_1 - 1) - (k - \beta)x]\frac{dt}{\tau} + \xi\sqrt{\frac{dt}{\tau}}. \quad (6)$$

The variable x behaves similarly to the decision variable in the classical diffusion model without drift variance, with $\nu = 2\rho_1 - 1$ corresponding to the diffusion drift. The standard deviation σ of the noise in the classical diffusion process corresponds directly to the standard deviation of the Gaussian random variable ξ for the leaky, competing accumulator model. The process described by Equation 6 is well known in statistical physics; it is called an *Ornstein–Uhlenbeck* (OU) process (see Ricciardi, 1977).

Unlike the classical diffusion model, the OU process is also characterized by an effective differential leakage or decay term $K = k - \beta$. Note that this term K extends beyond the effective unitwise leakage term k, which does not incorporate the inhibitory influences of other units, to provide a measure of the net leakage in the difference between the two units' activations. As with k, when $K = 0$, differential information integration occurs without loss. In this case, a classical diffusion process results. The OU diffusion with both negative and positive K value has been used by Busemeyer and Townsend (1993) in their model of decision making, where the sign of the K coefficient relates to approach–avoidance characteristics of the decision. A version with negative K arises in Smith's (1995) model of simple reaction time for luminance-increment detection.

To characterize the time–accuracy curves, a measure of d' has to be obtained as a function of the time t when the response is selected. According to signal-detection theory, d' is computed as the separation between the means of distributions associated with the two types of trials, divided by their standard deviation. Consider the case in which there are two kinds of trials: those in which Stimulus 1 is presented and therefore $\nu = \rho_1 - \rho_2 > 0$ and those in which Stimulus 2 is presented and therefore $\nu = \rho_1 - \rho_2 < 0$. For this situation, we can designate as a hit those cases in which $\nu > 0$, and at the moment a response must be made, $x(t) = x_1(t) - x_2(t) > 0$. A false alarm would correspond to those cases in which $\nu < 0$ and $x(t) > 0$. Let us denote the expected value of the x variable when $\nu > 0$ as $\mu_+(t)$ and as the expected value when $\nu < 0$ as μ_-. Due to the symmetry of our assumptions, $\mu_- = -\mu_+$, and the difference between the expected values just becomes 2μ (where μ represents the absolute value of μ_+ or μ_-). Thus, we can write an expression for $d'(t)$:

$$d'(t) = \frac{2\mu(t)}{SD(t)}. \quad (7)$$

The time evolution of the mean (μ) and the standard deviation of $x(t)$, as well as the time evolution of its distribution $P(x, t)$, can be obtained for the process described in Equation 6, using the theory of stochastic processes (Ricciardi, 1977; for computation of the probability for correct choice, $P(t)$, see Appendix B). If at time $t = 0$, the activation is initialized as $x(0) = 0$, then the distribution of x at a later time t is a Gaussian variable with distribution $P(x, t) = N[\mu_A(t), SD(t)]$, where N denotes the normal density distribution whose mean, μ, and standard deviation are given by

$$\mu(t) = \frac{\nu}{K}[1 - \exp(-Kt)],$$

$$SD(t) = \frac{\sigma}{\sqrt{K}}\sqrt{[1 - \exp(-2Kt)]}. \quad (8)$$

From this (and using Equation 7), the following expression of the time–accuracy function is obtained:

$$d'_{OU}(t) = d_{asy}\frac{1 - \exp(-Kt)}{\sqrt{1 - \exp(-2Kt)}}. \quad (9)$$

This determines the time evolution of an OU process, where d_{asy} is the asymptotic value of d' obtained for unlimited time:

$$d_{asy} = \frac{2\nu}{\sigma}\sqrt{\frac{1}{K}}. \quad (10)$$

We see in this expression that d_{asy} depends on the difference between the input to the two accumulators ν, the standard deviation of the noise, and the net decay K. When K is greater than 0, d_{asy} takes on a finite value, showing that the leakage or decay of information explains the pattern of data found in time–accuracy experiments.

The special case of $K = 0$ leads (using Equation 6) to a classical diffusion process. In this case, the distribution $P(x, t)$ is also a Gaussian $N[\mu(t), SD(t)]$, whose mean increases linearly with time ($\mu(t) = \nu t$) and whose standard deviation increases as the square root of time [$SD(t) = \sigma\sqrt{t}$]. In this case, as noted in the introduction, the accuracy increases with processing time to infinity, even for a highly degraded signal-to-noise ratio $\nu/\sigma \ll 1$. This is indeed expected because as the integration time increases, the noise is progressively averaged out; a noiseless estimation of the stimulus can thus be obtained with infinite integration time.

Comparison of the Linear Approximation With the Nonlinear Model

The OU function represents an approximation of the actual processes assumed to operate in our complete model, because it ignores the effects of the nonlinearity of the activation function. It therefore becomes important to understand whether the complete model adheres to the properties of the simplification. Our first goal was to determine whether the time–accuracy curves still corresponded to the shape expected under the OU equations and to

establish whether the differential effective leakage, $K = k - \beta$, still determines the shapes of these curves. To this end, we ran simulations of the nonlinear diffusion process in which negative activations were truncated to zero, as specified in Equation 4. The net unitwise leakage k was set at 0.2 throughout, and different values of β were used to produce different values of K in different simulations.

For each simulation trial, the activations of the units were initialized to 0. At every iteration of the simulation, units received an increment in activation according to the following expressions, following which negative values of x_1 or x_2 were set to 0:

$$dx_1 = [0.5(1 + \nu) - kx_1 - \beta x_2]\frac{dt}{\tau} + \xi_1\sqrt{\frac{dt}{\tau}},$$

$$dx_2 = [0.5(1 - \nu) - kx_2 - \beta x_1]\frac{dt}{\tau} + \xi_2\sqrt{\frac{dt}{\tau}}. \qquad (11)$$

In these expressions, we have replaced ρ_1 with $0.5(1 + \nu)$ and ρ_2 with $0.5(1 - \nu)$. This replacement preserves $\rho_1 + \rho_2 = 1$. As above, ν represents the difference $\rho_1 - \rho_2$ in input to the two accumulator units. For all the simulations in this section, dt/τ, the time step of the simulation, was set to 0.1, ν was set to 0.1, and σ, the standard deviation of the noise variables ξ_1 and ξ_2, was set to 1.58 (so that the effective noise per step when multiplied by $\sqrt{0.1}$ is 0.50).[4] For the sake of calculating time–accuracy curves, at every iteration, the simulation program tabulated whether $x_1 > x_2$. The number of times $x_1 > x_2$ was then divided by the number of simulation trials (5,000) to determine the probability correct at each time step.

Figure 4 shows that the time–accuracy curves do indeed depend on the differential effective leakage K. At $\beta = 0.2$, the inhibition and the leakage are balanced so that $K = 0$. In this case, the time–accuracy curve increases indefinitely toward perfect performance, as expected for a diffusion process without loss. For other values of β, inhibition does not perfectly balance the effective unitwise leakage, and there is information loss resulting in asymptoting curves. Therefore, we see that the curves depend only on the absolute value of the differential effective leakage, $|K|$. For example, virtually identical time–accuracy curves were obtained for $\beta = 0$ (square symbols) and for $\beta = .4$ (diamonds) as they correspond to $K = 0.2$ and to $K = -0.2$, respectively. Finally, the lowest performance was obtained for $\beta = .6$, which corresponds to $K = -0.4$ in Equation 10.

To test if the simulation results were consistent with the analytic time–accuracy curves, we fitted the simulated time–accuracy curves with the formula derived for the OU process in Equation 9. Two parameters, d_{asy} and the OU integration time constant, $1/K^{OU}$, were allowed to vary while the offset time T_0 was fixed at zero. We denote the parameter of the OU fit by K^{OU} to distinguish it from the value of K used in the simulation of the nonlinear process. The best fits, shown with solid lines in Figure 4, are consistent with the simulation results. Moreover, the fit parameters correspond closely to the values used in the simulation (see figure caption).

Note that values of K of equal magnitude but of opposite signs result in the same time–accuracy curves. A negative value of K in Equation 6 has the implication that the system is dominated by divergence instead of decay (this can happen with strong recurrent excitation or strong lateral inhibition). Strikingly, however, equivalent degrees of divergence and decay lead to the equivalent evolution of d' over time. This is due to the symmetry of the OU speed–accuracy curves to the inversion of sign of K (e.g., from 0.2 to -0.2).[5]

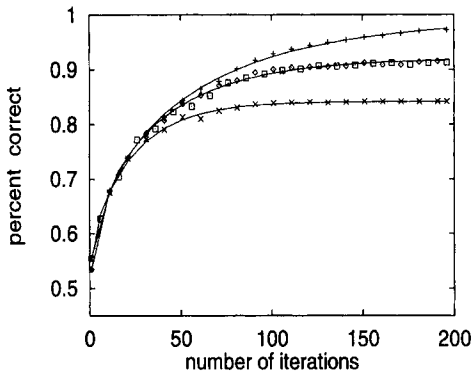

Figure 4. Time–accuracy simulations at four levels of lateral inhibition. The probability correct $P(t) = \text{prob}[x_1(t) > x_2(t)]$ is displayed as a function of time. In all simulations, the unitwise effective leakage factor, k, was 0.2, the stimulus discriminability corresponded to $\nu = 2\rho_1 - 1 = 0.1$, and the noise factor was $\sigma = 1.58$. Results for balanced inhibition and leakage ($\beta = .2$; + symbols [top curve]) and for three other cases: two in which $|k - \beta| = 0.2$ (next two superimposed curves: $\beta = 0$, square symbols, and $\beta = .4$, diamond symbols) and one in which $|k - \beta| = 0.4$ (bottom curve: $\beta = .5$, X symbols). The continuous lines are fits with the analytic time–accuracy curves for the Ornstein–Uhlenbeck (OU) process (Equation 9). Correspondence between the parameters of the fitted curves and their expected values based on the approximate linear model is very close. For example, the fitted time constant $1/K^{OU}$ would correspond to $\tau/dt/K$ from the actual stochastic simulation. For $|k - \beta| = |K| = 0.4$ and $dt/\tau = 0.1$, we expected $1/K^{OU}$ to be 25, which was the value obtained. For $|K| = 0.2$, we expected $1/K^{OU} = 50$, and a value of 48 was obtained. For the simulation with $|K| = 0$, we expected $1/K^{OU}$ to take a very large value, and 225 was obtained. This large value $1/K^{OU}$ would correspond to a very small $|K|$ (0.03), and differences between the curves for $|K| = 0$ and $|K| = 0.03$ would only show up at times much longer than the time interval covered by the simulation.

The symmetry of the OU speed–accuracy curves to K inversion implies therefore that the two strategies are equivalent in terms of what we call their *sensitivity dynamics* (i.e., the trajectory of d' as a function of time, where d' at time t is based on the probability that the difference variable x is greater than 0 at t. Despite this symmetry, however, the distributions of trajectories of the underlying activations are very different in these different situations. Busemeyer and Townsend (1993) have noted these characteristics of the OU process in their model of decision making, where the

[4] In the Monte Carlo simulation program, we first add to each accumulation the deterministic input (ρ_i) and the Gaussian distributed variable (ξ_i). The leak factor (multiplication by $1 - kdt/\tau$) and the lateral inhibition (subtraction of $\beta dtx_{j\neq i}$) are then applied to all units. This is equivalent to applying all the factors at once with the ρ and σ parameters scaled by $1 - kdt/\tau$.

[5] The symmetry can be verified by checking that the OU d' time–accuracy functions in Equation 9 are invariant under $[1 - \exp(-Kt)] \rightarrow [\exp(Kt) - 1]$: $[1 - \exp(-Kt)]/\sqrt{1 - \exp(-2Kt)} = [\exp(Kt) - 1]/\sqrt{\exp(2Kt) - 1}$.

sign of the K coefficient relates to approach or avoidance of the choice alternatives. To illustrate these characteristics of the system, we display in Figure 5 the density distributions of x derived from the full model, at three different times after the stimulus presentation, in three different situations, all of which involve $|K| = 0.2$. The first case involves what might be called *pure dissipation:* The unitwise effective decay leads to gradual information loss $k = \lambda - \alpha = 0.2$ and the inhibition β is set to 0. The second case involves what might be called *pure self-enhancement:* β is set to 0, as in the first case, and setting k to -0.2 (which could arise if the hidden self-excitation parameter α were slightly greater than the hidden passive decay λ), we see growth of activation through self-excitation. The final case involves dissipation coupled with lateral inhibition ($k = 0.2$ and $\beta = 0.4$).

We observed that for the purely dissipative case, the distributions remained bounded, whereas for the self-enhancement case, they diverged. In the second case, the mean of the distribution moved quickly toward the correct response side, but the standard deviation increased in a corresponding manner, so that their ratio (which determines d') was the same in the two cases. For the final case with dissipation and lateral inhibition, we observed a new phenomenon. Despite the fact that d' and P were the same in the previous two cases, the distribution became bimodal. This was because with inhibition, eventually one of the units is suppressed to zero. When this happens, owing to the nonlinearity, the effect of inhibition on the active unit disappears; thus, the active unit remains subject only to dissipation.

This analysis shows that a system incorporating leakage, competition, and a simple truncation nonlinearity is able to generate what amounts essentially to a binary decision. Although the time–accuracy curve is unaffected (the area at the right of the origin remains the same), the nonlinearity might make an important difference to actual RTs. The assumption under which the time–accuracy curves discussed above were obtained was that the participant simply chose the unit that was most active at the moment of decision. However, this assumption is an idealization; in cases where the difference is small, the response selection itself might be subject to error, or it might itself take time. Because the segregating effect of competition tends to eliminate cases in which the difference in activation between the two alternatives is small, this process might actually make it easier to make a quick decision. Note also that this separation is a process that takes place over time. Future work might examine whether the lack of separation early in processing could be partially responsible for the relatively long response latencies (measured from the response signal) at short signal lags that are typically observed in time–accuracy experiments.

Comparison of the Effects of Leakage and Drift Variance

As explained in the introduction, the bounded value of performance obtained with long information integration times can be addressed by assuming that the drift parameter v is subject to variance σ_d from trial to trial (Ratcliff, 1978). With this assumption, Ratcliff (1978) showed that the time–accuracy function becomes

$$d'_{DDV}(t) = \frac{d_{asy}}{\sqrt{1 + \frac{\sigma^2}{\sigma_d^2 t}}}. \tag{12}$$

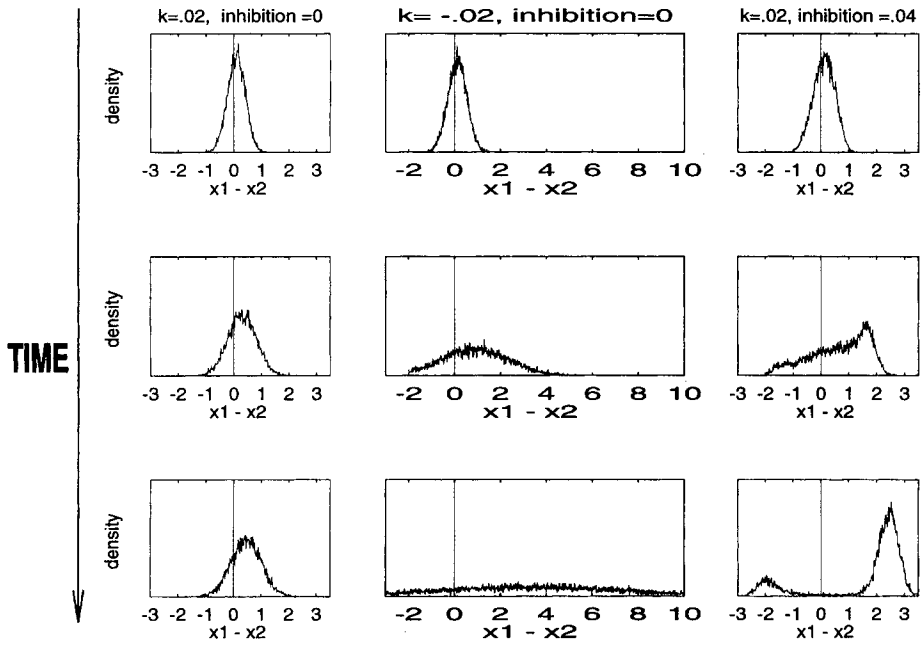

Figure 5. Density distribution for $x_1 - x_2$. (Left column) dissipative case ($k = 0.2$), (center column) self-expanding ($k = -0.2$, $\beta = 0$), (right column) dissipation and inhibition ($k = 0.2$, $\beta = .4$). (Top panels) $t = 10$ iterations, (middle panels) $t = 50$ iterations, (bottom panels) $t = 195$ iterations for the left and right panels (asymptotic distributions) and $t = 100$ iterations for the middle–low panel.

We emphasize that this function describes the sensitivity of the information accumulation process as a function of time and does not take into account the possible effects of use of decision boundaries that can terminate processing if one is reached before the presentation of a response signal (Ratcliff, 1988). For the present, we wish only to contrast the sensitivity dynamics that result from assuming drift variance to those that result from the assumptions of our model and to consider how well each approach can account for the shapes of time–accuracy curves without further assumptions. We return to a consideration of the role of decision boundaries below.

In Figure 6, we display the normalized time–accuracy function (i.e., d'/d_{asy}) for the OU process (Equation 9: long-dashed line) and for diffusion with drift variance (Equation 12: short-dashed line) for the same asymptotic value of performance. Although the two types of functions have some similar qualitative properties (a gradual increase to asymptotic value), one can clearly see that the OU curve is close to a pure exponential (solid line), whereas the DDV curve approaches the asymptote faster at the beginning and more slowly toward the end (compared with the exponential). A further comparison of these curves (using logarithmic scales) is displayed in Appendix C.

Shape of Empirical Time–Accuracy Curves

Studies that have examined accuracy of performance as a function of processing time have generally found that performance improves and then levels off, following a function that can be approximated as a shifted exponential approach to asymptote (Busey & Loftus, 1994; Corbett & Wickelgren, 1978; McElree & Dosher, 1989; Townsend, 1981; Wickelgren & Corbett, 1977). When processing time is controlled by a response signal or deadline procedure, it has frequently been observed that time–accuracy curves (measured in terms of sensitivity, d') are generally well approximated by a shifted exponential (Wickelgren, 1977). That is, responses made at very short times after stimulus onset have only chance accuracy. Accuracy begins to rise after some time lag T_0 and then follows an exponential approach to asymptote. Because the DDV model predicts a shifted curve that has a shape that is somewhat different from the exponential, it is worth asking whether the shifted exponential or the DDV curve provides a more accurate characterization of experimental data. McElree and Dosher (1989) compared the fit of the shifted exponential and the shifted DDV curve (Equation 12) to data from a time-controlled version of the Sternberg task. Consistently within all experimental manipulations, the exponential curves gave slightly better fits to the data. The DDV curves seem to approach asymptote more slowly than the data, leading to systematic deviations in the fit. Because OU-shaped curves are very similar to the exponential, the findings provide a possible indication that actual time–accuracy curves may be more consistent with our leaky, competing accumulator model. Of course, this is only the most tentative indication about the possible adequacy of the leaky, competing accumulator model. For one thing, there is literature that suggests that Wickelgren's shifted exponential may not provide the best fit to the early portion of time–accuracy curves. Because the OU and the DDV curves both deviate from the shifted exponential in this portion of the curves, it is important to understand whether either or both models can address this aspect of the data. More generally, the adequacy of the (shifted) OU and DDV curves has never been compared directly, and therefore it will be useful to provide such a comparison. Finally, the OU curve is only an approximation to the complete, nonlinear model, and therefore it is important to verify through stochastic simulations that the model can in fact simulate the obtained findings from actual experiments.

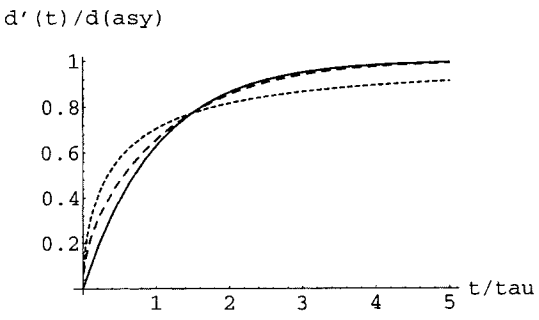

Figure 6. Normalized time–accuracy functions—$d'(t)/d(\text{asy})$, where d(asy) corresponds to d_{asy}, the asymptotic value of d', as described in the text—for the diffusion-with-drift-variance (DDV) model ($\sigma/\sigma^d = 1$; dotted line), the Ornstein–Uhlenbeck (OU) process ($K = 0.93$; dashed line), and the Wickelgren-type exponential curve (solid). Observe that the DDV curve is further away from the exponential curve than the OU curve, both when small and when large. No value of σ/σ_d can bring the DDV curve closer to the exponential.

