A Comparison of Two Response-Time Models Applied to Perceptual Matching

Trisha Van Zandt
The Johns Hopkins University

Hans Colonius
Universität Oldenburg

Robert W. Proctor
Purdue University

Running Head: TWO RESPONSE-TIME MODELS

Address correspondence to:

Trisha Van Zandt
Department of Psychology
The Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218

Abstract

Two models, a Poisson race and a diffusion, are fit to data from a perceptual matching task. In each model, information about the similarity or difference between two stimuli accumulates toward thresholds for either response. Stimulus variables are assumed to influence the rate at which information accumulates, and response variables are assumed to influence the level of the response thresholds. Three experiments were conducted to assess the performance of each model. In Experiment 1, observers performed under different response deadlines; in Experiment 2, response bias was manipulated by changing the relative frequency of *same* and *different* stimuli. In Experiment 3, stimulus pairs were presented at three eccentricities: foveal, parafoveal, and peripheral. We examined whether the race and diffusion models could fit the response time and accuracy data through changes only in response parameters (for Experiments 1 and 2) or stimulus parameters (for Experiment 3). Comparisons between the two models suggest that the race model, which has not been studied extensively, can account for perceptual matching data at least as well as the diffusion model. Furthermore, without the constraints on the parameters provided by the experimental conditions, the diffusion and the race model are indistinguishable. This finding emphasizing the importance of fitting models across several conditions and imposing logical psychological constraints on the parameters of models.

A Comparison of Two Response-Time Models Applied to Perceptual Matching

For close to fifty years, response time (RT) studies have been a major focus of attention in cognitive psychology (Hick, 1952; Hyman, 1953; Luce, 1986). Over this time, a great deal has been learned about how performance in cognitive tasks changes with such factors as stimulus intensity, response bias, and so forth. The relationship between RT and other behavioral variables, such as accuracy, is of considerable interest. For instance, it is well known that a person can decrease RT at the expense of decreasing accuracy; this is the ubiquitous speed-accuracy tradeoff that appears in most, if not all, cognitive tasks (Pachella, 1974).

The most successful models of RT and accuracy, and consequently the speed-accuracy tradeoff, are sequential sampling models. These models assume that, in a choice-response task, an observer engages a process of sequentially sampling from the stimulus that results in a gradual accumulation of information at some late stage of processing. A response is executed when the level of information exceeds some threshold or criterial amount required for that response. The speed-accuracy tradeoff arises as the thresholds are moved. If the thresholds are far from the starting point of the process, more information will be needed to reach them and more time will be required for that information to accumulate. The system will be less likely to accumulate a large amount of erroneous information, and so accuracy will be higher than when the thresholds are closer to the starting point. For closer thresholds, it will not take as long to accumulate the necessary amount of information; consequently, RTs will be faster. However, the probability that erroneous information causes the system to reach threshold will be larger, producing a higher error rate.

Sequential sampling models have been examined within a variety of experimental contexts, including signal detection, psychophysical discrimination, recognition memory, categorization, and perceptual matching. One result of their far-flung success is the general acceptance of the idea that these models provide an adequate description of the relation between RT and accuracy in simple and choice RT tasks. There are two major classes of models that have dominated the literature which will be examined in this paper: random walks and race models. The major difference between these two types of models lies in how information is stored at the critical response selection stage. The race models postulate the existence of "counters," modules that store separately the information for each response. In contrast, the random walk models require only a single counter that stores information that can be either positive or negative. In particular, we will compare the popular diffusion model (Ratcliff, 1978), which can be classified as a type of random walk, and a Poisson race model (Pike, 1973; Townsend & Ashby, 1983).

These models will be applied to data from a perceptual matching task, in which observers are asked to determine if two elements of a stimulus pair are the same or different. The perceptual matching paradigm was chosen to evaluate the accumulator models for two reasons. First, the pattern of RTs and accuracies provides an interesting challenge for modeling. Correct "same" responses are often faster than correct "different" responses. This is called the "fast-'same' effect". Yet, incorrect "different" responses tend to be more frequent than incorrect "same" responses. This is called the "false-'different' effect" (Bamber, 1969; Krueger, 1978). This pattern of RTs and accuracies suggests that fast "same" responses are not due simply to a bias to respond "same" (see, e.g., Proctor & Rao, 1983). If this were the case, "same" responses would be more frequently wrong than "different" responses. A successful model must not only illustrate how "same" responses can be speeded relative to "different" responses, but also how such a result could arise without response biases. However, as will be evident in this study, the fast-"same"/false-"different" pattern is not always obtained (e.g., Farrell, 1985; Krueger, 1978). Viable models of the matching process must also be able to predict the lack of effect of stimulus type on RTs and accuracies.