First, we consider the initial portions of empirical time–accuracy curves. As one can observe in Figure 6, at short times ($t < T = 1/K$), the OU (as well as the DDV) time–accuracy curve shows some deviation from a pure exponential (full and dashed lines). The OU has a slight deflection upward, which (at very brief times) could be detected in perceptual choice experiments with high discriminability. A large number of such studies performed using a speed–accuracy trade-off (SAT), where responses are required within a time window, have reported a linear relationship between the amount of transmitted information, $I = P \log_2(P) + (1 - P)\log_2(1 - P) + 1$, and the time available for response (Hick, 1952; Pachella & Fisher, 1972; Salthouse, 1981). In particular, Salthouse compared a number of performance measures such as the information transmitted, I, (Hick, 1952), raw proportion correct, P, d', d'^2, and log odds [$\log(P/1 - P)$], to find which one provides the best linear fit to SAT data. Although the differences were not big, the use of information transmitted led to the best linear fit. The OU model can explain the linear trend in information transmitted at short times. This arises directly from the upward initial deviation of the OU time–accuracy curve from a pure exponential. A similar behavior is obtained using a diffusion process without drift variance. The diffusion process with drift variance results in a linear trend over a somewhat smaller time range. This is shown in Figure 7, where we plot I (in bits of information) for three time–accuracy curves: OU (solid line), Wickelgren's (1977) exponential (long-dashed line) and DDV (dotted line).

One can observe that unlike the OU and the DDV curves, the exponential curve shows a strong deviation from linearity in I at short times. Because the exponential time–accuracy curve (Wickelgren, 1977) is linear at short times, it becomes bilinear when

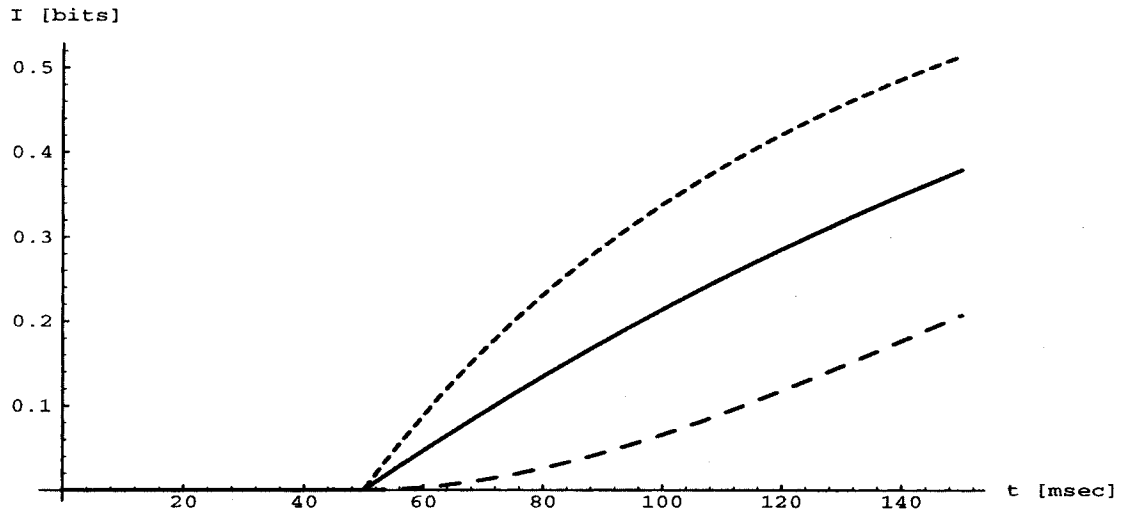

Figure 7. Transmitted information, I, for three time–accuracy curves: (a) Ornstein–Uhlenbeck process (solid line), (b) shifted exponential (long-dashed line), (c) diffusion-with-drift-variance (DDV; short-dashed line). For all three processes, the asymptotic accuracy was $d'_{asy} = 5$, the offset was $T_O = 50$ ms, and the time constant was $\tau = 300$ ms (σ^2/σ_d^2 for the DDV model, Equation 12).

transformed to a measure of transmitted information (Appendix D).[6] On the other hand, the OU and the diffusion models have characteristics that cancel out the bilinearity in the transmitted information expansion, resulting in a linear trend in I. For the OU model, the range over which the linear approximation holds is limited by the time constant, $1/K$. A typical time constant of 300 ms (in keeping with the range of values from fits of our model discussed below) provides a linear regime of about 150 ms, consistent with the range over which linear information SAT curves are reported in experiments (see, e.g., Luce, 1986, p. 243).

We next consider how well the OU and the DDV curves can be fitted to the actual shapes of complete time–accuracy curves. To our knowledge, such a comparison has not previously been carried out. Although McElree and Dosher's (1989) comparison of the fit of the shifted exponential and of the DDV time–accuracy curves suggests a possible advantage for the leaky, competing accumulators, the quantitative advantage of the exponential was very small and it is not known whether the advantage would actually still be found in a comparison of the fit of the DDV to the curves expected under the leaky, competing accumulator model, because these are not exactly exponential in shape. Moreover, the data fitted by McElree and Dosher come from memory search tasks and not perceptual choice, which is the focus of the present analysis.

To address these limitations, we have sought to explicitly compare the fits of curves based on the DDV function with those based on the OU function on time–accuracy data from a binary perceptual choice task. Because the difference between the OU and the DDV time–accuracy curves is larger at longer times, we need a set of data on binary perceptual choice in which the asymptotic level of performance is manipulated well below perfect performance levels and responses at long time intervals are obtained.

Experiment 1: Time–Accuracy Curves in Binary Perceptual Choice

The experiment we report here meets these requirements. The experiment made use of a difficult perceptual judgment: Similar to Swensson (1972), each participant was required to decide on each trial whether an almost-square tilted rectangle was longer toward the upper left or the upper right. Each participant was induced to respond at different times after the onset of the display by the use of response signals. Three levels of difficulty were tested at each of 10 different signal lags.

Method

Materials and apparatus. Participants were seated in front of a Tektronix 602 point-plotting oscilloscope, on which all of the displays were presented. A forehead rest bar was used to maintain head position at a distance of 37 cm from the oscilloscope face. The displays were six rectangles based on a square tilted 45%. Each side of the base square had length of 1,000 in the nominal units of the plotting system and subtended a visual angle of approximately 2 cm on the face of the oscilloscope. The six stimuli were formed by increasing the lengths of the two sides pointing to the upper left or of the two sides pointing to the upper right by 1, 3, or 5 points over the base length of 1,000 points.

Procedure. At the beginning of each trial, a fixation point was presented at the center of the display. The rectangle was presented 500 ms later, centered at fixation, and remained on the screen until the occurrence of the response signal. A two-button response box was used, and participants were instructed to respond by pressing the left (or right) button to indicate the judgment that the rectangle was tilted toward the upper left (or

[6] This can be understood (see also Appendix D) by developing the formula for the transmitted information, $I = P\log_2(P) + (1 - P)\log_2(1 - P) + 1$ (see, e.g., Hick, 1952) in a power series, in terms of the deviation δP around the guessing value of $P = .5$ and noticing that the lowest order term is bilinear in δP.

upper right). A response signal in the form of a beep occurred at one of 10 intervals after the onset of the stimulus. Participants were asked to respond within 200 ms of the onset of the response signal and, within this constraint, to be as accurate as possible. Their pay included a base amount per session, plus a bonus based on the number of correct responses within 200 ms of the response signal. Visual feedback was used to indicate whether the response occurred within the 200 ms, whether or not it was correct.

Participants. Three participants with normal or corrected-to-normal vision were tested in the experiment. They were tested for approximately an hour a day over a period of about 2 weeks.

Design. Each participant was tested on all three discrimination levels. For each discrimination level, there were 10 response-signal lags. Within each session, all combinations of discrimination level (1, 3, or 5 parts per 1,000), stimulus (longer toward the upper left or upper right), and signal lags were randomly intermixed. The lags used were 0, 50, 100, 200, 300, 400, 600, 800, 1,000, and 2,000 ms.

Data collection began after several sessions of practice, during which participants mastered the response-signal method and became familiar with the stimuli. In the data collection phase of the experiment, two participants received 190 trials per lag × discrimination level, amounting to a total of 5,700 responses, and 1 participant received 140 trials per lag × discrimination level, amounting to a total of 4,200 responses. For each trial, the computer software recorded both RT and response accuracy.

Results

Preliminary inspection of the data from this experiment revealed that participants' response times, measured from the onset of the response signal, tended to be relatively long for the shortest signal lags. Response times decrease to a minimum at intermediate lags and then increase very slightly again for the longest lags (Reed, 1976). The latter effect is small enough to be safely ignored, but the former effect can influence accounts of the data depending on how it is interpreted. One possibility is that with very short lags, participants actually continue integrating information for a longer period after the response signal, and the longer RT reflects this longer integration. Another possibility is that continuation of integration after signal onset is independent of lag but that the time it takes to actually emit the response is longer at short lags (perhaps because it takes longer to initiate a response based on weaker or less differentiated activations, as discussed above). According to the first view, information-integration time is equal to lag plus RT less a fixed offset; according to the second, integration time is equal to lag alone less a fixed offset. We do not have a theoretical stance on which of these possibilities is correct, so we analyzed the data both ways. Because the fits of both the OU and the DDV curves were somewhat better using lag + RT as the measure of integration time, we concentrate on this case, mentioning only summary results from the other case in passing.

We have plotted in Figure 8 the probability of a correct response against the mean time of occurrence of the response relative to the onset of the stimulus (lag + RT). With the experimental results, we display best fitting curves obtained using the OU function (Equation 9; solid curves) and the DDV function (Equation 12; dashed curves), assuming that processing time equals the temporal lag + RT, minus some fixed offset T_O. The data are shown with error bars reflecting the standard error of each data point, based on the binomial distribution, $SE = \sqrt{p(1-p)/N}$, where p is the probability correct and N is the number of trials contributing to the data point.

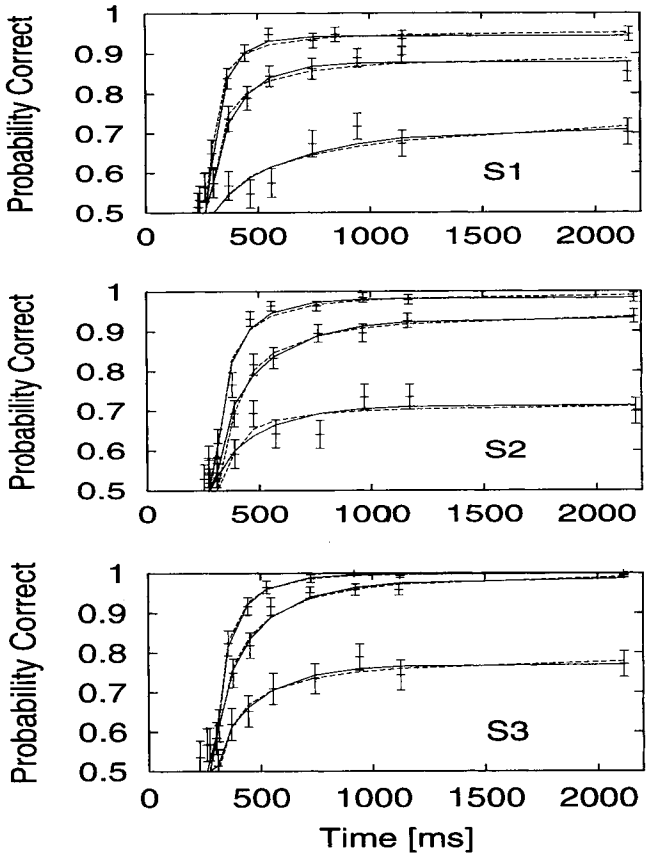

Figure 8. Time–accuracy curves for 3 participants in Experiment 1. Experiment data are represented by symbols and two optimal fits using the Ornstein–Uhlenbeck model (solid) and diffusion-with-drift-variance model (dashed) are shown for each discrimination-level condition. Error bars indicate the standard error of each data point, based on the binomial formula, as described in the text. S = subject.

Data Fits and Model Comparison

Our strategy in comparing the fit of the leaky, competing accumulator model with the DDV model makes use of the fact that the OU function provides a simple mathematical formula that gives a very good approximation to the form of the time–accuracy curves expected under the full nonlinear model. This greatly simplifies model fitting, because it is possible to derive exact shapes of time–accuracy curves immediately from given parameter values. Both the OU and the DDV time–accuracy predictions are characterized by simple mathematical formulas with three parameters (an asymptote, a rate variable, and an offset). Therefore, it is possible to perform a direct comparison between the goodness of fit of the two predictions, using standard curve-fitting methods. Also, the OU and the DDV curves have the same number of parameters, further simplifying comparison. Thus, we begin our analysis with a comparison of the fit provided by the OU and DDV functions. We then demonstrate that indeed the full, leaky, competing accumulator model can be fit to the data just as well as the OU model, using the obtained parameters from the OU fit to constrain the parameters of a final nonlinear simulation. Note that this final step is very time-consuming, because the simulation is stochastic, so

Table 1
Parameters of Fit to Individual Participant Data From Experiment 1: Nine Parameters per Participant

Level of difficulty	$\frac{P(\text{data}\|\text{OU})}{P(\text{data}\|\text{DDV})}$	Parameter					
		OU			DDV		
		d'_{asy}	$1/K$	T_O	d'_{asy}	σ^2/σ_d^2	T_O
Participant 1							
High	2.20	3.15	96	286	3.43	126	286
Medium	2.51	2.32	139	294	2.54	191	297
Low	1.54	1.11	393	337	1.47	1,175	339
Participant 2							
High	4.72	4.23	190	307	5.19	451	309
Medium	1.31	2.99	283	307	3.34	363	351
Low	1.29	1.12	209	302	1.14	120	363
Participant 3							
High	0.95	6.05	329	297	8.66	1,194	297
Medium	0.77	4.45	401	298	6.01	1,222	301
Low	1.25	1.47	205	306	1.65	298	328
All data	62.02						

Note. OU = Ornstein–Uhlenbeck; DDV = diffusion with drift variance; d'_{asy} = asymptotic value of the sensitivity statistic d'; $1/K$ = time constant for decay; T_O = time offset.

many simulation trials are required to get accurate estimates of the predicted results for a given set of parameters. The existence of the simple linear approximation becomes crucial in fitting the full model to data, because it allows us to find optimal parameters quickly and without having to rely on intrinsically noisy assessment of predicted results for given choices of parameters.

Comparison of OU versus DDV fits. In fitting and comparing the OU and DDV functions, two measures of goodness of fit have been used, providing consistent results. The first measure is the more standard least squares measure, $\Sigma_i(\text{data}_i - \text{model}_i)^2/\text{Var}_i$, where i indexes the various signal lags, and Var_i is the variance in the model's prediction for lag i. A second, more exact method consists of selecting for each model the parameters that maximize the likelihood of the data given the model, $P(\text{data} \mid \text{model})$, then comparing the likelihoods. For the special case in which the probability of the data given the model has a Gaussian form, this is equivalent to the least squares method. Because for this experiment, the data consist of counts of the number of correct trials per condition, it is possible to calculate directly the probability of obtaining n correct responses out of a sample of N trials, when the model's predicted probability correct is p, using the binomial distribution, $P(n,p,N) = [N!p^n(1-p)^{N-n}]/[n!(N-n)!]$. Because the analyses produced very similar results, we report only the results of the more exact method.[7]

For each function, optimal fits were found by doing a search within the three-dimensional parameter space of the model, to maximize either $\prod_i P(n_i,p_i,N_i)$ or $\Sigma_i(n_i/N_i - p_i)^2/p_i(1-p_i)$ (where p_i depends on the model's three parameters, and the index i runs over the data points). The optimization search was based on a metropolis algorithm, explained in Appendix E. First, we fitted the OU and the DDV functions by treating each time–accuracy curve for each participant and each discrimination level separately, with three parameters [d_{asy}, 1/rate ($1/K$ for OU and σ^2/σ_d^2 *for DDV), and the offset,* T_o] for each curve, producing nine fits, displayed in Figure 8. Both functions produce reasonably good fits, as can be seen by noticing that only a few of the data points fall more than 1 standard error from either model's predictions (by chance, as many as 32% of the points would be expected to fall more than 1 standard error from its true underlying value).

There is a small but systematic advantage for the fit obtained with the OU model (solid curves). Table 1 shows this by giving the ratio of the likelihood that each of the nine data sets (3 participants by three levels of difficulty) was generated from the OU function relative to the DDV function and the parameters for which the optimal fits were obtained.

Seven of the nine tests favor the OU function relative to the DDV function. Indeed, combining likelihood ratios across the nine conditions, we can see that the OU function is more than 60 times more likely to generate the combined experimental data relative to the DDV function (the value of 62.02 is obtained by multiplying together the likelihood ratios for all nine fits). Note also that the rate parameters are more constant across the discrimination conditions for individual participants in the OU case; in a few conditions, the optimal fit of the DDV function involves a negligible drift-variance parameter ($\sigma^2/\sigma_d^2 > 1{,}000$), corresponding to a classical diffusion process. Considering the data participant by participant, the OU function clearly provides a better fit to the data for Participants 1 and 2, but the two functions appear to be about equally good at fitting the data for Participant 3 (this is obtained, however, due to negligible drift variance, $\sigma^2/\sigma_d^2 > 1{,}000$). Similar results were obtained in the fit of accuracy versus lag only. In that case, all nine of the tests favored the OU function, although some did so only by a small margin, and overall, the OU function was 197 times more likely to have generated the combined data than the DDV function.

[7] For most p values (except where Np or $N(1-p)$ are very small), this distribution can be approximated by a Gaussian function. Thus, the least squares method can be seen as an approximation of the more exact method.

According to the assumptions of our model, the rate parameter of the information accumulation process $1/k$, and the time offset (T_o) parameter would be characteristics of the individual participant that would not vary between experimental conditions. Therefore, we examined a second way of fitting the data using only five parameters per participant: three values of d'_{asy} (one for each discrimination level) but only one rate and one time offset parameter. The optimal fits with five parameters per participant were used to calculate the likelihood of the data for both the OU function and the DDV function. These values and the corresponding optimal parameters [d'_1, d'_2, d'_3, 1/rate ($1/K$ or σ^2/σ_d^2), and T_o] are shown in Table 2. Here it is evident that the OU function provides a better fit to the data from all 3 participants, although the evidence is relatively weak in the case of Participant 3. Overall, the OU function is more than 43 times as likely to have generated that data than the DDV function. Again, similar results were obtained in the analysis of d' versus signal lag. The OU model fit the data for each participant better than the DDV model, and overall the OU model was 106 times more likely to have generated the data.

Applying the full model to experimental data. To this point, we have established that the OU-shaped time–accuracy curves predicted by the leaky, competing accumulator model provide a good characterization of the time–accuracy data from the reported experiment. Even though we had previously established that the OU-shaped curves were also produced by the full, nonlinear version of our model, it might be useful to verify that the full model could provide an adequate fit to actual experimental data.

The simulation is based on the following assumptions: Accumulators for the two responses are initialized to 0 at the beginning of each trial, then integration begins according to Equation 11. Sensory processes impose a fixed delay T_s in the onset of the information accumulation process. At some time, a response signal is presented. Integration may continue for some period after the onset of the response signal, at which point the accumulator with the largest activation is chosen. The response is emitted at this point and occurs after another delay T_r independent of signal lag. Given these assumptions, the integration time for a response occurring at the actual time T_a will be equal to $T_a - (T_r + T_s)$. Because they are not separately identifiable, $T_r + T_s$ collapse into a single variable equivalent to T_o in the OU fits.

On the basis of this model, we expect a stochastic simulation following Equation 11 to lead to a very close approximation to the data, when shifted by the fitted value of the parameter T_o. For a given participant, we assume fixed values of k, β, and T_o, but we allow different values of ν for each of the three different degrees of difference between the lengths of the sides of the rectangles. To show that this is indeed the case, we used stochastic simulations to fit the data from Participant 2.

To establish values for the parameters of these simulations, we began with arbitrary choices for $\tau = 100$ ms, $dt = 10$ ms (thus the time scale, τ, is on the order of magnitude of the process measured, i.e., 223 ms for S2 in Table 2, while the Euler integration step is 1/10 of it), and $\sigma = 1.58$ (the standard deviation of the random normal deviates ξ_1 and ξ_2). For each discriminability condition, we used the best fitting value of d' from Table 2 and solved Equation 10 to find the corresponding value of ν. The values of the parameters k and β are underdetermined by the fits, because the parameter K of the OU process is equal to $k - \beta$. Thus, if indeed the approximation is valid, we can choose arbitrarily a value of β and then calculate the corresponding value of k to achieve the desired value of K. In accordance with the assumptions of the model, we chose a value of β that was greater than 0 and large enough so that the activation of the accumulator that loses the competition tends to reach 0 if the simulation is allowed to run to asymptote. This then constrained the value of k to be equal to the chosen value of β plus the fitted value of K from Table 2.

On the basis of these assumptions, the simulation was run by initializing each accumulator to $x_i = 0$ for each trial, then beginning the stochastic information accumulation process at $t = T_o$ and continuing until a time equal to the lag + RT for that condition. For each of the three discriminability conditions, 4,000 trials of the simulation were carried out.