The second reason to choose perceptual matching is that, in similar paradigms, at least two random walk models have been investigated: Krueger's (1978) noisy operator theory and Ratcliff's (1981) diffusion model. The race model has not, until this point, been applied to data from perceptual matching. Because of the number of and relationships between parameters in the random walk and race models, the fast-"same"/false-"different" pattern of data can be captured quite easily by both types of models. The issue, then, is whether both types of models can accommodate changes in RT and accuracy through reasonable changes in parameters (see Proctor, 1986, for discussion of this issue with respect to the diffusion model). In the present sequence of experiments, we manipulated conditions designed to influence the rate of information accumulation as well as the thresholds for “same”/“different” response selection.

Although both the race and random walk models may capture the relationship between speed and accuracy in a single condition, to be considered adequate models of behavior in choice-reaction tasks, they must also be able to explain differences between conditions through appropriate changes of their parameters. Townsend and Ashby (1983) have called this the “Principle of Correspondent Change.” For example, a race model may be able to fit the patterns in RT and accuracy under different levels of response bias, but perhaps this account requires changes in the parameters associated with accumulation rates. This would be an unsatisfactory fit of the model, because the mechanisms of the process require that thresholds, not accumulation rates, change with bias.

In what follows, we discuss the random walk and race models. We present the random walk models that have already been proposed to account for matching data. We also discuss the Poisson race model in some detail, because it has not previously been considered in this context. Then we present the results from three experiments designed to test the models. Our analyses show that both random walk and race models have some success accounting for the data. In particular, the Poisson race model does at least as well as the diffusion model and provides a computationally simpler characterization of the response-selection process. Although we are concentrating only on perceptual matching in this paper, it should be emphasized that we do not intend our results to be limited to this paradigm. We hope, by presenting a comprehensive description of our modeling efforts, to help guide future research in other choice-reaction tasks.

Models of Information Accumulation

Sequential sampling models have been studied extensively for simple and choice reactions, beginning with the earliest counter models (Audley & Pike, 1965; LaBerge, 1962) and random walk models (Laming, 1968; Link, 1975; Link & Heath, 1975; Stone, 1960). Occasionally the race and random walk classes of models have been pitted against each other (Diederich, 1995; Smith & Vickers, 1989). In the following sections, we present several sequential sampling models of response selection. We focus in particular on the random walk models that have been applied to matching, and race models that have been applied to other types of choice reaction tasks.

Random walk models

Noisy Operator Theory

Krueger's (1978) noisy operator theory proposes that the presence of noise in the information-processing system perturbs stimulus features with some small probability. Some differences will be perceived even between identical stimuli, but, on average, *same* pairs will have fewer perceived differences than *different* pairs. The comparison process must determine if the difference registered by a single glance at the display (one "pass") is large enough to conclude that the pair is different. It counts the number of mismatching features between the two stimuli, rechecking or recounting when the number of mismatches is neither small enough to conclude "same" nor large enough to conclude "different." The process keeps a count of the total number of differences perceived over passes of the display. Because of the tendency for noise to make some *same* pairs look different, "different" errors will occur more frequently, but *different* stimulus pairs will be more likely to be rechecked, resulting in slower "different" RTs: the fast-"same"/false-"different" pattern.

Krueger's model can be characterized as a random walk in discrete state space (the number of mismatches) and discrete time (the number of passes). The noisy operator theory differs from other random walk models in that, on every pass, the distributions of the number of perceived mismatches change according to the number of differences perceived on the previous pass. It is therefore *nonstationary*, and explicit expressions for the RT distributions have not been derived. Krueger (1978) showed good fits of the model to the RT histograms for several experimental conditions in single and multielement matching. The model also predicted the false-“different” pattern, although, in absolute terms, the model predicted accuracies that were higher than those observed.

The Diffusion Model

A diffusion model was proposed by Ratcliff (1978) to explain RTs and accuracy in recognition memory. Since its first appearance, it has been applied to a range of tasks, including visual search (Ward & McClelland, 1989), typing (Heath & Willcox, 1990), detection (Diederich, 1995; Smith, 1995), simple two-choice tasks (Ratcliff, Van Zandt & McKoon, 1999), and matching (Ratcliff, 1981). The diffusion model assumes that the process begins with no information about the appropriate response. Over time, stimulus properties drive the information level up or down, toward one response boundary or another (see Figure 1, top right). When the information level reaches one of the two boundaries, the process ends and the response is selected according to the boundary that was crossed.

This model was originally applied to matching with sequential presentations, in which the two elements of a stimulus pair are presented one at a time, imposing a memory load on the process. The memory representation of the first stimulus (a letter string) is assumed to decay over time. Upon presentation of the second stimulus, the amount of overlap between the stimulus pair is computed. Pairs that are identical will overlap a lot, whereas pairs that are different will not overlap as much. This overlap determines the rate at which evidence “drifts” toward one response boundary or the other. If the overlap is less than some critical value (equal to zero without loss of generality; see Figure 1, top, left), then the state of the system drifts on average in the negative direction, toward the boundary for a “different” response. If the overlap is greater than that critical value, then the system drifts in the positive direction, toward the boundary for a “same” response. To account for the fast-“same”/false-“different” pattern, the model assumes changes in the placement of the response boundaries as well as the critical value for the amount of overlap.

Some important characteristics of the diffusion process should be noted. It is the limiting case for a random walk in which the time steps of the walk and the size of the steps up or down become infinitely small. The position of the diffusion at any point in time is normally distributed with a mean equal to the product of the drift rate and time. In Ratcliff's applications of the diffusion model (e.g., Ratcliff, 1978; Ratcliff, 1981; Ratcliff, Van Zandt & McKoon, 1999), the drift rate is also a random variable, normally distributed with some mean and variance. Each stimulus type (*same* and *different* in the matching paradigm) gives rise to its own drift rate distribution. For example, *same* pairs will give rise to higher mean overlap than *different* pairs. Determining a critical overlap value is therefore equivalent to the placement of a criterion in signal detection theory. The RTs and accuracies predicted by the model are produced not by a single diffusion process, but by a mixture of diffusions with varying drift rates.

The diffusion model has been applied to a wide range of choice-RT tasks, whereas the noisy operator theory is a theory of perceptual matching only. Moreover, the diffusion model is a well–studied and well–understood statistical process, while the noisy operator theory is a nonstationary simulation model for which an explicit mathematical characterization does not exist. Therefore, we focused on the diffusion model in this paper.

Race Models

Race models were first explored by LaBerge (1962) and then by Pike (Audley & Pike, 1965; Pike, 1973). The clearest distinction between a race and a random walk model is the number of mechanisms on which information accumulates. For a random walk, information is summed on a single mechanism and can increase and decrease over a trial. Race models assume that information toward alternative responses accumulates in parallel and monotonically over the course of a trial. The counters storing the evidence may be completely independent (Pike, 1973; Townsend & Ashby, 1983) or correlated to some degree (Mordkoff & Yantis, 1991). The first counter to accumulate information to a threshold "wins" the race and determines the response. To compare between the race and random walk (or diffusion) representations, it is convenient to think of the random walk as reflecting the momentary difference between two counters in a race. The random walk, then, is equivalent to a race between two perfectly and negatively correlated counters.

Instance theory

Perhaps the most familiar example of a race model was presented by Logan (1988; 1992) in his theory of automaticity. Logan proposed that the performance of a task depends initially on an algorithm, or series of steps, that takes the performer from the stimulus to a correct response. Each exposure to a stimulus/response pair results in the storage of an instance of that pair in memory. The completion of the algorithm occurs in parallel with the retrieval of the appropriate response from memory. As more and more instances enter the race, the more and more likely it is that the memory process will finish first. Skilled performance of a task thus develops with practice and relies more heavily on the memory of particular stimulus-response pairs than unskilled performance. Logan (1992) demonstrated good fits of this model to the entire RT distribution and explained how not only the mean RTs decrease as a power function of practice (Newell & Rosenbloom, 1981), but also how the entire RT distribution decreases as a power function of practice. Colonius discussed the mathematical assumptions of instance theory and more clearly laid out the conditions under which the RT distributions can change as they do (Colonius, 1995; Logan, 1995).

Vickers' accumulator model

Another well-studied model is Vickers' (1970; 1979) accumulator model, which has been applied to psychophysical discrimination data in expanded judgment tasks. In these tasks, partial information about a stimulus is presented to an observer at time steps over the course of a trial. The motivation for this paradigm is based on the idea that sequential sampling models accumulate information over time by repeatedly sampling from the stimulus or a mental representation of the stimulus. So, for example, if an observer's task is to determine if a presented tone is of a high or low frequency, and if the perceptual effect of the tone varies according to signal detection theoretic principles, sequentially sampling the tone gives rise to a series of random perceived frequencies. One way to gain experimental control over this process is to explicitly give the observer a sequence of tones, each of a random frequency selected from a distribution with a high or a low mean.