The time–accuracy curves obtained in the simulations (solid, squiggly curves) as well as the time–accuracy curves based on the best fitting OU curves (dashed curves) are shown in Figure 9, with the experimental data. The OU and the nonlinear fits are very similar, as expected on the basis of our earlier analysis. The variation of the stochastic simulation around the OU curve is expected because the simulation is the result of a probabilistic sample and so is subject to fluctuations around its expected value. In the limit of a large sample size, the simulation is expected to converge to the OU curves to the point of indistinguishability.

Table 2
Parameters of Model Fits to Individual Part Data From Experiment 1: Five Parameters per Participant

		Parameter									
		OU					DDV				
Participant	$\frac{P(\text{data}\|\text{OU})}{P(\text{data}\|\text{DDV})}$	d'_1	d'_2	d'_3	$1/K$	T_O	d'_1	d'_2	d'_3	σ^2/σ_d^2	T_O
1	6.25	3.34	2.28	0.80	135	284	3.73	2.54	0.90	211	287
2	4.76	4.41	2.82	1.14	223	307	5.34	3.43	1.38	503	308
3	1.45	6.04	4.13	1.60	325	297	8.41	5.74	2.19	1,101	298
All	43.14										

Note. OU = Ornstein–Uhlenbeck; DDV = diffusion with drift variance; $1/K$ = time constant for decay; T_O = time offset.

Discussion: Leakage and Drift Variance in Two-Alternative Time–Accuracy Experiments

In summary, it appears that the OU function provides a slightly better fit to the shapes of empirical time–accuracy functions than the DDV function. However, note that the proposals of Ratcliff (1988) include the assumption that participants may use decision boundaries that sometimes terminate processing before the response signal occurs. With these response boundaries in place, the full DDV model produces curves that are closer to the exponential form that characterizes the experimental data. Consequently, it is still possible that drift variance rather than leakage governs the approach to asymptote at least in some time–accuracy experiments, even when the time–accuracy curve is approximately exponential. On the other hand, our analysis shows that there is no need to invoke drift variance in accounts of the shapes of time–accuracy curves. Indeed, it appears that leakage of information, which imposes a limit on the temporal window of information integration, provides a good account of the shapes of time–accuracy curves without invoking additional factors such as the use of decision boundaries in time-controlled experiments.

It seems likely that both leakage and drift variance are factors limiting performance in time–accuracy experiments. The relative magnitudes of the leakage factor that operates in our model and of the drift-variance factor of the DDV model are likely to be task dependent. It might be suggested, for example, that the drift variance would have a material impact in memory experiments of the sort Ratcliff (1978) originally addressed, where each trial corresponds to a different item. In this case, differences in the strength of items in memory would naturally lead to variations in drift within a given experimental condition.[8]

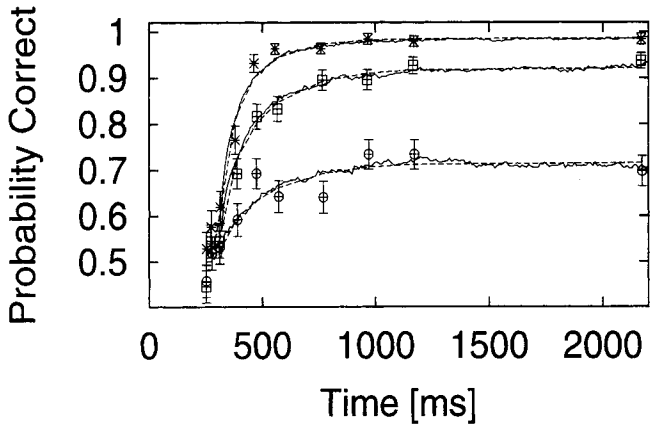

Figure 9. Nonlinear simulation fit (solid lines) and Ornstein–Uhlenbeck (OU) fit (dashed lines) for the time–accuracy data in Experiment 1 (with symbols); 4,000 Monte Carlo simulation trials in each discriminability condition. The parameters correspond to the five-parameter OU fit (Participant 2) in Table 2 ($d'_{asy1} = 4.41$, $d'_{asy2} = 2.82$, $d'_{asy3} = 1.14$, $\tau/K = 223$, $T_O = 307$). The simulation parameters are calculated according to the OU linear approximation for the process described in the text ($T_O = 307$, $\sigma = .16$, $K = 0.44843$ [$k = 0.79843$, $\beta = .35$]). The drift values, ν, in the three discriminability conditions are .2333, .152, and .0603 (for the high, medium, and low conditions, respectively). Error bars represent the standard error of each data point, as discussed in the text.

Information-Controlled Reaction Time Tasks With Two Alternatives

In this section, we analyze our model's behavior in standard two-choice RT tasks. Modeling these tasks differs from modeling time-controlled experiments, in that participants have to determine themselves the moment when a response is generated. In most models, this is done by assuming an activation-based response criterion that is subject to strategic control. In accordance with this scheme, we assume that responses are generated when either of the accumulators reaches some preset criterion value.

The first point to note is that a special case of our model, in which there is neither leakage nor inhibition (that is, in which $k = \beta = 0$), is equivalent to a continuous-time version of what we call the *classical* accumulator model (Vickers, 1970, 1979; Vickers et al., 1971; Wilding, 1974; see also Audley & Pike, 1965; LaBerge, 1962, for previous versions of this model; and see Van Zandt, 2000; Van Zandt, Colonius, & Proctor, 2000, for a recent version based on a race of Poisson processes).[9] According to this class of models, choice RT responses depend on a race between two independent accumulators. There is an absolute response criterion for each accumulator, and as soon as either accumulator reaches its criterion, the response is triggered, regardless of the state of the other accumulator(s) at that time. This is equivalent to the assumption of our model.

The classical accumulator model has been the subject of intense analysis and comparison with the family of random walk models (Ashby, 1983; Laming, 1968; Link & Heath, 1975; Stone, 1960) or its diffusion variants. Unlike the accumulator model, in which there are two separate cumulative values, the random walk–diffusion model is defined as a process involving a single diffusion variable, representing the difference in the amount of evidence for each of the two alternatives. In this case, the decision to respond is made when the value of the diffusion variable reaches either of two decision boundaries, each representing a certain level of relative evidence for one of the alternatives compared with the other. This difference between the two approaches has major consequences for the relationship between RT and the discriminability of presented stimuli. In part for this reason, a major focus of earlier research on choice RT models was an examination of the ability of each of the approaches to account for the effects of stimulus discriminability on response accuracy and RTs for both correct and

[8] Introducing a Gaussian distribution of drifts (Ratcliff, 1978) into the OU model leads to $d'(t) = \nu(t)/\sqrt{\sigma_d \nu(t)^2 + \sigma^2(t)}$, where $\nu(t)$ and $\sigma(t)$ are given by Equation 8.

[9] Several related models have been labeled as *accumulator* models. For example, Audley and Pike (1965) used this label to refer to a countermodel, where at discrete moments in time a unit is added, either to the counter representing the first-choice alternative (with probability p) or to the counter corresponding to the second-choice alternative (with probability $1 - p$) and a decision is reached when one of the counters reaches an absolute criterion. Later, Vickers (1970) used the same label for a model where, in addition to the assumptions described above, the amount of activation that is added to the counters is not a constant but rather a stochastic Gaussian distributed variable (of variance, say, δ). A discrete-time accumulator model with probability p and noise variance δ is equivalent (in the limit of small time steps) to a two-dimensional diffusion process (as in Equation 5), with drifts $\rho_1 = p$, $\rho_2 = 1 - p$, and a noise standard deviation of $\delta + p(1 - p)$.

incorrect responses. As we will see, neither of the classical models was fully adequate as originally formulated, but a more recent implementation of the DDV model (Ratcliff & Rouder, 1998) that incorporates variability from trial to trial in the starting point of drift (as previously proposed by Laming, 1968, and by Ratcliff, 1981), in addition to variance in the direction of drift itself, has been able to account quite well for data on the effects of discriminability on response probability and latency.

In this section, we consider the role of leakage and lateral inhibition in accounting for data from studies of the effect of discriminability on RT. We see that lateral inhibition provides a mechanism that allows sensitivity to relative evidence in conjunction with the use of a simple absolute criterion for response initiation. For some participants in some experiments, it appears that an additional mechanism is necessary to account for the fact that error responses to highly discriminable stimuli may sometimes be faster than correct responses. Starting point variability provides one possible mechanism that can be used to account for such effects.

Effects of Stimulus Discriminability on Reaction Time

For concreteness, we consider a typical binary task that has been used in the literature on the effect of discriminability on RT (Vickers, 1970): On each trial, the stimulus was a pair of vertical line segments, and the participant was required to indicate with a button press which of the two line segments was longer. In this case, discriminability was manipulated by varying the length difference between the two line segments. One could tabulate the probability of choosing, say, the right line segment as a function of its length minus the length of the left line segment. This difference variable would be thought of as being at least monotonically related to the extent to which the input favors the right over the left response. One could then examine the mean RT (and other statistics) associated with the "right" responses in each of the length-difference conditions. A separate analysis could be performed for the left responses as a function of its length minus the length of the right line segment. However, in what follows, we follow typical procedure in this literature and combine the results of these two analyses. Response probabilities and RT statistics are thus averages over the responses to the two stimuli representing the same amount of evidence for each. In these combined analyses, negative length differences always correspond to errors, and positive length differences always correspond to correct responses. In the following, the curves showing response latency as a function of a discriminability parameter are called *latency-discriminability* (LD) functions. Such functions are similar to what are usually called *latency-probability* (LP) functions (Audley & Pike, 1965; Ratcliff & Rouder, 1998; Vickers et al., 1971), in which latency is plotted against the probability of making a given response, because the response probability is a monotonic function of the discriminability parameter.

Previous Models

The classical random walk model and the classical accumulator model differ in their predictions for the shapes of LD functions. Although it seems counterintuitive at first, the classical random walk model and the classical diffusion model with Gaussian noise in the diffusion process predict that the LD functions will be symmetrical when the two boundaries of the decision process are equidistant from the starting point of the process. The RT is slowest for the most ambiguous stimuli and gets faster toward the extremes, with both the high- and the low-probability responses occurring more quickly (see Figure 10). On the other hand, the accumulator models predict that the mean RT will increase monotonically as the probability goes down (Audley & Pike, 1965; Vickers et al., 1971).[10] In the case where continuous, Gaussian increments are added to both accumulators on each time step, this same general pattern holds, but with a small downturn for the responses with the very lowest probabilities (Vickers et al., 1971).

These observations show that the choice between the accumulators and the random-walk models can have empirically testable consequences. The contrasting predictions of these two models have led to a series of articles that examine LD functions in binary-choice RT experiments (Audley & Mercer, 1968; Jamieson & Petrusic, 1977; Petrusic & Jamieson, 1978; Pike 1968; Vickers, 1970; Vickers et al., 1971; Wilding, 1974), and recently Ratcliff and Rouder (1998) and Ratcliff et al. (1999) have reopened this line of investigation. In general, these studies report asymmetric LD functions that increase as the response probability decreases down to a fairly low value and then turn downward. Representative data from a relevant experiment are shown in Figures 11–14 from Vickers et al. (1971). The LD functions are shown in Figure 12. The pattern shown by Participant 4 in this figure is representative of data reported by Audley and Mercer (1968), Vickers et al., and Wilding (1974). An additional observation (Vickers et al., 1971) is that the LD curves have a more asymmetric characteristic for the slower participants, so that RTs tend to increase as probability decreases until turning down again at the very lowest probabilities. The degree of symmetry is higher for faster participants. Participant 4 of Ratcliff et al. (1999) appears to represent the extreme case of this generalization, in that this participant was by far the fastest overall in their experiment and produced a highly symmet-

[10] The symmetric form of LD functions under the random-walk model can be understood as follows. Consider a random walk initialized at the origin. On every time step, the process moves either to the right (with probability p) or to the left (with probability $q = 1 - p$). The process is terminated when it has progressed N steps, either to the right (A) or to the left (B). Audley and Pike (1965) have observed that the probability of reaching A after exactly N time steps is p^N, whereas the probability of reaching B in N steps is q^N. Thus, the ratio of these probabilities is $[P(\mathrm{A}|N)/P(\mathrm{B}|N] = (p/q)^N$. For any larger number $n > N$ of steps to reach either boundary, exactly half of the additional steps will have been toward A, and half will have been toward B. As a result, the probability of these additional steps will be the same in both cases, and so the ratios of the probabilities, $P(\mathrm{A}|n)/P(\mathrm{B}|n)$, will remain constant at $(p/q)^N$. Thus, the two RT distributions differ only by a factor that reflects the fraction of correct responses, and the probability of reaching either boundary in n steps, relative to the overall probability of reaching that boundary, is the same for both boundaries. In the discrete accumulator model, at every time step, a unit is added either to a left counter or to a right counter. Thus, the ratio of the conditional probabilities given that the process has finished after n steps is $P(\mathrm{A}|n)/P(\mathrm{B}|n) = (p/q)^{N-n}$. Assuming $p > q$, the relative probability for a correct response decreases with n, and thus errors tend to be on average slower than correct responses. This effect becomes more and more pronounced as p/q grows larger, so that responses become slower and slower as they become less and less probable.

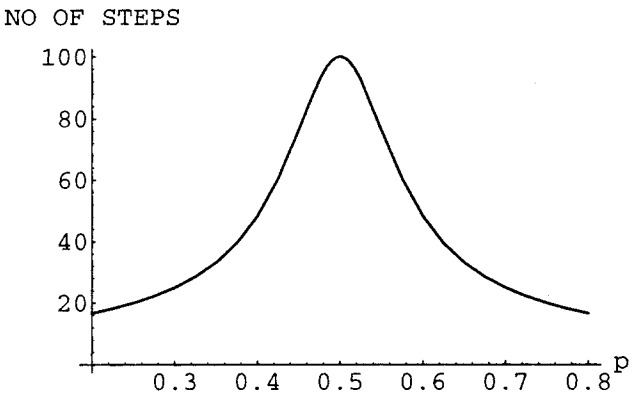

Figure 10. Mean number of steps to reach the upper boundary at $N = 10$, as a function of the random-walk probability, p. The mean number of steps, calculated following Cox and Miller (1965), is $N(p^N - q^N)/[(p - q)(p^N + q^N)]$, where $q = 1 - p$.

ric LD curve. Symmetry appears to increase as RT decreases with practice (Petrusic & Jamieson, 1978).

The LD functions from some very fast participants may be consistent with the predictions of the classical diffusion model, and the LD functions of some very slow participants may sometimes be consistent with accumulator models using Gaussian increments. However, the full range of shapes of participants' LD functions cannot easily be encompassed by either of the classical models. The classical random-walk and diffusion models predict that LD functions will always be symmetric, whereas the classical accumulator model predicts LD functions that increase toward the low response probabilities, with only a small downturn at low probabilities (when Gaussian noise is added to the accumulators; Vickers et al., 1971). The pattern of highly asymmetric and U-shaped curves exhibited by most participants seems inconsistent with both approaches, and the range of variation of shape also seems difficult to encompass by either approach alone.

Modifications can be introduced into these classical models to produce U-shaped but asymmetric LD functions. The addition of drift variance into the classical diffusion model, producing Ratcliff's (1978) DDV model, can significantly increase the error RTs, leading to LD functions corresponding to the experimental data (Ratcliff & Rouder, 1998).[11] Asymmetric LD functions can also be obtained assuming asymmetric drift rates (Ratcliff, 1985), non-Gaussian noise (Link & Heath, 1975), or bias in the decision process (Ashby, 1983).

Starting point variability is another factor that has been considered in stochastic information accumulation models (Laming, 1968). In a recent article, Ratcliff and Rouder (1998; see also Ratcliff et al., 1999) have used a diffusion model that includes both drift variance and starting point variance and provides good fits for a variety of LD profiles. Starting point variability appears to be necessary to account for data from some participants, in which error responses to very highly discriminable stimuli can be faster, on average, than the correct responses to these same stimuli: The effect of starting point variability is to make it possible for the diffusion variable to be very close to the incorrect response threshold at the start of the information accumulation process, thereby allowing a few very fast incorrect responses. Such very fast responses will occur in all discriminability conditions, but they will exert their greatest effects on the mean RT only in those conditions that otherwise produce very few incorrect responses.

Fitting the Leaky, Competing Accumulator Model

Our effort has focused on an investigation of how well the leaky, competing accumulator model can account for the patterns of data seen in LP experiments. As a starting place for this analysis, we examine how well this model can do without incorporation of trial-to-trial variability either in drift or in the starting point of the information accumulation process.

The data we have fitted with our model were reported by Vickers (1970). In this experiment, participants performed choice RT ("respond as accurately and quickly as possible") in a length-discrimination task involving two vertical line segments. The left segment was longer on half the trials, and the right segment was longer on the other half; this manipulation was crossed with five levels of discriminability involving length differences of 0.3, 0.5, 1.0, 1.5, and 2.5 mm, and trials of all 10 resulting types were randomly intermixed within each block of trials. Response probabilities and various moments of the RT distributions (mean, standard deviation, skewness, and kurtosis)[12] were calculated for the various length-difference conditions, combining the left and right responses for each length-difference condition in which *length difference* refers to the difference in length between the correct and the incorrect choice, so that responses associated with negative differences constitute errors. Five participants performed the experiment, each one producing approximately 500 trials per discriminability condition. On the basis of this large number of responses, each participant provided 50 data statistics (accuracy and 4 RT moments, in five discrimination conditions, for correct and incorrect responses). Those are displayed with symbols in Figures 11–14.

We modeled this task using Equation 5 and truncating negative activations after each unit was updated. In keeping with our earlier formulation of the model, we assumed that accumulators are initialized to 0 at the beginning of each trial and a response is chosen when either accumulator reaches the response criterion θ. The actual RT simulated is the number of cycles to criterion times a fixed 10 ms per cycle, plus an offset T_o, corresponding to residual processes. To map the length-difference condition to the drift ν of the diffusion process, we assumed here a monotonic

[11] This arises from the fact that a mixture of different drifts is generated across trials within each experimental condition. Consider the probability of the error response associated with the upper decision boundary when the drift is negative. The probability of such an error decreases as the drift becomes more negative while the mean RT for such errors paradoxically decreases (as previously seen in Figure 10). When drift variance is introduced, the less negative drifts result in a higher proportion of errors than the more negative drifts, and the RTs of errors in the less negative cases are longer. The less negative trials thus contribute more to the error RT distributions, increasing mean RT. For correct responses, on the other hand, the averaging arising from drift variance leads to a faster mean RT compared with the classical process with the same drift, for similar reasons.

[12] We thank Douglas Vickers for providing these data. The measure of skewness is $\gamma_3/\gamma_2^{3/2} = \gamma_3/SD^3$, and the measure of kurtosis is $(\gamma_4/\gamma_2^2) - 3$, where γ_i is the ith moment about the mean.

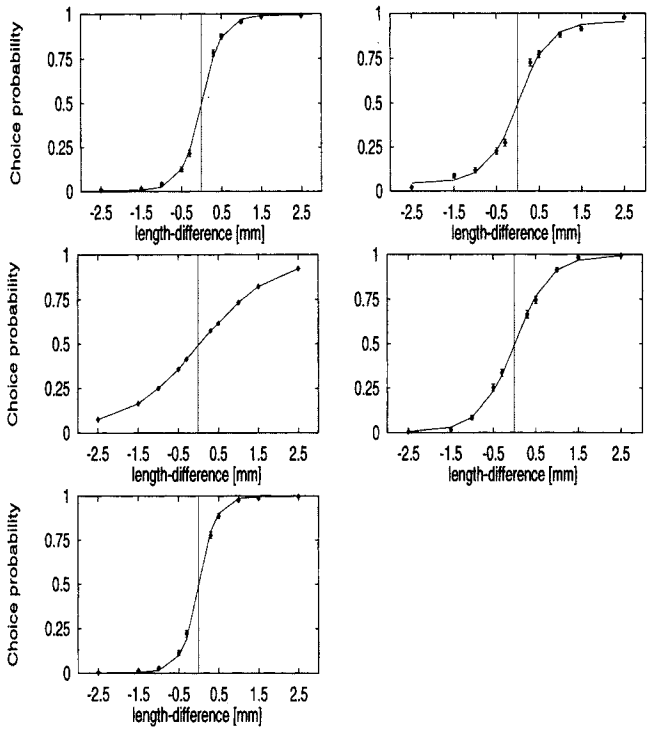

Figure 11. Choice probabilities from the latency-discriminability experiment of Vickers (1970). Symbols with error bars represent the experimental data, and the curves show fits of the leaky, competing accumulator model. Responses are collapsed over the two alternatives: Responses to positive length differences correspond to correct responses, and responses to negative length differences correspond to errors. Data (left to right, top to bottom) are for Participants 1–5, respectively. Raw data were provided by D. Vickers (personal communication, September 1997) and are presented here with his permission.

sigmoidal relation given by the cumulative Gaussian integrate, $v = Erf(\epsilon L)$ (see also Vickers, 1970, for further considerations), where L is the length difference and ϵ is a participant-dependent coefficient representing the accuracy of spatial encoding. We used five positive and five negative values of L corresponding to the values used in the experiment, and we recorded the occurrence and RT of cases in which Unit 1 reaches the criterion first, corresponding to selection of the first alternative. Note that when the value of L is positive, such cases correspond to correct responses and when L is negative, they correspond to errors.