The accumulator model assumes that the presentation of a stimulus sample results in a percept that is compared to an internal referent (a criterion, as in signal detection theory). The difference between the stimulus and the referent is added to one counter if it is positive and to an alternative counter if it is negative. This difference is a random variable, normally (Vickers, 1970, 1979) or exponentially distributed (Smith & Vickers, 1988). The model operates in discrete time, like the standard random walk and the noisy-operator theory, and in continuous state space. This model has been quite successful in accounting for RT distributions, accuracy and confidence judgments in the expanded judgment task.

The Poisson race model

The Poisson race model, the focus of attention in this paper, was originally proposed by Pike (1973), and later generalized by Townsend and Ashby (1983). It assumes that counters are independent and accumulate evidence in parallel over the course of a trial (see Figure 1, bottom half). Evidence arrives at each counter in unit increments and is summed until one counter reaches a threshold level of information. The time between units is an exponentially distributed random variable with some rate that depends on the stimulus. For *same* stimulus pairs, the accumulation rate for the "same" counter will be high, whereas for the "different" counter it will be low. When a *different* stimulus pair is presented, the rate for the "different" counter will be high and the rate for the "same" counter will be low. Because the time between units is exponentially distributed, the accumulation of counts is a Poisson process. The time to select a response is determined by the time required for the fastest counter to reach threshold. It is a continuous time, discrete state process, in comparison to Vickers' (1979) discrete time, continuous state accumulator model and the continuous time, continuous state diffusion model. The race model can readily generate the fast-“same”/false-“different” pattern through the relative differences between rates of accumulation and thresholds of the two counters.

An important aspect of the Poisson race model is the fact that the counters are represented by two renewal processes. A renewal process is one in which the interarrival times (the times between the arrival of information units) are independent and identically distributed. So, the mean and variance of the interarrival times do not change regardless of the duration of the process or the number of units that have arrived. Such processes typically are used to describe patterns of events that occur over time, like light bulb replacements or traffic flow. The discrete state assumption, which follows from the idea that information arrives in units and so an integer number of these are accumulated, is made only for convenience and can be relaxed without affecting any of the discussion and results to follow (see, e.g., Dzhafarov, 1993; Rumelhart, 1970).

The choice of the exponential as the interarrival time distribution is not strictly necessary but can be defended as a starting point. First, the exponential yields the computationally simplest case. Second, as shown by Khintchine (1960) and more generally by Grigelionis (1963), under general conditions the superposition (i.e., the pooling of the output) of a number of renewal processes with arbitrary interarrival time distributions becomes a Poisson process when the number of contributing processes becomes very large. If a counter is thought of as a neural module receiving a flow of information arriving from many concurrent processes, then the Poisson process provides an appropriate description of the behavior of the accumulator mechanism over time. Finally, Ashby (1983) has shown that if the distribution of any processing time component in a sequence is exponentially distributed, and if the other components are stochastically independent and have nondecreasing hazard functions, then the RT hazard function must asymptote to a constant. Most empirical RT hazard functions asymptote to a constant, consistent with the hypothesis that at least one component of the process is exponential.

Besides choosing a different interarrival time distribution, there are various other ways to modify and/or generalize the basic Poisson race model that could be taken into consideration (see, e.g., Cox & Isham, 1980). These include the "delayed" process, in which unknown factors (which might be incorporated into a residual or base time) are included in the first interarrival time, a "dead-time" model, in which each counter is shut down for a (possibly random) amount of time after registering an arrival (Schwarz, 1990), and an imperfect model where with some probability π each counter fails to register an arriving bit of evidence. It turns out that π would not be an observable parameter, because the only change in the model would be that the accumulation rate would be reduced by a factor π. The delayed process is equivalent to the case where the residual time is a random variable (Dzhafarov, 1992) – a reasonable assumption that would create more than a few mathematical difficulties. The plain Poisson race model is very flexible, and these variants of the model should only be considered if suggested by empirical findings.

A final important aspect of the race model is the assumption that the counters act in a stochastically independent way. Interestingly, it turns out that the independence assumption is not that strong at all. Marley and Colonius (1992) have shown that if only the response time of the winner in the race and its identity (i.e., the "same" or "different" counter) is observable, which is the case when only RT and the response executed are collected as data, then any race between imperfectly correlated counters can be represented as a race between independent counters. This nonidentifiability result, due to the limited observability of the race, discourages any elaborate modeling of some form of stochastic dependence between the counters in the absence of empirical evidence to demand it.