Note that the time step used for mapping simulation steps to real time (10 ms in our simulations, corresponding to the time scale, τ, of 100 ms) is arbitrary. However, once fixed, the time constant obtained from the free parameter k is relative to it. Thus, once the time step is set, the model is specified by six parameters. For the information accumulation process itself, we have the leakage parameter k, the lateral inhibition parameter β, and the standard deviation of the noise σ. In addition, we have the criterion θ, the time offset T_0, and the encoding coefficient ϵ.

For each participant, the six parameters were varied to capture that participant's data. To find parameters that provided a good fit to each participant, a metropolis optimization program was run (see Appendix E), minimizing a cost function that included factors for the discrepancy between the accuracy, mean, and standard deviation of RT of the data and those of the model. No attempt was made to fit the model to the skew and kurtosis of the data. Because the parameter space was large (six-dimensional) and assessing predicted values required stochastic simulations of all length conditions for every choice of parameter set, we were able to run only 8,000 swaps (as defined in Appendix E) of the metropolis algorithm per participant. As a result, there was some noise in the fitting process, so the fits obtained are unlikely to be optimal. The final parameters that were used to fit the data of the five participants are provided in Table 3. Using those parameters, each discriminability condition was simulated by performing a final assessment block of 10,000 simulation trials per discriminability condition.

Comparison of Fit to Data

As shown in Figures 11, 12, 13, and 14, the model provides a good fit to the accuracy data and the mean and standard deviation of the RT for most participants and also captures the trends apparent in the data on the skewness and kurtosis of the RT distributions. The model does not fit the LD functions of two of the participants (1 and 5). However, both of these LD functions are quite unusual, in that both show a dip in the LD function for moderately discriminable stimuli (moderately negative length differences), contrary to the predictions of all models that we know of. These data are anomalous, and we have no explanation for them. Otherwise, there are no large or systematic deviations in the fit to the accuracy, mean, and standard deviation data. In particular, the model reproduces closely the sigmoidal accuracy functions, and the nonmonotonic mean RT function for Participant 4 with the much flatter RT function for Participants 2 and 3. In addition, the fits reproduce the pattern of longer RT for incorrect than correct responses seen in Participants 1, 3, 4, and 5 and the nonmonotonic, inverse-U-shaped functions for the standard deviations seen in Participants 4 and 5.

The skewness and the kurtosis functions seen in the behavioral data are quite noisy. This is as expected, because these statistics are extremely sensitive to outliers, as Ratcliff (1979) has shown. No error bars are available for these data, and it is likely that individual data points could change a great deal if only one or a few trials were removed from the calculation of the statistic. We consider this data nevertheless because there are fairly clear increasing trends in both statistics at least in some participants (1, 4, and 5). Similar, monotonically increasing skewness and kurtosis functions have also been reported in two other experiments (Vickers et al., 1971; Wilding, 1974). Interestingly, these trends are also apparent in the fits provided by the model, even though the parameter-fitting process did not explicitly consider these variables in assessing the goodness of fit.

By inspecting the fitted parameter values for each participant, one can see how each of the parameters affects the LD functions. For example, it seems that the inhibition parameter increases the U-shapedness of the mean-RT function (Participant 3, whose mean function is perhaps the least U-shaped, has the lowest value of the inhibition parameter, corresponding to a classical accumulator, whereas Participant 4, whose mean function is the most U-shaped, has the largest value for the inhibition parameter). In addition, the leakage and the inhibition parameters increase the value of the

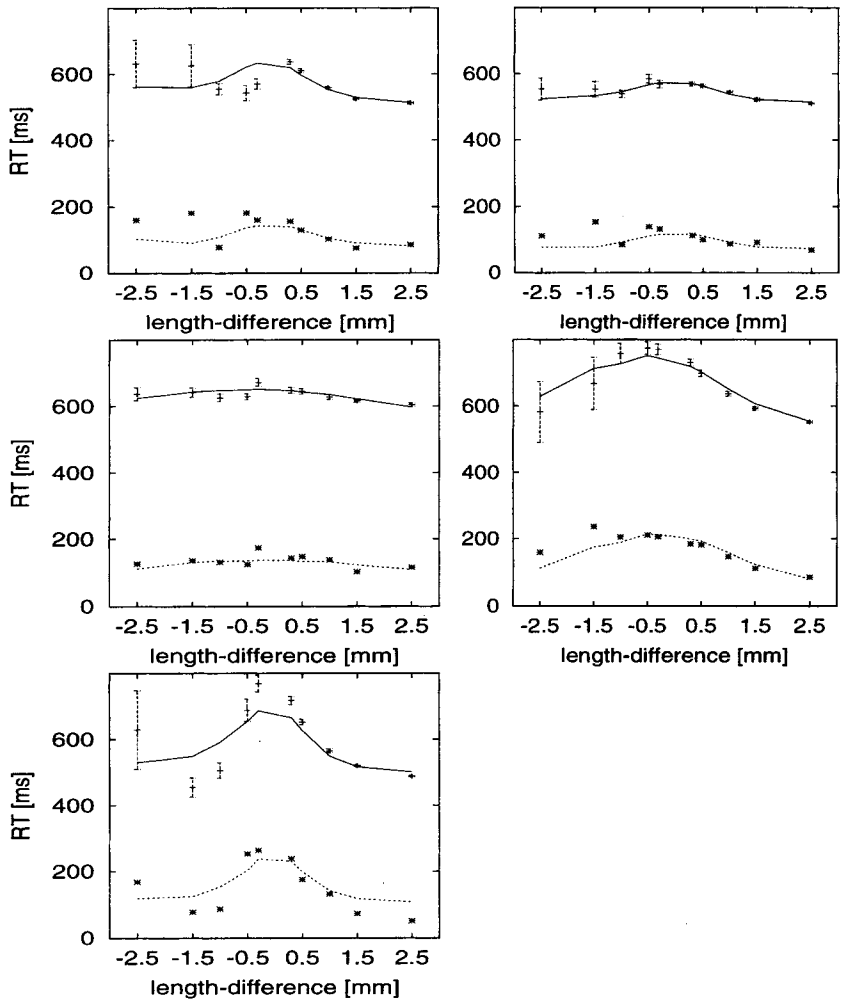

Figure 12. Mean (crosses, solid curves) and standard deviation (stars, dashed curves) of reaction times (RTs) from the latency-discriminability experiment of Vickers (1970). Symbols with error bars represent the experimental data, and the curves show fits of the leaky, competing accumulator model. Responses are collapsed over the two alternatives: Responses to positive length differences correspond to correct responses, and responses to negative length differences correspond to errors. Data (left to right, top to bottom) are for Participants 1–5, respectively. Raw data were provided by D. Vickers (personal communication, September 1997) and are presented here with his permission.

skewness and kurtosis (without leakage and inhibition, we obtained very small skewness values, scattered around 0, whereas in the presence of leakage and inhibition, we obtained skew values typically between 1 and 2, as in the data, with increasing trends).[13]

To better understand how the model parameters affect the shapes of the LD or LP functions, we performed a set of simulations in which, starting from the classical accumulator model with Gaussian increments (Vickers et al., 1971), we modified one parameter at a time. In particular, we examined the effects of introducing lateral inhibition, leakage, and starting point variability (see Figure 15). Although the latter variable was not used in our fits, we considered its effects because it has been useful in fitting data sets when many of the error responses are very fast, as previously discussed (Laming, 1968; Ratcliff & Rouder 1998; Ratcliff et al., 1999). Each of these simulations was performed for three values of the response threshold (chosen so as to approximately equate the mean RT for the corresponding conditions across the four panels), to see how speed–accuracy tradeoffs are reflected in the vari-

[13] Although the parameters used in Table 3 are not thought to represent the best possible fits to the data, it might be instructive to compare the predicted time constants with those found in optimizing the OU time–accuracy curves in the fits to the data from Experiment 1. Whereas the time-controlled optimal fits produced rate parameters within the range of 100 to 400 ms (Table 1), the rate parameters that could be computed from the information-controlled fits ($\tau = 10/K = 10/|k - \beta|$) were 2,000, 312, 76, and 714 ms for Participants 1, 2, 4, and 5, respectively. (Participant 3 had balanced inhibition and leakage resulting in $K = 0$.) Although the range here was larger, note that there was some degree of overlap, despite the fact that the fits involved different groups of participants performing different experiments.

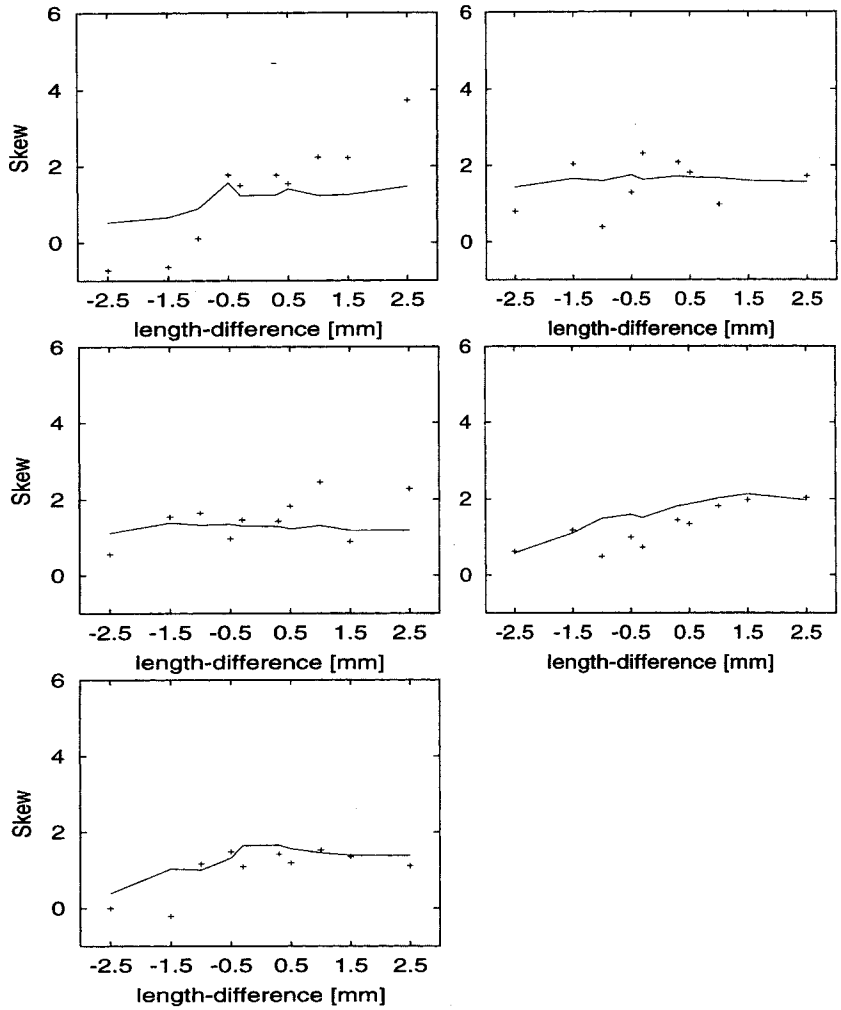

Figure 13. Skewness of reaction time distribution from the latency-discriminability experiment of Vickers (1970). Symbols with error bars represent the experimental data, and the curves show fits of the leaky, competing accumulator model. Responses are collapsed over the two alternatives: Responses to positive length differences correspond to correct responses, and responses to negative length differences correspond to errors. Data (left to right, top to bottom) are for Participants 1–5, respectively. Raw data were provided by D. Vickers (personal communication, September 1997) and are presented here with his permission.

ous models. Two additional simulations (shown in Figure 16) were performed to examine the characteristics of combining the factors considered separately in Figure 15. The first examined the combined effects of lateral inhibition and leakage, and the second examined the combined effects of these variables plus starting-point variability. Once again, the simulation was performed at three different values of the response threshold.

For the classical accumulator (upper left in Figure 15), we observed that RTs increased toward low probability (i.e., errors were slower than correct responses) but with a small downturn at the lowest probability. This downturn was due to the Gaussian increments, as previously shown by Vickers et al. (1971). The effect of lateral inhibition (upper right) was to make the functions more curved (or less flat) and to increase the downturn (see, in particular, the conditions with medium and faster RTs) while leakage (lower left) affected the shape of the LP functions to a lesser degree (although slight modifications were visible). The strong impact of the lateral inhibition can be understood intuitively, because, as discussed earlier, it provides a way to interpolate between the accumulator models that generate more monotonic LP functions with error responses slower than correct responses (Audley & Pike, 1965) and random-walk models that generate strongly peaked and symmetrical LP functions (as shown in Figure 10). One can also see that with both leakage and lateral inhibition (left panel in Figure 16), LP functions remain asymmetrical, with error of responses slower than correct responses. One can also observe that in all the models, faster responses (induced by lower response criteria, at the expense of accuracy) result in flatter and somewhat more symmetrical LP functions, as previously noticed by Vickers et al. (1971) and by Ratcliff and Rouder (1998).

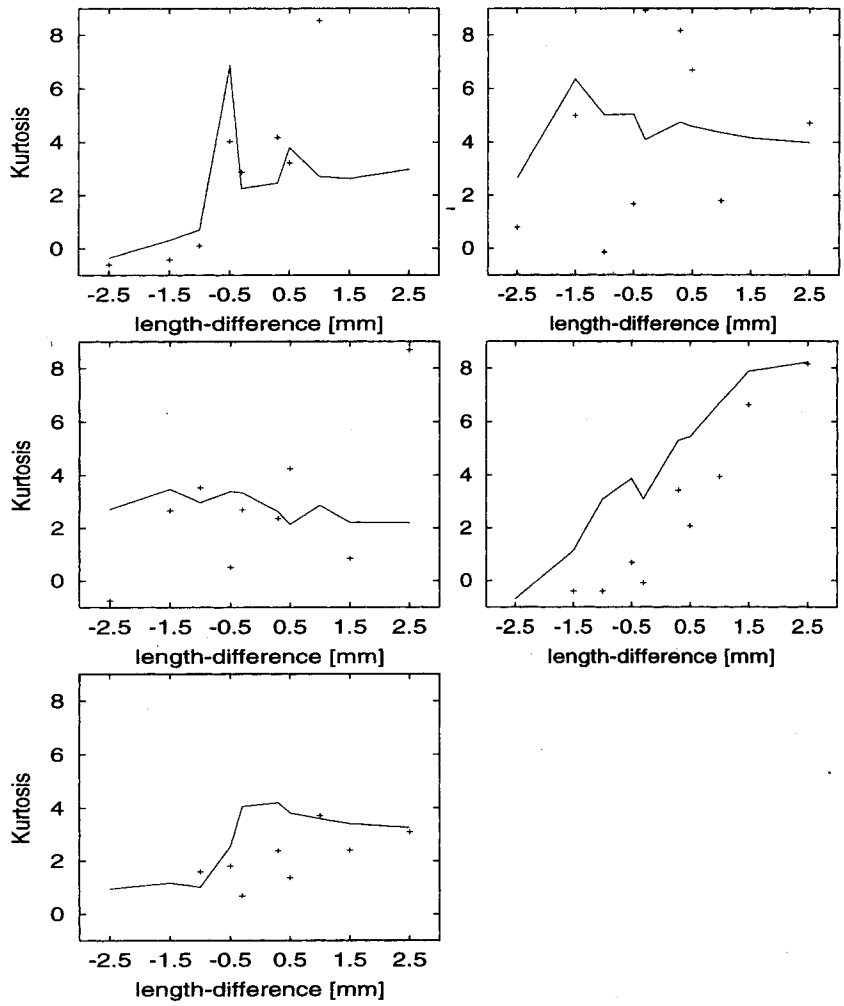

Figure 14. Kurtosis of reaction time distribution from the latency-discriminability experiment of Vickers (1970). Symbols with error bars represent the experimental data, and the curves show fits of the leaky, competing accumulator model. Responses are collapsed over the two alternatives: Responses to positive length differences correspond to correct responses, and responses to negative length differences correspond to errors. Data (left to right, top to bottom) are for Participants 1–5, respectively. Raw data were provided by D. Vickers (personal communication, September 1997) and are presented here with his permission.

The effect of starting-point variability is to pull down the left (error) end of the LP function, relative to the right (correct response) end. This is as expected based on the work of Laming (1968), Ratcliff and Rouder (1998); and Ratcliff et al. (1999). When added to the classical accumulator (compare upper left with lower right in Figure 15), the result is fairly symmetric LP functions, or, with extreme values of starting-point variability (not shown), a pattern in which errors are generally faster than correct responses across all difficulty levels. In the leaky, competing accumulator model, moderate values of starting-point variability (not shown) produce relatively symmetric LP functions, as with the accumulator model. A large value of starting-point variability (as in the case shown in Figure 16; see caption for details) can result in a crossover effect, so that errors to difficult stimuli are slower than correct responses, but errors to the easiest stimuli are faster than correct responses. This pattern is seen in the data of some participants in Ratcliff and Rouder (1998), Ratcliff et al. (1999), and Smith and Vickers (1988). The crossover effect is apparent for the highest threshold, which results in long overall RTs, in line with

Table 3
Parameter Values for Fits to Individual Participant Data From Vickers (1970)

Participant	θ	σ	β	k	ϵ	T_O
1	1.99	.520	0.10	.05	.74	300
2	1.14	.560	0.46	.14	.73	384
3	2.00	.553	0.05	.05	.23	340
4	1.20	.360	1.41	.11	.33	379
5	2.05	.577	0.24	.10	.78	259

Note. T_O = time offset.

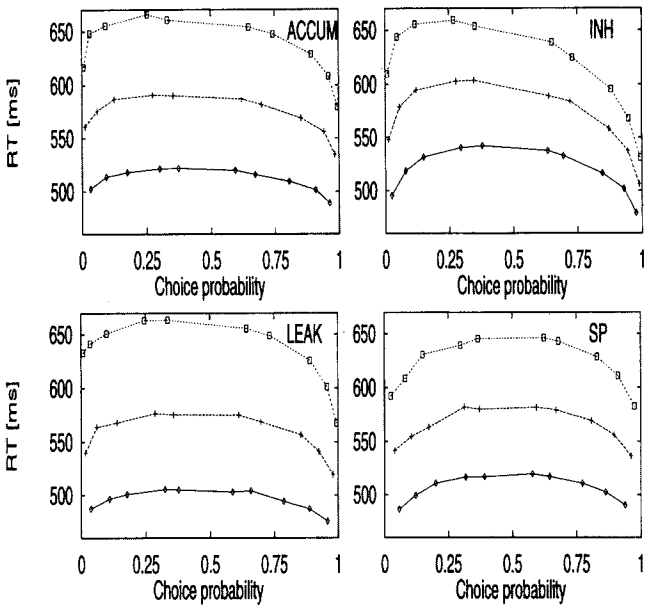

Figure 15. Latency-probability (LP) functions (mean latency as a function of the response probability) for the classical accumulator (ACCUM; $\theta = \{1, 1.2, 1.8\}$, $\beta = 0$, $k = 0$), the accumulator with lateral inhibition (INH; $\theta = [0.8, 1, 1.2]$, $\beta = .9$, $k = 0$), the accumulator with leakage (LEAK; $\theta = \{0.8, 1.1, 1.4\}$, $\beta = 0$, $k = .2$), and the accumulator with starting point variability (SP; $\theta = \{1.5, 2.1, 2.7\}$, $\beta = 0$, $k = 0$; the starting point of each accumulator was distributed uniformly between 0 and ⅔ of θ), for three levels of the response criterion. Response criteria were chosen so as to obtain LP functions within the same range of mean reaction time (RT) for each model. All other parameters were as for Participant 4 in Table 3.

effects of speed versus accuracy instructions in Ratcliff and Rouder.

Experiment 2: Reaction Time Distributions

Although we have fit the leaky, competing accumulator model to the effects of discriminability on the accuracy and the various moments of RT (means, standard deviations, skew, and kurtosis), we have not shown that the model can fit the exact shape of the RT distributions (for the data of Vickers, 1970, only the RT moment statistics were available). A motivation for considering the RT distributions in some detail comes from the fact that existing stochastic information accumulation models differ in their predictions for the shapes of these distributions. The DDV provides very good fits to such distributions, whereas Wilding (1974) has reported that the classical accumulator model predicts distributions that fail to exhibit the long tail to the right that is characteristic of most RT distributions. The question arises whether the leaky, competing accumulator model is more like the diffusion model or more like the accumulator model in this regard. Although it might seem that our model's fit to the skew statistic might address this, higher RT moments are noisy and may be oversensitive to individual data points (Ratcliff, 1979) and may fail to capture all aspects of the shape of RT distributions (R. Ratcliff, September personal communication, 1999). For this reason, an explicit test for our model's ability to fit RT distributions was required. To provide data for this, we performed a new experiment where participants were required to perform a difficult-choice RT task with stimuli similar to those used in Experiment 1.

Method

Materials and apparatus. Participants were seated in a dark room, in front of a computer monitor where stimuli were presented. Stimuli were bright rectangles, with nearly equal sides, on a dark background, tilted by 45°, so that they appeared to be of a diamond shape. One of the sides was always 200 pixels large, whereas the other was smaller by either, 1, 2, or 3 pixels. The direction in which the larger side was presented, as well as the level of discriminability, was randomized across the trials.

Procedure. At the beginning of each trial, a fixation point was presented at the center of the display. The rectangle was presented 500 ms later, centered at fixation, and remained on the screen until the occurrence of the response. A two-button response box was used, and participants were instructed to respond by pressing the left (or right) button to indicate the judgment that the rectangle was tilted toward the upper left (or upper right). Both the accuracy and speed were emphasized ("try to respond accurately but as fast as you can"). Two seconds after their response, the next trial was presented. Trials were presented in blocks of 96 (32 trials at each level of discriminability). Participants could rest in between the blocks before continuing with the next block. A session consisted of 1 practice block followed by 10 experimental blocks (and took about 1 hr to perform). A 5-min break was enforced after 6 blocks. Each participant performed three sessions, generating a total of 960 experimental responses at each level of discriminability.

Participants. Two participants with normal or corrected-to-normal vision were tested in the experiment. They were paid UK£5 (approximately US$7) for each of the three sessions they performed. Two participants who attempted the experiment were not used due to failure to produce sufficiently rapid or accurate responses. (One of these participants had mean RTs greater than 1.5 s and many very slow responses; the other achieved an overall level of accuracy of less than 70% correct.)

Model-fitting method. The data were fit using the same optimization procedure as in the previous study. The cost function included terms for the accuracy, the mean and standard deviation of the RTs, and the skew of the correct RT, at each level of discriminability.

Experimental results and model fits. In Figure 17, we present results and model fits for the accuracy and for the mean and standard deviation of

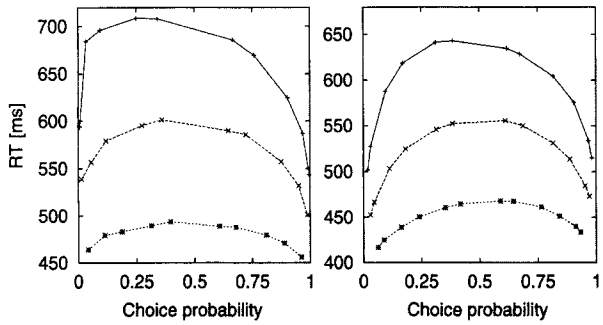

Figure 16. Latency-probability (LP) functions for accumulator with both leakage and lateral inhibition, for three different values of the response threshold. Top curve: threshold = 1.2; middle curve: threshold = 0.9; lower curve: threshold = 0.6. Left panel: LP functions without starting-point variability. Parameters as for Participant 4 in Table 3, except of β = .88. Right panel: LP functions with starting-point variability. The starting point of each accumulator was initialized to a value chosen at random from a uniform distribution from 0 to 0.8 of the threshold.

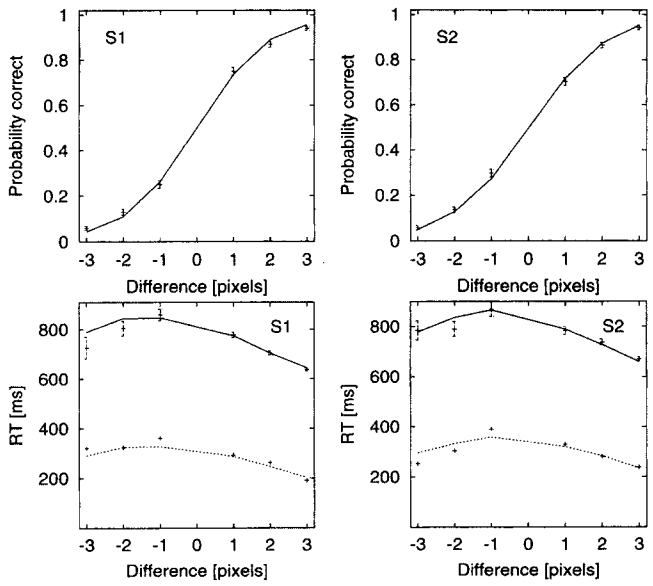

Figure 17. Data (points with error bars) and model fits (lines) for accuracy (top) and latency (bottom, solid line) and standard deviation (bottom, dotted line) for 2 participants in Experiment 2. Error bars represent 1 *SE* of the accuracy and latency data. The model parameters used in the fits for Participant 1 are $\theta = 1.608$, $\sigma = .426$, $\beta = 2.23$, $k = .06$, $\epsilon = 0.1575$, and $T_0 = 305$. Parameters for Participant 2 are $\theta = 1.14$, $\sigma = .322$, $\beta = 4.0$, $k = .12$, $\epsilon = 0.10$, and $T_0 = 300$. RT = reaction time. *S* = subject.

the RT, as a function of discriminability for the two participants (as before, negative values on the *x*-axis denote incorrect responses). The data show that the subtle changes in stimuli (differences of one or two pixels) resulted in detectable and systematic effects on accuracy and RT, reproducing the LP patterns reported in the Vickers (1970) experiment (U-shaped RT distributions with error RT slower than correct RTs).

In Figure 18, we show the RT time distributions for correct responses of the two participants at the three levels of discriminability and the corresponding model fits. These RT distributions are similar to those that have been reported in many other experiments. The model was not explicitly fit for the shape of the RT distributions. Nevertheless, once its parameters were fit to the accuracy, mean, standard deviation, and skew of the RT data, it generated RT density distributions very similar to those seen in the experiment.

Discussion

One of the most stable patterns of the results is that the distributions are strongly skewed (skew larger than 1) and that the skew values increase with discriminability. Specifically, for difficult, medium, and easy, respectively, the skew of the correct RT distribution is 1.3, 1.8, and 2.2 for Participant 1 and 1.5, 1.8, and 2.2 for Participant 2. These results are well predicted by the model and indicate that the trends of increasing skew with discriminability discussed in the previous sections and reported in other studies (Wilding, 1974) are not simply artifacts of noisy estimation of high moment RTs.

The leaky, competing accumulator model does not suffer from the same limitations as the classical accumulator models in fitting the shapes of RT distributions. Overall, it appears that the leaky, competing accumulator model can provide a good account of the patterns of data found in LD experiments.

Experiment 3: Effects of Information Arrival Time

Up to this point, we have shown that the leaky, competing accumulator model accounts for experimental data from two broad types of experiments. But models that do not adopt the principles of leakage and competition may also be consistent with much of this data, and so it is important to consider whether the principles lead the model to predict new findings. Here we describe an experimental paradigm that allows a test for predicted effects of leakage and lateral inhibitory interactions. This can be done by presenting fast sequences of visual stimuli chosen from two classes (say, red or green colored lights or the letters H or S) and requiring participants to estimate the more numerous or predominant stimulus in the sequence. Experiments of this sort have been done previously by Erlick (1961) and more recently by Vickers (1995; Pietsch & Vickers, 1997).

Consider a situation in which the two response alternatives receive equal support overall (e.g., same number of red and green flashes, in total) but one type of event predominates at the beginning of the sequence and the other predominates at the end. According to perfect accumulator models, such a manipulation would not lead to any choice bias; counts are integrated whenever they arrive, adding up to the same total. In contrast, a model based on a leaky accumulation process predicts a choice bias for the type of event that predominates at the end. Earlier counts are less effective because their impact on the activation of the accumulators has more time to decay. Lateral inhibition has the opposite effect. The impact of early counts is larger because they inhibit the activation of competing accumulators activated later on in the sequence. Depending on the balance of inhibition and leakage, a variety of effects can be obtained. These points can be obtained analytically for the mean trajectory (averaging over the noise). An illustration based on simulations with noise included is shown in Figure 19.

In most of the existing experiments of this type, each flashed stimulus may be separately perceived, and the task extends over several seconds. Therefore, the decision process might operate at a different time scale than ordinary perceptual processes and may tap accumulation of information in memory rather than information integration in perception (Pietsch & Vickers, 1997). Vickers (1995; Pietsch & Vickers, 1997) recently reported strong individual differences in studies of this type. Most participants showed a bias toward the stimulus predominating at the end of the sequence, but a few participants showed a bias toward the stimulus predominating at the beginning. Therefore, in this paradigm, it is essential to examine individual rather than averaged RT data. Finally, it is also important to ensure, using an objective measurement, that participants are indeed estimating the predominant stimulus. To address these issues, we carried out a new experiment using fast sequences presented at a high rate of 60 stimuli/s, collecting enough data per participant to examine patterns of performance on a participant-by-participant basis. To ensure that participants attempted to estimate the predominant stimulus, half of the trials in each block had a predominant stimulus, whereas the other half were balanced with equal number of S and H stimuli. The critical

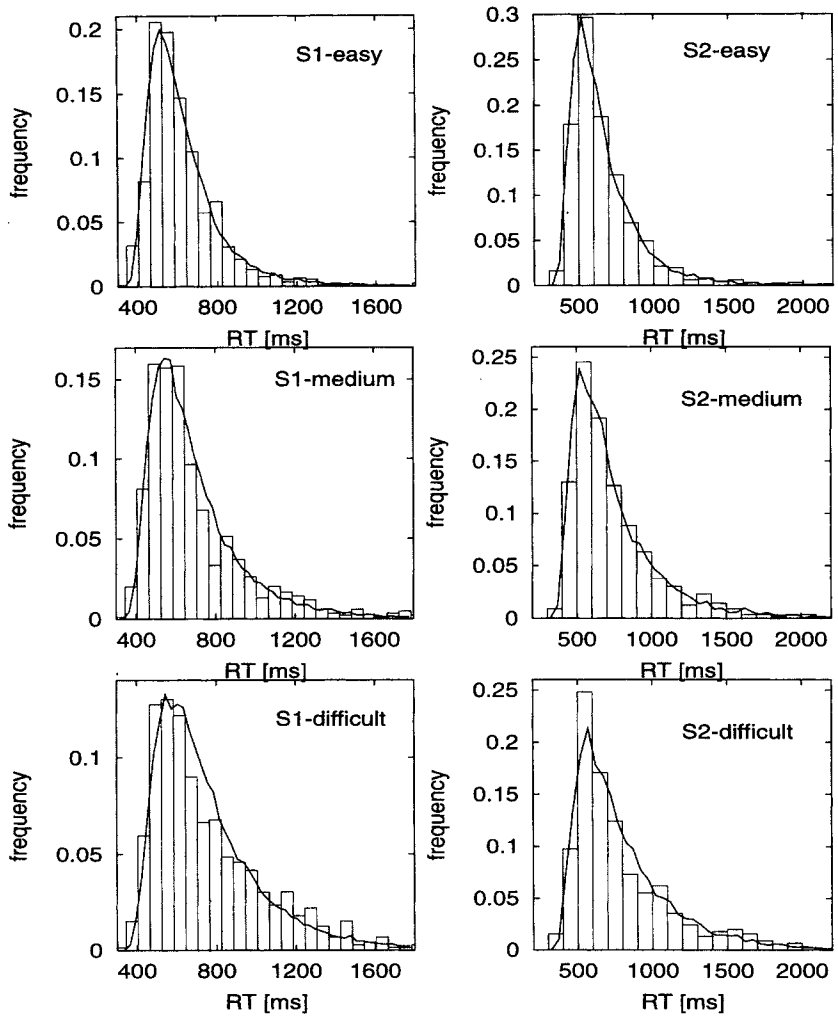

Figure 18. Fits to the reaction time (RT) density distributions at the three levels of accuracy. S1 = Subject 1; S2 = Subject 2.

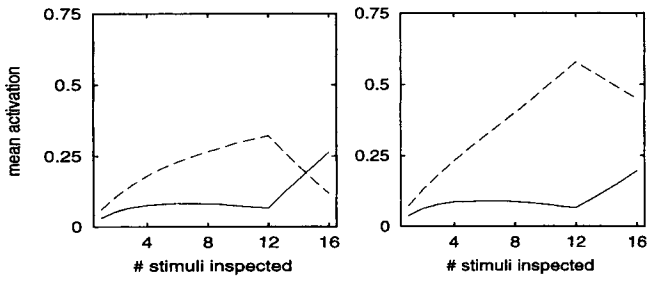

Figure 19. Mean activation of the noisy trajectories, averaged across 4,000 trials of random sequences with equal number of stimuli of each kind (eight) and with a cluster of at least four stimuli of type I at the end of the 16-stimulus sequence. The solid curve corresponds to the alternative receiving support from the last four stimuli. The dashed curve corresponds to the other alternative, which is supported by two thirds of Stimuli 1–12. Left: High leakage (K = .22, β = .1), corresponding to Participant 1. Right: High inhibition (K = 0, β = .25), corresponding to Participant 5.

trials of the experiment contained an equal number of Ss and Hs overall but had a short cluster of Hs or Ss at the very end of the sequence, creating a slight predominance of the other stimulus over the remainder of the sequence. The critical sequences contained 16 elements, so they required only 256 ms to present, well within the period that appears necessary for asymptotic information integration, as indicated by the time–accuracy curves from time-controlled experiments.

In a recent article, Pietsch and Vickers (1997) suggested that the primacy effect observed for some participants in a similar task might be because they selected a response before inspecting the whole sequence. To discourage such a strategy, the critical trials were randomly intermixed with background trials in which one of the two stimuli predominated. These sequences in the background trials had 40 elements, resulting in a duration of 640 ms, with one of the stimuli occurring 25 times in the sequence. Because participants could not know in advance the duration of a stimulus, we reasoned that they would be discouraged from selecting a response on only a fraction of the elements in the critical sequences, because

such a strategy would result in a very low performance level on the longer sequences.

Method

Materials. Stimuli were presented and controlled on a Silicon Graphics (O2) computer with double buffer mode, using C code developed for the experiment. The S and the H stimuli were white on a gray background and at low contrast. The size of stimuli extended for about 2° of visual angle, and they were viewed from a distance of about 1 m, in a dark room. The two types of trials were randomly intermixed. In the critical trials, two cluster sizes were used ($n = 4$ and $n = 2$). The order of stimuli within a sequence was randomized. Thus, for a specified cluster of size 4, at least the last four stimuli were of the S (or H) type.

Procedure. Participants were told to press the left mouse button if the stimulus was predominantly S and to press the right mouse button if they thought it was predominantly H. Participants were not told to respond rapidly, and RTs were not recorded. They were told to make this decision in a global way, based on the entire sequence, and to guess if they were not sure. Participants were not told that some of the trials contained clusters. They were told that a beep signal would be provided after error responses to help learn the task. Once a response was generated, the computer program presented the next trial after a delay of about 2 s. The beep signal was provided after errors only on the background trials (with predominant stimuli). Participants first performed a 20-trial practice block, which consisted only of sequences with predominant stimuli (no clusters). If their performance on this training set was better than 80%, they proceeded to the experimental blocks. Otherwise, they repeated the practice block with a new random series of trials. After that, each participant performed six blocks of 80 trials. Within each block, there were equal numbers of critical and background trials. Each stimulus (S or H) predominated in half of the background trials, and there were also equal numbers of critical trials ending with clusters of 2 or 4 Ss or 2 or 4 Hs in each block. Finally, participants completed two control blocks in which the critical trials did not contain any predetermined clusters ($n = 0$).

Participants. Six participants all with normal or corrected to normal vision, performed in the experiment for credit or for a payment of UK£3 (approximately US$5).

Data collection and analysis. The computer software collected the total number of correct responses on background trials and the number of responses in favor of the end-cluster letter on the critical trials. These scores were pooled over blocks. Despite the response symmetry in the design (equal number of S and H trials), participants sometimes showed response biases to one of the responses. To obtain bias-free measures of accuracy and cluster choice, the geometric average of the probability of cluster choice for S and for H trials was computed separately for each participant and experimental condition.

Results

All the participants performed the task on the background trials with an accuracy larger than 75%; individual accuracy data on these trials are given in Column 4 of Table 4 (chance level is 50%). Columns 1–3 in Table 4 show the fraction of trials in which the participants chose the end-cluster letter, for cluster size of 4 and 2 and for the no-cluster control trials. Participants varied in the pattern of cluster-choice probability, showing three qualitatively different patterns (participants are grouped and numbered to bring these patterns out). Two of the participants (top row in Figure 20) show a choice preference for the end cluster (Participant 1: $p < .0001$, $z = -4.47$, and Participant 2: $p < .01$, $z = -2.68$), whereas 2 other participants (bottom row) show a preference against the end cluster (Participant 5: $p < .001$, $z = 3.76$, and Participant 6: $p < .0001$, $z = 4.47$). The remaining 2 participants (middle row) showed a more neutral pattern of approximately balanced choice (no significant trend was present in either case).

It seems likely that the pattern of choice found in the cluster trials reflects the same process of perceptual estimation that is used to achieve fairly accurate performance on the background trials. Participants were able to correctly discriminate the most numerous stimuli from the longer sequences of 40 stimuli on the background trials, and these trials are randomly mixed with the critical cluster trials. Also, the fact that the clusters of length 4 produce more marked effects than clusters of length 2 suggests that performance on cluster trials does not simply reflect a tendency to choose (or not to choose) the response associated with the last element of the sequence.

Model Fits

To model performance in this task, we presented the model with sequences of input that obeyed the same statistics as the stimulus sequences in Experiment 3. We assumed that when one of the stimuli was presented, the accumulator corresponding to that stimulus would receive an increment, which is a Gaussian random variable whose standard deviation is a model parameter. The additional parameters are the leakage and the lateral inhibition, which operate as described in the previous sections. Choice is determined by the accumulator that reaches the highest level of activation at the end of the sequence (note that our experimental manipulation should have ruled out early selections). In Figure 20, we show the end-cluster probability for each participant, with a fit of the leaky, competing accumulator model.

The patterns of performance shown by the individual participants are consistent with the parameter values found in fitting the model to the data. The first pattern (preference for the end cluster) is obtained when the leakage is dominant over the mutual inhibition; at high leakage, the effect of the beginning is fading at the end of the sequence, and therefore the last cluster in the sequence has a larger probability to determine the choice (Figure 19, left). The opposite pattern occurs when the inhibition is dominant over the leakage; in this situation, the activation received at the end of the sequence is inhibited relative to activation received earlier (Figure 19, right). The neutral pattern is found where there is an approximate balance of leakage and inhibition.

To illuminate in more detail the processes leading to the different choice patterns, we show in Figure 19 the activations of the two accumulators, averaged across 4,000 trials using a four-stimulus

Table 4
Proportion of Choices of Response Corresponding to End-Cluster Letter and Background-Trial Accuracy in Experiment 3

	End-letter cluster			
Participant	$n = 4$	$n = 2$	Control	Accuracy
1	.75	.61	.48	.83
2	.64	.57	.53	.77
3	.54	.44	.48	.88
4	.40	.44	.49	.96
5	.28	.46	.51	.88
6	.25	.41	.50	.82

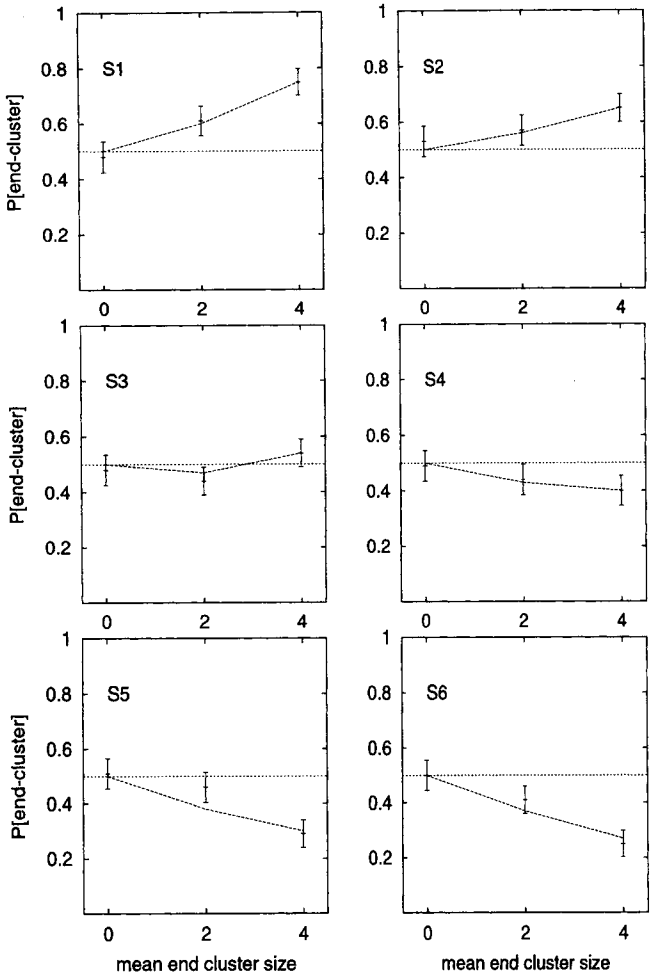

Figure 20. Individual data for Experiment 3. The probability of choosing the end-cluster stimulus is shown for cluster size of 4, 2, and 0 (control). Error bars correspond to the standard error in probability. The solid lines represent fits of the leaky, competing accumulator model. Three parameters were varied: leakage (K), lateral inhibition (β), and the standard deviation of the noise in the accumulator increment. S = subject.

end cluster, for parameters corresponding to Participant 1 (left) and Participant 5 (right). The solid curve shows the activation for the unit corresponding to the end cluster, and the dashed curve shows the activation of the unit for the other choice alternative.

The set of parameters used to fit the individual data in Figure 20 were used a second time to generate model predictions for the accuracy in the predominant (40-stimuli sequence) trials. These predictions are shown in Figure 21, plotted against the experimental data. There is a close correspondence between the predicted and obtained results. In particular, as predicted by the model, participants with the more neutral pattern on the critical trials achieved the highest accuracy on the background trials.

Discussion

Overall, it appears that the data are consistent with models that incorporate leakage and lateral inhibition and that interesting patterns of the relative values of the leakage and inhibition parameters can account for aspects of individual participants' performance. However, it is worth considering other possible accounts for the patterns seen in the data. As previously discussed, Pietsch and Vickers (1997) have reported individual differences in performance on a similar task, where participants attended to sequences of two different colored light flashes and estimated which color occurred more frequently, and they offered different sorts of interpretations. In their experiment, the sequences were long, and each item could be individually perceived, and this may have allowed participants to selectively attend either to the beginning or the end of each sequence or to rely on postperceptual memory processes. However, in our experiment, the presentation speed was much faster, and the stimuli were presented at very low contrast. Inspection of the stimulus sequences indicated that the elements could not be individually perceived and on the critical trials the sequences were only 256 ms long, well within the period that seems to be required for integration of perceptual information based on fits of our model to the results of other experiments (e.g., for Participant 1 from Experiment 1, the participant with the fastest integration rate in that experiment; 256 ms allows only enough time to reach 85% of asymptotic accuracy). Therefore, it seems likely that the task tests primarily integration of perceptual information. Nevertheless, it is still possible that despite the very short exposure time, some postperceptual form of memory is also involved. In addition, participants could attend more to the beginning or the end of the sequence. Such processes could be incorporated into models that do not assume leakage or lateral inhibition, and it seems likely that with such assumptions, these models could also account for the present data. No such additional assumptions are necessary in our model, which inherently incorporates both leakage and lateral inhibition among the choice alternatives.

Models based on differential attention might easily predict both a primacy and a recency effect. In contrast, the mechanism we have explored cannot produce such a U-shaped effect. The bias toward early versus late information is a function of the difference between the inhibition and the effective decay, and when these factors are in balance, the approach predicts that information arriving at all times will have the same impact. Furthermore, our

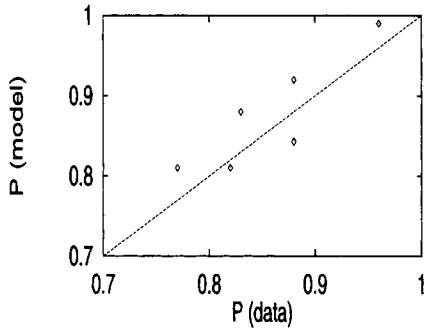

Figure 21. Accuracy in the predominant trials for the individual data (6 participants) on the x-axis versus the corresponding model predictions (y-axis). The participants that achieved the higher accuracy levels are the ones who have a relatively flat cluster probability curves (Participants 3–4), corresponding to a balanced profile of leakage and inhibition.

account explicitly predicts that accuracy is highest for participants who show no bias for or against the final cluster and falls off as bias increases in either direction. This pattern is an intrinsic consequence of the dynamics of imbalance of leakage versus competition and makes it subject to disconfirmation by later experiments. Such a relationship is not necessarily predicted if the bias arises from uneven allocation of attention. Note that we do not wish to claim that such uneven allocation cannot occur. Our only claims are that many effects that might have been attributed to such factors may arise from the balance of inhibition and effective decay and that such a mechanism is sufficient to account for the bias effects seen in the present experiment.

Multialternative Choice

Up until now, we have compared our model with data from experiments with only two choice alternatives. In this section, we consider generalization of the model to the case of multiple distinguishable alternatives, such as letters or positions around a circle (Hick, 1952; Merkel, 1885).[14] The model of leaky, competing accumulators applies directly to choice among any number of alternatives, so the extension is automatic. Classical accumulator models also extend naturally to any number of alternatives, but it is less clear how to extend random walk and diffusion models. The diffusion model with its single variable is equivalent to a model in which there are two diffusion variables, one for each alternative, and the decision to respond is based on the difference between their values. This way of thinking about the model suggests a generalization to N alternatives in which there is an accumulator for each alternative and the decision to respond is based on the difference in activation between the most active accumulator and the next most active. This most-minus-next criterion has been suggested by a number of investigators, including Ratcliff and McKoon (1997) and Ratcliff and Rouder (1998).

Hick's Law

In this section, we consider which if any of these approaches leads to an account for the well-known linear relationship between RT and the log of the number of alternatives (Hick, 1952; Merkel, 1885; see Figure 22). We have chosen to focus on this relationship, widely known as *Hick's law,* because it is a well-established and general fact and directly confronts the information-processing consequences of variations in the numbers of response alternatives (see Luce, 1986; Teichner & Krebs, 1974, for review). Originally, Hick interpreted this regularity according to a sequential and hierarchical model of choice, which later was strongly criticized (Laming, 1968). A different explanation of this law on the basis of N parallel and exhaustive processes also has been developed (Christie & Luce, 1956; Laming, 1966). Although Vickers (1979, chap. 8; see also Vickers & Lee, 2000) has developed an extension to the case of multiple stimuli arrayed on a single dimension, stochastic information accumulation models have not previously been applied to the multialternative case with higher dimensional stimuli such as letters or digits, as far as we know.

In the following, we address RT to make a choice response among N alternative stimuli and consider how such RTs vary as a function of the number of alternatives N. In developing an account of these findings, a crucial initial consideration is the effect of set size on the input to units in the network. We have adopted the assumption that the input to an accumulator produced by a particular stimulus is unaffected by the number of response alternatives. This seems to be a natural assumption to make within the context of our model, where adding more accumulators would not change the connections between the input units and existing accumulators, but in other frameworks, other assumptions may be very natural, and indeed others have been suggested. For example, in a counter model of multichoice decision making, Ratcliff and McKoon (1997) propose that on each time step, only one accumulator gets a count. Either a valid count is added to the correct accumulator with probability p; or with probability $1-p$, a count is assigned to any one of the alternatives (including the correct one) at random. The effect of this is that the rate of accumulation of counts at a given accumulator varies with the number of alternatives.

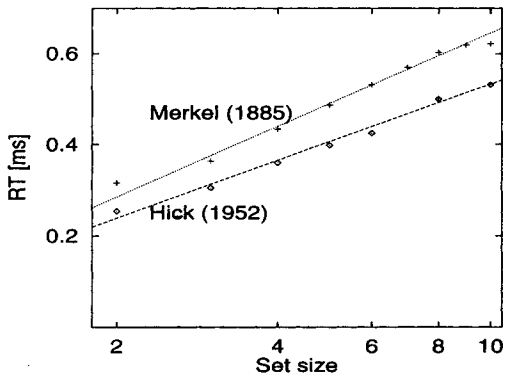

Figure 22. Multichoice reaction times (RTs): pluses (+) signify experimental data from Merkel (1885; choice stimuli are the set of numbers 1–5 and I–V); diamonds (◇) signify data from Hick (1952; choice stimuli are lights arranged in an irregular circular array arranged to prevent grouping). Responses are made by pressing response keys. Data from Hick (1952, Figure 1); Merkel's data from Woodworth (1938, p. 333).

For present purposes, we work within the assumption that in all conditions of a particular experiment, the input to the correct accumulator consists of a signal ($\rho_c > 0$) plus noise, whereas the input to each of the incorrect accumulators is a smaller signal ($0 \leq \rho_i < \rho_c$) plus noise. Thus, the mean value of the input from the feature units arriving at any incorrect accumulator is the same for all such accumulators and is independent of the number of alternatives. Obviously, this is an approximation to the actual conditions of any real experiment, in which confusability of the stimuli is unlikely to be completely homogeneous. The policy just described is consistent with the policy used in our previous analysis of the two-alternative case. For present purposes, there was no reason to constrain $\rho_c + \rho_i$ to sum to 1, because we are not simulating effects of manipulating input to the alternatives along a continuum, as is done in LP experiments.

[14] We do not consider cases in which the stimuli are arranged along a single psychological dimension that introduces limits on the number of alternatives considered (Miller, 1956; see Lacouture & Marley, 1991, for a connectionist model, and Vickers, 1979, chap. 8, for a stochastic computational model of Hick's law in unidimensional absolute identification tasks).

As a further constraint on the modeling effort, we take note of a feature of the procedure that Hick used in his 1952 experiment. Because he was interested in errorless performance, he required participants to undergo preliminary blocks with each set size. During these blocks, participants were instructed to adjust their performance to eliminate errors while still responding as rapidly as possible. In Hick's experiment, several blocks of trials were run at each set size, and the data reported were taken from the block after the first block of the given set size in which no errors were made. Given this aspect of the procedure, we adopt the assumption that participants adjust the criterion as a function of number of alternatives, so as to keep accuracy constant at a high level when the number of stimuli increases. Vickers (1979, chap. 8; Vickers & Lee, 2000) used a similar approach based on confidence (rather than accuracy), which is self-regulated, in his account of Hick's law as it is applied to stimuli arrayed along a single dimension.

It is helpful to begin by considering what would happen if the criterion for responding remained at a fixed activation value (or a fixed activation-difference value) as a function of set size. When the number of choice alternatives increases, so does the number of units that integrate noise. With a fixed criterion, the rate of errors increases for all of the models, because there are more incorrect units whose activation can reach criterion before the correct alternative through the accumulation of noise alone. To compensate for this, participants can increase the criterion as the number of alternatives increases.

Here we report computer simulations showing that the policy of maintaining a fixed error rate leads to a logarithmic relationship between number of alternatives and RT according to all three of the models mentioned above (see Figure 23): (a) the classical accumulator model with an absolute response criterion (i.e., first accumulator to reach criterion is chosen as the response, regardless of activations of other accumulators); (b) the leaky, competing accumulator model with an absolute criterion, and (c) the classical accumulator model with a criterion based on the difference in activation between the most active and the next most active (most-minus-next criterion). In all cases, the value of the criterion is adjusted for each set size to achieve a specified level of accuracy that is held constant across all set sizes. To explore the generality of the results, we consider two different choices of the accuracy level and two degrees of confusability of the alternatives. In a separate project, we have demonstrated analytically that such a policy of maintaining a fixed error rate leads to a logarithmic relationship between number of alternatives and RT (Usher, Olami, & McClelland, 2001) for the special case of Model A, in which the mean of the input + noise is greater than 0 only for the accumulator representing the correct alternative; for all of the other accumulators, the input is simply 0-mean Gaussian noise.

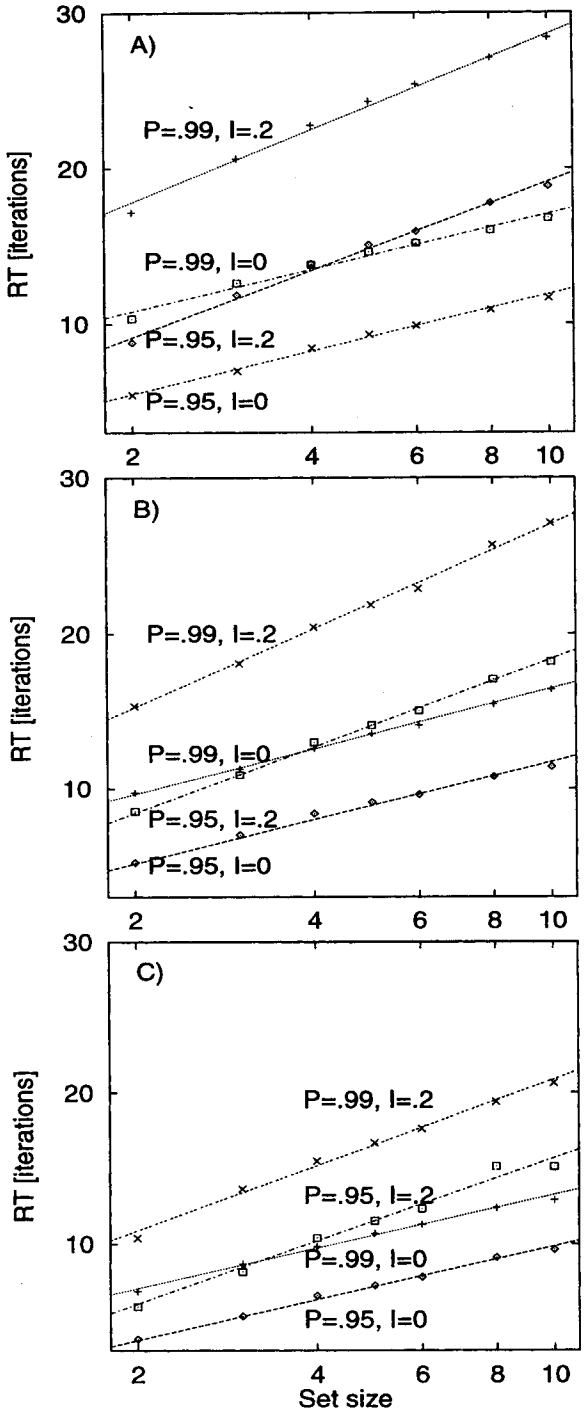

Figure 23. Multichoice reaction times (RTs) in stochastic models as a function of set size (on logarithmic scale) for two levels of accuracy (p = .99 and p = .95) and for two levels of discriminability, ρ_i (0 and 0.2). (A) Classical accumulator; (B) accumulator with leakage (k = .5) and lateral inhibition (β = .2); (C) random-walk model (difference criterion). The straight lines represent logarithmic fits of the form, RT = a + $b$log(n). Each data point corresponds to a block of 10,000 trials of simulations that produced an accuracy of .95 or .99, respectively, with a precision of .001.

Simulation Procedure

For each model and within each model for each combination of accuracy criterion and confusability, the goal of the simulations was to find, for each set size, the value of the response threshold that produced a particular constant level of accuracy and to measure the corresponding RT resulting when that threshold was in use. Hick (1952) reported that participants in his experiment did not generate more than 4% errors in any condition, so they were never less than 96% correct. In the

reported simulations, we chose two values, 99% and 95%. The parameter ρ_c was fixed at 1; confusability was varied by considering two values of ρ_i, namely 0 and 0.2. The remaining parameters of the model were held constant throughout the simulations ($k = .5$, $\beta = .2$, $\sigma = .38$, $dt/\tau = .1$).

Each simulation trial for each model began as usual by setting the values of all accumulators to 0. Information accumulation occurred in a series of time steps. In each time step, the input ρ_c was applied to one accumulator, and ρ_i was applied to all other accumulators, and values of the accumulators were adjusted using Equation 4. As soon as the value of any accumulator reached the threshold (see below), the trial ended, and the corresponding response and the number of time steps taken to reach threshold were recorded. For each combination of model, accuracy level, and confusability, we ran a series of blocks of 10,000 trials, adjusting the response threshold after each block until we found a value that produced an accuracy level within an interval of 0.05% around 95% or 99% (i.e., between 94.95% and 95.05% or between 98.95% and 99.05%). We recorded the average number of iterations needed for correct trials with this criterion.

Simulation Results

In Figure 23, we show the average number of iterations each of the models needed on correct trials to reach the threshold that was found to achieve the desired level of accuracy. The RT data are plotted against the log of the number of alternatives to display the logarithmic dependency of Hick's law as a straight line. The times to threshold can be converted to predicted RTs by specifying a value for the number of milliseconds per time step and a value for T_0, reflecting fixed delays associated with input and output processes, but we have not done this because the main focus of interest is on the linearity of the relationship between log set size and RT, which is not affected by such a conversion. For all three models and all combinations of accuracy level and confusability, the simulation results are very close to linear, indicating a logarithmic relation between set size and RT.

It is interesting to consider the effects of confusability and accuracy level in the various models. As one might expect, both higher confusability and higher accuracy tend to lead to longer responses (remember that the accuracy level is maintained by adjusting the response threshold, so that confusability does not reduce accuracy directly, as it would if the threshold were fixed). Interestingly, changing the accuracy level appears to affect only the intercept of the linear function relating RT to $\log(n)$, whereas changing the confusability affects the slope and the intercept. Although the slope effect appears to be independent of accuracy level, the effect of confusability on the intercept tends to be larger with the higher accuracy criterion.

We have experimented with a range of different parameter values in simulations, and generally the results shown here are representative. There are, however, conditions in which the linear relationship breaks down for the leaky, competing accumulator model. This can occur when the leakage and inhibition are not balanced ($K \neq 0$), and the response threshold is relatively high compared with the expected value of the asymptotic activation of the correct response unit. Under these conditions, it can sometimes take a very long time for the activation of the most active unit to fluctuate to a value consistent with the criterion, and this effect is amplified at large set size (because owing to competition, the activation decreases with set size). We emphasize that this problem reflects considerations that arise as discriminability goes down and desired accuracy level goes up; because discriminability is generally very high in Hick's law experiments (i.e., accuracy would be expected to be 100% given unlimited viewing time), we would not expect this issue to arise in fitting data from standard Hick's law experiments. In our simulations, we have not otherwise encountered discrepancies from Hick's law.

An interesting implication of our results is that it is not necessary to assume any sort of limitation on information-processing capacity or resources to account for the logarithmic increase in RT with the number of response alternatives. Hick's law instead appears as a compensation for the increased opportunity for error that arises from an increase in the number of accumulators. Indeed, the results reported above depend on the use of a constant accuracy policy, under which the criterion is adjusted for each set size to maintain a fixed level of accuracy. We do not observe Hick's law behavior in these models when a constant activation criterion is used independent of the number of alternatives. All of the models show an increase in error rate as the number of alternatives increases. The classical accumulator actually shows faster mean RTs as set size increases, because trials that would otherwise have produced longer RTs with a smaller number of incorrect alternatives are more likely to end with errors as the number of alternatives is increased. In the other two cases, RT increases with set size, but the increase in RT is not logarithmic. The use of a constant accuracy criterion is thus a general principle that leads to logarithmic dependency of RT on response set size, for several stochastic self-terminating models.

It seems likely that participants do not always maintain the same fixed level of accuracy as a function of set size; in very many cases, the error rate tends to increase in conditions that lead to larger RTs, and we would expect that to be the case in Hick's law type experiments, unless stringent precautions are taken. On the other hand, we doubt very much that participants would not make some compensatory adjustment of the response criterion for the different set-size conditions, because otherwise they would either be excessively slow with small set sizes or make excessive numbers of errors with large set sizes. (Vicker's 1979 principle of self-regulation could provide a way to implement such compensatory adjustments even in the absence of feedback, on the basis of self-confidence.)

Our analysis highlights the importance of considering the effects of experimental conditions on response initiation criteria; otherwise, it is impossible to make definite statements about the dependence of RTs on experimental conditions. This indeterminacy has been a long-standing source of criticism of RT experiments (Pachella, 1974; Wickelgren, 1977). Thankfully, it is becoming far more common to use models that make explicit predictions of both RTs and errors and to consider both types of data in assessing the adequacy of model predictions.

Hick's law is consistent with a wide range of different models. In addition to the models we have considered here, it is consistent with parallel exhaustive models (Christie & Luce, 1956; Laming, 1966) and possibly several other approaches (Lacoutre & Marley, 1991; Vickers, 1979). Thus, the fact that the leaky, competing accumulator model is consistent with it does not provide differential support for this approach. As in some other cases we have

looked at in this article, the most we can say at this point is that there are several models that are consistent with Hick's law, and the leaky, competing accumulator model is one of the contenders. Also note that there are qualifications on the conditions under which this statement is true. We know it is true under the particular conditions of our simulations (threshold adjusted with set size to maintain accuracy, input to accumulators independent of number of alternatives, and threshold below expected asymptotic activation of the correct response unit).

Visual Word Recognition: The Intersection Principle

Word recognition is a typical task that requires selection among a very large number of possible targets, and as such, it is a good candidate for assessing the adequacy of lateral inhibition as a basis for response selection. Previously, we have considered the benefits of lateral rather than bottom-up inhibition in the interactive activation model of visual word recognition. Another interesting characteristic of lateral inhibition is brought out by a consideration of the data from an experiment by McClelland and O'Regan (1981).

In their experiment, McClelland and O'Regan (1981) tested the benefit in RT that participants can obtain in a word identification task from two different sources of information: a very brief parafoveal preview of the target stimulus and prior semantic context. In the experiment, there was a strong preview condition, in which the preview was the same word as the target; a weak preview condition, in which the preview was a letter string visually similar to the target, and a control preview condition, in which the preview was a row of *X*s. Similarly, there was a strong context condition, in which a prior partial sentence context strongly suggested the actual target word (e.g., "I like coffee with cream and . . . "); a weak context condition, in which the target word tended to be one of many that were consistent with the context (e.g., "The county sheriff had a new . . . "); and a neutral context condition, which provided no constraint on the possible identity of the target (see Figure 24).

McClelland and O'Regan (1981) found that neither the weak preview paired with a neutral context nor the weak context paired with a neutral preview was sufficient to speed the recognition RT, compared with the neutral-context/neutral-preview baseline. However, a combination of a weak preview and a weak context did speed the identification of the target word. A recent experiment by J. Hinton, Liversedge, and Underwood (1998) produced a similar effect. They used ambiguous partial-word primes with ambiguous semantic primes and showed that each type of prime alone did not produce significant priming but when combined they produce a reliable priming effect. Such a pattern might be expected if the word recognition system exploits the intersection of information sources (Humphreys, Wiles, & Dennis, 1994). According to this idea, each weak source activates several alternatives in addition to the target. When two such sources intersect, the correct alternative will receive support from both sources. Given that context selects a set based on semantic relationships and a preview selects a set based on visual relationships and given that semantic and visual relationships are approximately independent (i.e., visual similarity is only at best very weakly correlated with semantic similarity), other words will rarely be consistent with both sources of information, so the correct alternative will receive two sources of support and therefore can win out over all the other alternatives. As we show next, a lateral inhibition scheme can implement this intersection principle and can explain why the correct target is boosted when two sources of support intersect, even though there is little or no benefit when only one source is available.

Consider a system of n leaky, competing accumulators based on Equation 4, where n is the size of the vocabulary. We assume that each weak source alone sends input to the correct word alternative and to $m - 1$ other alternatives within the word's neighborhood. For simplicity, we choose all these inputs to be equal to each other and denote their strength by Δ. Second, we assume that the two sources add their input contributions to the target units linearly. Thus, when two sources of weak input exist, each one projects input of strength Δ to $m - 1$ different word units, whereas the correct alternative, activated by both sources, receives an input of 2Δ.

To estimate the priming benefit under each condition, we assume that when the word is presented, a strong signal projects to the target unit. The starting point of this unit activity, however, depends on its prestimulus activation, which is a function of the preview and context information. We focus attention here only on the four conditions crucial to the interaction, namely, those involving neutral or weakly constraining sources of information, because a single strong constraint appears to be sufficient to produce facilitation on its own. To obtain the simplest qualitative account for the effect, we neglect here the dynamics of the preview and context information and assume that when the word target is presented, the corresponding unit is in a steady state of activation that characterizes the prestimulus condition.

It is easy to calculate the steady state solutions (obtained by requiring that $dx_i/dt, = 0$, for all $0 < i \leq N$) in Equation 4. Consider first the case of a single weak source. All the units that do not receive input are inhibited and are thus suppressed to

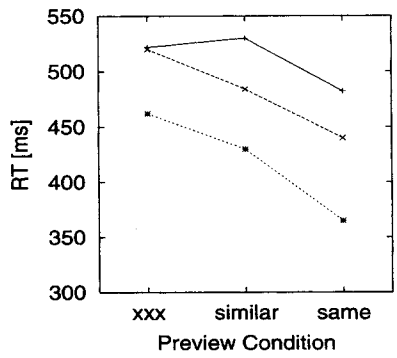

Figure 24. Joint effects of context and preview information on the time it takes to read a word target aloud. Solid line (pluses) = neutral semantic context; dashed lines (*X*s) = weak semantic context; dotted line (asterisks) = strong semantic contexts. Visual preview condition is shown on the *x*-axis and recognition reaction time (RT) is shown on the *y*-axis. Adapted from "Expectations Increase the Benefit Derived From Parafoveal Visual Information in Reading Words Aloud," by J. L. McClelland and J. K. O'Regan, 1981, *Journal of Experimental Psychology: Human Perception and Performance, 7,* p. 639. Copyright 1981 by the American Psychological Association.

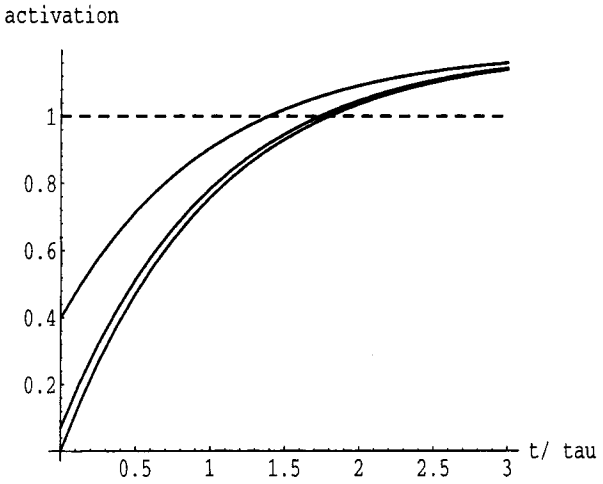

Figure 25. Qualitative illustration of priming effect in McClelland and O'Regan (1981) experiment. The word *input* has an asymptotic value of 1.2, a time constant of 1, and the response criterion is 1. The lower curve, which starts at 0, represents the control. The second curve represents a single weak information source, characterized by the function $x(m) = \Delta/[1 + \beta(m - 1)]$, $\beta = .5$ $\Delta = .2$, and $m = 5$ (Equation 13). The third curve represents a combination of two weak sources leading to a starting point activation of 2Δ.

zero (we assume threshold linear units, so that a negative synaptic current x_i results in a zero activation, or firing rate). All the units that receive input will have the same steady state (due to the equal input values) of $x(m)$, which satisfies

$$x(m) = \frac{\Delta}{1 + \beta(m - 1)}. \tag{13}$$

Without lateral inhibition, each unit would have a steady state activation of Δ that reflects the input. The effect of the lateral inhibition is a reduction in the magnitude of this activation in proportion with the number of activated units and with the strength of the inhibition.

Consider next the steady state of the units in the condition of two weak intersecting sources. Again, the units that receive no input are suppressed to zero. All the nonintersecting units that receive input will have a steady state denoted as x_N and the intersecting unit has a steady state of x_I, which satisfy

$$X_I = 2\Delta - \beta(M - 1)x_N,$$
$$X_N = \Delta - \beta x_I - \beta(M - 2)x_N. \tag{14}$$

(Here, $M = 2k - 1$.) The solution for the nonintersecting units, x_N is $X_N = x(m)(1 - 2\beta)/(1 - \beta)$. Thus, the presence of the intersecting unit leads to an additional suppression of the M nonintersecting units by a factor of $(1 - 2\beta)/(1 - \beta)$ (the value of 2 in this expression reflects the fact that we use only two intersecting sources). Thus, for any value of $.5 < \beta < 1$, all the nonintersecting units are suppressed to zero (negative synaptic current leads to zero firing rate), and as a consequence, the steady state activation of the intersecting unit is restored to its value that reflects the input alone, $x_I = 2\Delta$ (Equation 14).

In Figure 25, we display the deterministic time course of activation for a leaky accumulator, under the impact of the same stimulus (reflecting a common asymptote and time constant) but from different starting positions that reflect the prestimulus condition. We observe that the first curve (which reflects a control condition; zero initial activation) and the second curve (reflecting a single weak information source) are very similar and intersect the response criterion at almost the same time. For two weak sources, a significant speedup results.

Because the sensory input corresponding to the target is above threshold, the mean RT arising from the nonlinear stochastic information integration process can be well approximated by deterministic (noiseless) OU trajectories (see, e.g., Figure 7 in Usher & Niebur, 1996). We confirmed this for the present case by performing simulations using the stochastic system characterized by Equation 15:

$$dx_i = [I_i - kx_i - \beta \sum_{j \neq i} x_j]\frac{dt}{\tau} + \xi_1 \sqrt{\frac{dt}{\tau}}. \tag{15}$$

Three conditions were simulated, which are distinguished by the inputs to the network during the first 50 simulation steps (corresponding to the prime): (a) a control condition, in which the input to all units remains zero during the first 50 simulations steps; (b) a weak prime condition, in which $m = 5$ units are activated with the same input, $I = .15$ for 50 simulation steps; (c) a condition corresponding to two independent primes (visual and semantic context) where $2m - 1 = 9$ units are activated with an input of $I = .15$ and Unit 1 is activated with a larger input of $I = .3$ during the first 50 simulation steps. In all cases, following the initial 50 simulation steps, the input to Unit 1 is increased to $I = .5$, and the simulation is continued until the activation of this unit reaches a criterion (chosen arbitrarily as 1). The parameters k and β are chosen as .3 and .65, respectively. The standard deviation of the Gaussian noise is $\sigma = .079$, which ensures that no errors are generated (at this response criterion). The Euler integration time step dt/τ is .1. Blocks of 500 simulations have been performed in each of the three conditions. The mean number of simulation steps needed to reach the response criterion (and their standard deviations) are shown in Table 5.

Thus, indeed, the control and the weak prime result in equivalent RT whereas the intersection of two sources of priming results in a considerable speedup. A similar result should take place according to the countermodel of Ratcliff and McKoon (1997), which uses a criterion based on the difference in activation between the most and next most active units to trigger a response.

Table 5
Mean Simulated Time Steps Needed to Reach Criterion in Simulation of McClelland and O'Regan (1981)

Control		One weak prime		Two weak primes	
M	*SD*	*M*	*SD*	*M*	*SD*
42.9	10.0	42.8	10.4	34.7	10.7

General Discussion

We have presented a model within the classical tradition of stochastic information accumulation models. Like the classical models, our model assumes that information processing takes place through the gradual accumulation of intrinsically noisy signals. In previous work, this framework has been exceptionally productive, both in terms of its ability to provide a fruitful context within which various principles of information processing can be explored and for its ability to capture experimental data.

A fundamental motivation for our work has been the idea that the human information-processing system may not be a perfect integrator of information. Instead, we have explored the possibility that the information accumulation process may be subject to effects such as leakage or amplification of differences (including differences attributable to noise) due to recurrent interactions among the accumulators. This amplification, as we explored early on in the article, can arise either from lateral inhibition between the accumulators or through recurrent activation processes that feed back a portion of an accumulator's activation onto itself.

These assumptions contribute directly to two classes of findings that we have reported: (a) successful fitting of the general shape of time–accuracy curves and of the specific shape of such curves seen in our own perceptual identification experiment (Experiment 1) and (b) successful tests of two specific predictions following from the presence of these imperfections: that participants might be differentially influenced by early or late elements of a very rapid sequence of elements, and that those participants who show such differential influence will in general be less accurate than those whose performance more nearly indicates the ability to maintain a balance between leakage and recurrent amplification.

Note that the classical random walk and pure accumulator models cannot address either of these two types of data. Such models predict that accuracy will always grow without bound as information accumulation continues, resulting in time–accuracy curves that never asymptote. Such models also predict that information will have the same impact on the outcome of information accumulation, whether it comes early or late. This prediction is consistent with the behavior of 2 of the 6 participants in Experiment 3 but not with the behavior of the other 4 participants in this experiment.

We have given extensive consideration to one alternative to our way of accounting for the fact that time–accuracy curves tend to level off below 100% accuracy levels. This alternative is the possibility that the overall direction of drift may vary from trial to trial within each condition of an information-processing experiment, as suggested by Ratcliff (1978). As noted earlier, we view the idea that there may be such drift variance as a very plausible one, particularly when it is applied to experiments involving highly inhomogeneous items, such as items from a memory experiment. In such cases, item differences could easily arise from a number of sources, including adequacy of study or relative similarity of a test item with other studied items, and these would very likely play an important role. Similarly, in perceptual identification experiments, relatively slow-varying factors such as variation in attention could play a role in altering the processing systems' response to a given input.

On the basis of the considerations raised above, we suggest that a reasonable approach may be to assume that stochastic information accumulation processes incorporate both within-trial imperfections in the accumulation process, as we have suggested, and between-trial variation in the overall direction of information accumulation. An interesting issue for further research is to determine which factor dominates in particular experimental situations. However, it may be difficult to obtain clear conclusions on this matter, due to the subtlety of the differences between the predictions that arise from imperfect accumulation versus drift variance, especially within the context of newer versions of the DDV, which assume both starting-point variability and the use of decision boundaries to terminate processing before the response signal on some trials in time-controlled experiments (Ratcliff, 1988; Ratcliff & Rouder, 1998). These factors tend to bring the shapes of empirical time–accuracy curves predicted by the DDV closer to the shifted exponential shape found in the experimental data and make it difficult to distinguish from our account based on leakage of information. Another factor is cascading of several stages of processing, resulting in nonstationary drift (Heath, 1981; Ratcliff, 1980), which tends to obscure the details of the shape of individual stages of processing, especially those that are not rate limiting (McClelland, 1979).

A key feature of our model is its reliance on lateral inhibitory or competitive interactions between accumulators as part of the information accumulation process. This mechanism may contribute to the effects discussed above. More fundamentally, it provides a very direct way of allowing the activations of accumulators to reflect the relative support provided to them by the input. We have offered a number of arguments motivating the exploration of this mechanism, including (a) the existence of physiological evidence that such a mechanism is used during information processing in the brain, (b) the naturalness of the generalization of this mechanism to any number of alternatives, and (c) the computational advantages of this mechanism relative to the use of feed-forward inhibitory mechanisms in the presence of ambiguous input. Given these motivations, two questions that arise are (a) whether the use of such a mechanism is consistent with the data and (b) whether there are any behavioral findings that clearly support the lateral inhibition approach relative to other proposals.

Regarding the first question, it does appear that lateral inhibition is consistent with all of the different kinds of data that we have considered here. We have shown that models incorporating lateral inhibition can be used to fit a range of different kinds of data, including time–accuracy curves, LP data, instances of greater dependence on early arriving information in a rapid sequence of inputs, data from experiments on the effects of the number of alternatives on choice RT, and data from experiments on the effects of multiple sources of information on word-identification latency.

Regarding the second question, a number of aspects of the available data appear to indicate the existence of some form of competition between alternatives, without indicating whether the mechanism producing the competition effect is lateral inhibition or the use of a relative evidence criterion. The pattern of highly symmetric and deep LP functions that has been reported for some participants (Ratcliff et al., 1999, and Figures 12 and 16, this article) as well as the interaction between visual primes and semantic contexts in word identification (McClelland & O'Regan, 1981; see also Hinton et al., 1998) both receive very natural interpretations under the view that some form of competition is occurring. In the McClelland and O'Regan (1981) word-identification experiment, for example, a single, weak type of prior information-either context alone or a visual prime similar to the target-produced no priming when presented alone, but priming did occur when both types of prime were presented together. We showed

that a competitive model can explain this by assuming (a) that each type of prime alone activates a number of alternatives, so that the competition among them prevents an advantage for the correct alternative, and (b) that two primes together typically leave only the correct alternative in the set of items primed by both sources. In a pure accumulator model without competition, one would expect the effects of the primes to be simply additive. Similar logic applies to the recent experiment by J. Hinton et al. (1998). They used primes consisting of all but one letter of a target word and found significantly more priming when the target was the only word consistent with the prime than in cases where there were several words consistent with the prime. Again, the result follows naturally from models that include some form of competition but is not predicted by noncompetitive race or accumulator models.

Distinguishing between the mechanism of lateral inhibition and the use of a relative evidence criterion may be more difficult. However, there are three relevant points. First, classical models that use a relative evidence criterion, either in the standard form of the random walk, where only the difference variable is maintained, or in the form of independent accumulators with a best-minus-next criterion, cannot account for the finding we obtained with the two participants in Experiment 3 whose responses were biased toward the item that predominated early in a stimulus sequence. Lateral inhibition may not be the only way to account for this effect (differential attention to the beginnings of sequences, and recurrent self-excitation are other possibilities), but if lateral inhibition were the mechanism used to account for response competition, no additional mechanism would be required to account for a bias toward stimuli occurring early in a sequence.

Second, depending on the exact level of lateral inhibition, one might predict that primes that are equally consistent with at least one incorrect alternative in addition to the correct alternative would result in a diminished, but not a null, priming effect, whereas a null effect is predicted by the best-minus-next relative-evidence criterion. The only experimental results we know of that directly address this issue are not completely consistent: Ambiguous primes show a significant priming effect (relative to the control) in two out of four conditions tested (J. Hinton, personal communication, April 6, 1998). Further evidence would be useful on this point, but the data do suggest that ambiguous primes do at least sometimes produce a significant effect, and given that the effect might be small, it is understandable that it might not always be statistically significant.

Third, Vickers (1979) has noted some problems that arise for relative evidence models in experiments where participants are asked to indicate their confidence immediately after their choice response. Specifically, he has noted that classical models using a relative evidence criterion have difficulty accounting for both of the following findings: (a) In standard RT tasks, longer RTs correspond to lower confidence ratings (and vice versa). (b) In tasks that control processing time, confidence increases with processing time. Diffusion or random walk models that use a single random walk process for all responses of all confidence levels are not able to account for the first finding, because the relative evidence is assumed to be at the same criterial value at the moment of response for both fast and slow responses. One might suggest that confidence judgments are based on the RT itself in standard RT tasks, with higher confidence going to shorter RTs, but Van Zandt (2000) presents data that contradict this explanation: The RT distributions for the responses given at different confidence levels have considerable overlap, and thus RT could not be sufficient to separate the various confidence responses.

In the classical accumulator model, an absolute criterion is used to generate the response in an information-controlled task, and the response signal or other source of timing information determines the time for generating the response in a time-controlled task. Vickers (1979) has noted that under these circumstances, one could compute confidence based on the difference in activation of alternative accumulators at the moment of response selection, and this would lead to correct predictions for the confidence findings (a) and (b) above. In the accumulator model, longer correct RTs would be associated with cases in which counts arrive relatively slowly at the correct accumulator, allowing more time for activation to build up at the incorrect alternative, leading to a smaller relative difference that would be the case for faster responses. Thus, the evidence difference would be greater for faster responses. On the other hand, when responding is based on elapsed time, more time would lead on the average to a larger difference in activation of the two accumulators, accounting for the increase in confidence with time in this case.

When juxtaposed with the reasons given above for thinking that some form of competitive mechanism must be used, Vickers's (1979) argument poses a dilemma within the classical framework: Some of the evidence seems to suggest that relative evidence is the basis for responding in RT tasks, but other evidence suggests that the absolute criterion might be used. The use of lateral inhibition in the leaky, competing accumulator model provides a way out of this dilemma. The model uses lateral inhibition to account for relative evidence effects while still using an absolute criterion for the timing of decisions in standard RT experiments. This allows the model to account for (a) and (b) above in the same way as the classical accumulator model, that is, by assuming that the difference in activation between the two accumulators at the time the choice is made corresponds to the degree of confidence. To show this, we took the two-alternative version of our model, using leakage and inhibition parameters in the range used in the fits reported above, and examined the activation difference between the two accumulators (a) as a function of time to reach criterion in a standard RT condition and (b) as a function of processing time allowed in a time-controlled condition. As with the pure accumulator model, we found that the difference in activation of the two accumulators decreases for larger RT in the first case but increases for longer processing time in the second case. Thus, our leaky, competing accumulator model may inherit the advantages of both the relative-evidence and the absolute-evidence models. Like the former, it accounts for effects of competitors on RT, but like the latter, it allows a simultaneous account for confidence and RT data. This is because the suppressive effect of one alternative on another is only partial. Such a partial suppressive effect was also essential in Heuer's (1987) account of the interaction between stimulus discriminability and response compatibility. His model uses negative coupling equivalent to feed-forward rather than lateral inhibition but shares with our approach the idea that the inhibitory effect may be partial, interpolating between the classical accumulator and the classical random walk.

Conclusion

We suggested at the outset of this article that the leaky, competing accumulator model might be a useful extension of the stochastic information accumulation framework, allowing (a) for imperfection in the information accumulation process and (b) for a choice mechanism sensitive to relative evidence that generalizes readily to any number of alternatives. The analysis presented above suggests that this framework is quite viable as a way of addressing a wide range of relevant findings from perceptual choice experiments. As noted earlier, the work of Smith (1995) and of Busemeyer and Townsend (1993; Roe, Busemeyer, & Townsend, 2001; see also Diederich, 1997) suggests that the principles we have explored here also have relevance across a number of other domains. We also noted earlier that there is evidence from neurophysiology of a competition process in attentional selection tasks (cf. Figure 2 from Chelazzi et al., 1993). Partly on the basis of this evidence, Desimone and his colleagues have proposed that attentional effects on response selection are mediated by competitive interactions between neural representations (Desimone, 1998; Reynolds et al., 1999). Some aspects of such findings have been addressed in models incorporating competition and leakage (Cohen, Servan-Schreiber, & McClelland, 1992; Usher & Niebur, 1996). In general, it appears likely that the principles of leakage and competition apply broadly across a wide range of information-processing tasks and are applicable at both the psychological and neurophysiological levels.

Nevertheless, there are important elements of other approaches within the stochastic information accumulation framework that cannot be neglected. The assumption that there is variance in the drift rate from trial to trial within the same condition of an experiment is very likely to be correct in many cases. Furthermore, recent work in models incorporating diffusion with drift variance (e.g., Ratcliff & Rouder, 1998; Ratcliff et al., 1999) shows that a stochastic information integration model that includes variability in the starting point of the diffusion process (Laming, 1968), as well as drift variance, can readily incorporate a number of apparently divergent patterns of behavior with relatively few assumptions. Other factors include drift biases (Ratcliff, 1985, or equivalently, parallel drift of the response boundaries during the information integration process, Ashby, 1983) and self-regulation of response criteria on the basis of confidence (Vickers, 1979). In our view, the most satisfactory model is likely to be one that combines these additional principles with those captured by the leaky, competing accumulator model. A model integrating all of these principles would hold great promise for addressing a very broad range of findings within a single, unified account.

References

Abbott, L. F. (1991). Firing rate models for neural populations. In O. Benhar, C. Bosio, P. Giudice, & E. Tabet (Eds.), *Neural networks: From biology to high energy physics* (pp. 179–196). Pisa, Italy: ETS Editrice.

Ahmed, B., Anderson, J. C., Douglas, R. J., Martin, K. A. C., & Nelson, C. (1995). Comparison of current-discharge relationships of pyramidal neurons from the visual-cortex—In-vitro and in the anesthetized cat. *Journal of Neurophysiology, 485,* 10–11.

Ahmed, B., Anderson, J. C., Douglas, R. J., Martin, K. A. C., & Whitteridge, D. (1998). Estimates of the net excitatory evoked by visual stimulation of identified neurons in cat visual cortex. *Cerebral Cortex, 8,* 462–476.

Amit, D. J. (1989). *Modeling brain function: The world of attractor neural networks.* Cambridge, England: Cambridge University Press.

Amit, D. J., Brunel, N., & Tsodyks, M. V. (1994). Correlations of cortical Hebbian reverberations. *Journal of Neuroscience, 14,* 6435–6445.

Amit, D. J., & Tsodyks, M. (1991). Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate—Spikes, rates and neuronal gain. *Network, 2,* 259–273.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Ashby, F. G. (1983). A biased random-walk model for 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 stochastic models of choice. *British Journal of Mathematical and Statistical Psychology, 18,* 207–225.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory. *Psychological Review, 100,* 432–459.

Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review, 101,* 446–469.

Chelazzi, L., Miller, E., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363,* 345–347.

Christie, L. S., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics, 18,* 89–112.

Cohen, J., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes. *Psychological Review, 97,* 332–361.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review, 99,* 45–77.

Cohen, J., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology, 105,* 239–269.

Corbett, A. T., & Wickelgren, W. A. (1978). Semantic memory retrieval: Analysis by speed and accuracy trade-off functions. *Quarterly Journal of Experimental Psychology, 30,* 1–15.

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

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London, Series B, 353,* 1245–1255.

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

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

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A., & Suarez, H. (1995). Recurrent excitation in neocortical circuits. *Science, 269,* 981–985.

Douglas, R. J., & Martin, K. A. (1990). Neocortex. In G. M. Shepherd (Ed.), *The synaptic organization of the brain* (pp. 389–438). New York: Oxford University Press.

Edelman, G. M. (1982). Group selection and phasic reentrant signaling: A theory of higher brain function. In G. M. Edelman & V. B. Mountcastle (Eds.), *The mindful brain* (pp. 51–100). Cambridge, MA: MIT Press.

Erlick, D. E. (1961). Judgements of the relative frequency of a sequential series of two events. *Journal of Experimental Psychology, 62,* 105–112.

Ermentrout, B. (1994). Reduction of conductance-based models with slow synapses to neural nets. *Neural Computation, 6,* 679–695.

Georgopoulos, A. P., Kettner, R. E., & Schwartz, A. B. (1986). Neural population coding of movement direction. *Science, 23,* 1416–1419.

Granit, R., Kernell, D., & Shortess, K. S. (1963). Quantitative aspects of repetitive firing of mammalian motoneurons, caused by injected currents. *Journal of Physiology, 168,* 911–931.

Gratton, G., Coles, M. G., Sirevag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 331–344.

Grice, G. R. (1972). Application of a variable criterion model to auditory reaction time as a function of the type of catch trial. *Perception & Psychophysics, 12,* 103–107.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding. *Biological Cybernetics, 23,* 121–134.

Grossberg, S. (1978). A theory of visual coding, memory and development. In E. L. J. Leeuwenberg & H. F. Buffart (Eds.), *Formal theories of visual perception* (pp. 7–26) New York: Wiley.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science, 11,* 23–63.

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

Hebb, D. (1949). *The organization of behavior.* New York: Wiley.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Redwood, CA: Addison-Wesley.

Hess, R., Negishi, K., & Creutzfeld, O. (1975). The horizontal spread of intracortical inhibition in the visual cortex. *Experimental Brain Research, 22,* 415–419.

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

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11–26.

Hinton, G., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.

Hinton, J., Liversedge, S. P., & Underwood, G. (1998). Neighborhood effects using a partial priming methodology: Guessing or activation? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1294–1305.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, USA, 79,* 2554–2558.

Humphreys, M. S., Wiles, J., & Dennis, S. (1994). Toward a theory of human memory—Data-structure and access processes. *Behavioral and Brain Sciences, 17,* 655–667.

Jagadeesh, B., Gray, C. M., & Ferster, D. (1992). Visually evoked oscillations of membrane potential in cells of cat visual cortex. *Science, 257,* 552–555.

Jamieson, D. G., & Petrusic, W. M. (1977). Preference and time to choose. *Organizational Behavior and Human Performance, 19,* 56–67.

Kim, C., & Myung, I. J. (1995). Incorporating real-time random effects in neural networks: A temporal summation mechanism. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Meeting of the Cognitive Science Society* (pp. 472–477). Hillsdale, NJ: Erlbaum.

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

Lacouture, Y., & Marley, A. A. J. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science, 3,* 401–433.

Laming, D. R. J. (1966). A new interpretation of the relation between choice reaction time and the number of equiprobable alternatives. *British Journal of Mathematical and Statistical Psychology, 19,* 139–149.

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

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

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

Mason, A., & Larkman, A. (1990). Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. *Journal of Neuroscience, 10,* 1415–1428.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology, 21,* 398–421.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology, 23,* 558–614.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McClelland, J. L. (1991). Stochastic interactive activation and the effect of context on perception. *Cognitive Psychology, 23,* 1–44.

McClelland, J. L. (1993). Towards a theory of information processing in graded, random, interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention & Performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* (pp. 655–688). Cambridge, MA: MIT Press.

McClelland, J. L. (2001). Failures to learn and their remediation: A Hebbian account. In J. L. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 97–121). Mahwah, NJ: Erlbaum.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 634–644.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effect in letter perception. *Cognitive Psychology, 23,* 1–44.

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science, 2,* 387–395.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology, 118,* 346–373.

Merkel, J. (1885). Die zeitlichen verhaltnisse der willensthatigkeit [The timing of the volitional process]. *Philosophische Studien, 2,* 73–127.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16,* 5154–5167.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing informaiton. *Psychological Review, 63,* 81–97.

Miyashita, Y. (1988). Neuronal correlate of visual associative long term memory in primate visual cortex. *Nature, 335,* 817–820.

Movellan, J. R., & McClelland, J. L. (2001). The Morton-Massaro law of information integration: Implications for models of perception. *Psychological Review, 108,* 113–148.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantrowitz (Ed.), *Human information: Tutorials in performance and cognition.* (pp. 41–82). Hillsdale, NJ: Erlbaum.

Pachella, R. G., & Fisher, D. (1972). Hick's law and the speed–accuracy trade-off in absolute judgement. *Journal of Experimental Psychology, 92,* 378–384.

Petrusic, W. M., & Jamieson, D. G. (1978). The relation between the probability of preferential choice and the time to choose changes with practice. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 471–482.

Pietsch, A., & Vickers, D. (1997). Memory capacity and intelligence: Novel techniques for evaluating rival models of a fundamental

information-processing mechanism. *Journal of General Psychology, 124,* 229–339.

Pike, A. R. (1968). Latency and relative frequency of response in psychophysical discrimination. *British Journal of Mathematical Psychology, 21,* 161–182.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

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 modeling accumulation of information when the rate of the accumulation changes over time. *Journal of Mathematical Psychology, 21,* 178–184.

Ratcliff, R. (1981). A theory of order relation in perceptual matching. *Psychological Review, 88,* 59–108.

Ratcliff, R. (1985). Theoretical interpretations of the 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, 92,* 212–225.

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., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104,* 319–343.

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

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

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

Reed, A. V. (1976). List length and the time course of recognition in immediate memory. *Memory & Cognition, 4,* 16–30.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19,* 1736–1953.

Ricciardi, L. (1977). *Diffusion processes and related topics in biology.* New York: Springer-Verlag.

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.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processes. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 45–77). Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representation by error propagation. In D. L. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–363). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & The PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition,* (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press.

Salthouse, F. A. (1981). Converging evidence for information processing stages. *Acta Psychologica, 47,* 39–61.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical T organization. *Current Opinion in Neurobiology, 4,* 569–579.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review, 102,* 567–593.

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

Smolensky, P. (1986). Neural and conceptual interpretation of PDP models. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstrucutre of cognition* (Vol. 2, pp. 390–431). Cambridge, MA: MIT Press.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience, 13,* 334–350.

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

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

Tanaka, K., & Saito, Y. (1991). Coding visual images of objects in the inferotemporal cortex of macaque monkey. *Journal of Neurophysiology, 66,* 170–189.

Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review, 81,* 75–98.

Townsend, J. T. (1981). Some characteristics of visual whole-report behavior. *Acta Psychologica, 47,* 149–173.

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

Usher, M., & Cohen, J. (1999). Short-term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 78–81). Berlin, Germany: Springer-Verlag.

Usher, M., & McClelland, J. L. (1994, July). *Stochasticity, non-linearity and competition in reaction time models.* Paper presented at the meeting of the Mathematical Society, Seattle, WA.

Usher, M., & McClelland, J. L. (1995). *Time course of perceptual choice* (Tech. Rep. No. PDP.CNS.95.5). Pittsburgh, PA: Carnegie Mellon University, Center for the Neural Basis of Cognition.

Usher, M., & Niebur, E. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience, 8,* 311–327.

Usher, M., Olami, Z., & McClelland, J. L. (2001). *Hick's law in a stochastic race model with speed–accuracy tradeoff.* Manuscript submitted for publication, Birkbeck College, University of London, London, England.

Usher, M., Stemmler, M., Koch, C., & Olami, Z. (1994). Network amplification of local fluctuations causes high rate variability, fractal firing patterns and oscillatory local field potentials. *Neural Computation, 6,* 795–836.

Usher, M., Stemmler, M., & Olami, Z. (1995). Dynamic pattern formulation leads to 1/f noise in neural populations. *Physical Review Letters, 74,* 325–329.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 582–600.

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.

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

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

Vickers, D. (1995). The frequency accrual speed test (fast)—A new measure of mental speed. *Personality and Individual Differences, 19,* 863–879.

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

Vickers, D., & Lee, M. D. (2000). Dynamic models of simple judgments: II. Properties of self-organizing PAGAN (Parallel, Adaptive, Generalized Accumulator Network) model for multi-choice tasks. *Non-Linear Dynamcs, Psychology and Life Sciences, 4,* 1–31.

Wald, A., & Wolfowitz, J. (1948). Optimum characteristic of sequential probability ratio test. *Annals of Mathematical Statistics, 19,* 326–329.

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

Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes–no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 189–202.

Wilding, J. M. (1974). Effects of stimulus discriminability on the latency distribution of identification responses. *Acta Psychologica, 38,* 483–500.

Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biological Cybernetics, 12,* 1–24.

Wolfram, S. (1988). *Mathematica: A system for doing mathematics by computer.* Reading, MA: Addison-Wesley.

Woodworth, R. S. (1938). *Experimental psychology.* London: Methuen.

Appendix A

Numeric Integration of Nonlinear Equations

We show in Figure A1 numerically integrated trajectories, for Equation 2, according to two schemes. In the first scheme, which corresponds more closely to the underlying physiology, the leakage (λ) and the recurrent excitation (α) terms are kept separate. Thus, when the x_i variable becomes negative, the terms $\alpha f(x_i)$ and $\beta x f(x_j)$ vanish and only the leakage term λx_i is integrated with the noise. Because in the principal neurons in the neocortex, the synaptic decay (characteristic for glutamate synapses) is very short (5 ms), the value of λ on the time scale τ = 100 ms used in our simulations is very fast, λ = 8.7, corresponding to a decay by a factor of 0.13 within one iteration step of 10 ms (dt/τ=0.1). This results in a very fast decay of the activation toward zero, when $x_i < 0$. At positive values of x_i, the decay is determined by the value of $k = \lambda - \alpha = 0.2$, which is much slower, corresponding to the time constants in the time–accuracy curves.

According to the second scheme, (Eq. 4) which we adopt as a simplified approximation to the first one, leakage and recurrent excitation are lumped together. The fast decay at negative x values is approximated by resetting x to zero at every iteration when a negative value is obtained. Because x is then bounded $x \geq 0$, one can simplify the equations by eliminating one variable, as leakage and recurrent excitation are lumped together. As shown in Figure A1, the numerically integrated trajectories obtained by using each of these two schemes are virtually indistinguishable.

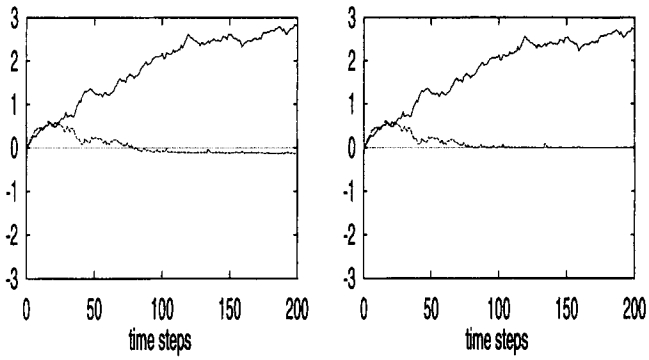

Figure A1. Trajectories of the two response units for a leaky, competing accumulator model, according to two schemes for incorporating the nonlinearity. In both schemes, the same noise is generated according to the same seed, and the inhibition parameter is β = .4. The left panel depicts the nonlinear model with κ = 8.7 and α = 8.5 (K = .2). The right panel depicts the approximation with bounded activation at $x > 0$ and K = .2.

Appendix B

Derivation of d' and Probability Correct, P

Assume a binary choice with two target stimuli A and B. Denote by x the difference in activation accumulated by the two response units (positive x reflects A choices, and negative x reflects B choices). Due to noise, the activation for A and that for B stimuli at time t after the input presentation are both normally distributed random variables. The means of their distributions $\mu(t)$ and their standard deviations are given in accordance to the solution of the Fokker–Plank equation (see Ricciardi, 1977) by Equation 8. Thus, d', the separation between the two distributions, is $d' = 2\mu(t)/SD(t)$. The probability for a correct response is the integral from zero to infinity over the A-target Gaussian (if $x > 0$, A is chosen), which relates to d' as

$$P(t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left[-\frac{[x - \mu(t)]^2}{2SD^2(t)} \right] dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-d'/2}^\infty \exp(-x^2) dx. \quad \text{(B1)}$$

The probability corrected for guessing is

$$P_c = \frac{2}{\sqrt{2\pi}} \int_0^{d'/2} \exp(-\chi^2) dx. \quad \text{(B2)}$$

Appendix C

Exponential Behavior in d' at $t > 1/\lambda$

In Figure 6, we have shown that for $Kt > 1$, $d'(t)$ for the Ornstein–Uhlenbeck (OU) process is very close to an exponential. A more convenient way to test graphically the fit to the exponential is to use a log-linear graph. If some function $g(t)$ approximates an exponential approach to asymptote of the form $1 - \exp(-Kt)$, then $-\log[1 - g(t)]$ will be linear with time. Thus, the adequacy of the exponential approximation can be assessed by inspecting the linearity of $-\log[1 - g(t)]$. In Figure C1, we display $-\log(1 - d'(t)/d_{asy})$ for the OU diffusion and for the diffusion-with-drift-variance (DDV) model. The graph shows a linear trend for the OU process. This can be understood by observing that for $t > 1/K$, d' in Equation 9 is dominated by the numerator leading to an exponential approach to asymptote. Clearly, the approach to asymptote is not exponential in the DDV case.

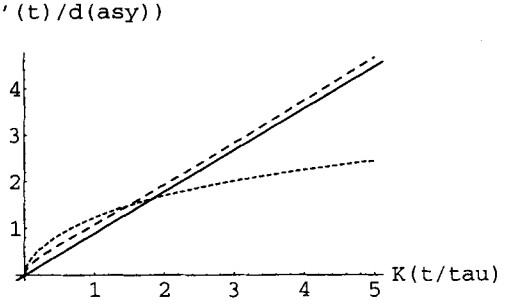

Figure C1. Time–accuracy curves for the Ornstein–Uhlenbeck (OU) model (dashed line) and the diffusion-with-drift-variance (DDV) model (dotted line). The magnitude of $-\log(1 - d'(t)/d_{asy})$ is plotted as a function of time. d_{asy} is shown on the figure as d(asy). The stimulus discriminability was $d_{asy} = 2$.

Appendix D

Transmitted Information

The amount of transmitted information for a binary choice without bias is (see, e.g., Hick, 1952):

$$I(t) = P(t) \log_2[P(t)] + [1 - P(t)] \log_2[1 - P(t)] + 1, \quad \text{(D1)}$$

where $P(t)$ is the probability for a correct response after a time t. At zero time, this probability reflects guessing, $P(0) = .5$ and therefore (Equation D1) $I(0) = 0$. At short times, t, the transmitted information, $I(t)$ can be approximated by developing the formula in Equation D1, with respect to a small increment, δP, from guessing, $P = .5 + \delta P$, and expanding Equation D1 in a power series around $P = .5$, which results in a bilinear formula in δP (first-order terms cancel):

$$I(t) = 2.89(\delta P)^2 + O[(\delta P)^3].$$

Therefore, any formula for $P(t)$ (or $d'[t]$; at small t, $P[t]$ and $d'[t]$ are linearly related), which is linear in t around $t = 0$, such as an exponential function, will result in a bilinear form of $I(t)$, illustrated by the long-dashed curve in Figure 7. It is thus interesting that the square-root dependence of $P(t)$, according to the diffusion and the OU processes, conspire with the bilinearity of the transmitted information formula, to result in a linear increase in $I(t)$ with processing time. This linear approximation is valid only at short times, after which saturation effects appear.

Appendix E

Metropolis Algorithm for Optimization

The aim of the algorithm is to find a set of parameters that generate an optimal match between a predicted model and experimental data. A swap of the algorithm begins with a set of parameters for which it computes the model prediction. The prediction is matched against the data using a cost function (sensitive to the discrepancy, see below). At the next swap, new values for the parameters are simultaneously selected by using independent Gaussian random variables, distributed around the set of parameters that obtained the best fit (in terms of the cost function) from all the previous swaps. At every 100 swaps, the standard deviations for the Gaussian distributions are made smaller by a factor of 0.99. Thus, at the beginning, the algorithm makes large random steps in its search of best parameters, and the search is refined around the best previous fit as the algorithm progresses.

The algorithm was tested with third-order polynomial functions (using a least squares cost function) and was shown to converge within 0.01 of the values of its coefficients, beginning from random values within less than 50,000 swaps. Consistent results have been obtained by using a Levenberg–Marquardt algorithm for optimization in Mathematica (Wolfram, 1988).

In fitting the time–accuracy data from Experiment 1 to the Ornstein–Uhlenbeck and diffusion-with-drift-variance models, 10,000 swaps of the algorithm were used to find optimal parameter values. For both models, the predicted accuracy at each time point could be deterministically calculated.

In fitting the latency probability (LP) data from Vickers (1970), predicted values cannot be analytically computed. Instead, time-consuming and intrinsically noisy simulations had to be performed instead. Within each swap, 500 simulation trials were performed for each of the 10 drift values (corresponding to the five length conditions for both correct and

(*Appendix continues*)

incorrect responses), except for the 3 most negative drift values in which the number of trials was larger (1,000, 2,000, and 3,000, respectively) to compensate for the otherwise lower statistical power at low response probability. At each swap (performed with a set of parameters), the accuracy, the mean, and the standard deviation of the reaction time (RT) were compared with those of the data generated by the participant and were used to calculate the cost, as indicated below.

In the time–accuracy optimizations, the cost function was the conditional probability, $P(\text{data}|\text{model})$, which the algorithm tried to maximize. An alternative cost function was the minimization of the euclidean distance (least squares method) between the model and the data.

In the LP optimization, the cost function was a weighted euclidean distance between the model prediction and the data, for the accuracy (P), the mean RT ($\langle \text{RT} \rangle$), and the standard deviation of the RT. Explicitly, the cost function used was

$$400(P_{\text{model}} - P_{\text{data}})^2 + \frac{(\langle \text{RT} \rangle_{\text{model}} - \langle \text{RT} \rangle_{\text{data}})^2}{SD^2_{\text{data}}} + 0.0002(SD_{\text{model}} - SD_{\text{data}})^2.$$

The weight coefficients were chosen so as to bring the three terms to approximately similar scale while emphasizing the accuracy and the mean RT.

Received July 20, 1999
Revision received August 29, 2000
Accepted August 29, 2000 ■