The Poisson race (or, simply, race) model is presented in this paper as an alternative to the diffusion model. The race model has not been studied in any detail, and yet has much to recommend it. One primary appeal of the model rests in the relative tractability of its mathematics. The equations required to derive predictions from the diffusion model are complex and it can be difficult to simulate (see Luce, 1986). However, the purpose of this paper is not to pit the two models against each other to show that one is "right" and the other "wrong." Indeed, it may not be possible to do so (Dzhafarov, 1993; Marley & Colonius, 1992). The purpose of this study is to demonstrate that the much discussed but never tested race model is able to account for RT and accuracy data at least as well as the more complex diffusion model, and so it might be an attractive modeling option for perceptual matching, as well as other two-choice reaction tasks. The extent to which we can discriminate between the race and diffusion models will be discussed later.

Having discussed sequential sampling models and their application to perceptual matching, we will now present data from three experiments. These experiments were designed to exploit the "Principle of Correspondent Change" (Townsend & Ashby, 1983). Using this principle, we can evaluate the race and diffusion models by examining how parameters change with changes in experimental conditions, while simultaneously evaluating goodness of fit. The equations describing the density and distribution functions for each model, as well as accuracies, are given in Appendix A.

Experiments

Three experiments were conducted to determine whether the race and diffusion models can produce the observed RT and accuracy patterns when the parameters of the models were constrained to vary in ways appropriate for the experimental conditions. The first experiment examined the behavior of the response thresholds under different levels of speed stress induced by response deadlines. The hypothesis was that the observers should increase their thresholds when the deadline is increased. That is, as much evidence as possible should be accumulated before a response is selected. Consequently, mean RT and accuracy should increase with increasing deadline, as long as the observers know the deadline for a given set of trials. The race model should fit these data through elevations in the response thresholds, and the diffusion model should fit these data through increases in the distances of the boundaries from the starting point of the process. If the deadlines are intermixed within a set of trials, the observers should not be able to adjust response criteria appropriately. The race model and the diffusion model should fit these data with a single set of parameters. For both experimental conditions, the rate at which information accumulates should remain constant over all deadlines.

In the second experiment, bias was manipulated by varying the relative probability of *same* and *different* pairs. When *same* pairs are more likely than *different* pairs, the "same" threshold in the race model should be lowered, or, equivalently, the "different" threshold should be elevated. For the diffusion model, the starting point should move closer to the upper ("same") boundary when *same* pairs are more likely, and closer to the lower ("different") boundary when *different* pairs are more likely. The race model and the diffusion model must be able to fit the data across bias conditions with a single set of rate parameters and by varying the thresholds in the appropriate directions.

The third experiment used different display conditions intended to influence accumulation or drift rates. Letter pairs were presented with different separations, increasing the distances of each letter from the fovea, and increasing the rate at which spurious information accumulates (Eriksen & Schultz, 1977; Krueger & Allen, 1987). Under these conditions, response thresholds should be elevated to avoid errors based on unreliable evidence. This elevation should produce an increase in mean RT, due to the longer time required to accumulate evidence to criterion. An increase in errors should be observed at the widest separation where the effects of perceptual noise, and hence the amount of spurious information, is greatest. For the race model, the rate parameters should therefore increase for the "incorrect" counter and potentially decrease for the "correct" counter. For the diffusion model, the drift rates should decrease as stimulus width increases. When different stimulus conditions are presented in distinct sets of trials, participants will adjust response criteria to compensate (Krueger, 1985; Krueger & Allen, 1987; Proctor, Van Zandt & Watson, 1990). To avoid such shifts of criteria, stimulus conditions were intermixed within sets of trials. The race and diffusion models should therefore fit these data through changes in the accumulation and drift parameters alone, using a single set of thresholds and boundaries for all conditions.

Methods

Participants

Three graduate students from Purdue University, three graduate students from The Johns Hopkins University, and six undergraduate students from Purdue University volunteered to participate in Experiments 1, 2, and 3, respectively. The six Purdue University undergraduates participated to fulfill a course requirement. Each participant was naive to the purposes of the experiments and had normal or corrected-to-normal vision.

Apparatus

All stimuli were generated by and presented on PC-style microcomputers running in text mode. The screens were refreshed at a 60-Hz rate, and the stimuli appeared light on a dark background. Stimulus onset and offset, deadlines, and intertrial intervals, as well as accuracy and RT information, were controlled and recorded by software. All responses were made by pressing the "Z" key in the lower left corner of the computer keyboard with the left index finger (for "same" responses) or the "/" key in the lower right corner of the computer keyboard with the right index finger (for "different" responses).

Stimuli

Stimulus items were the letters "K" and "X" appearing in pairs. Thus, four displays were possible: K K, X X, K X, and X K. Letters were presented in the center of the screen, separated by one or more blank spaces. A "-" centered immediately below the blank spaces served as a warning stimulus and fixation point. All characters were composed of points in a 9x8 matrix, subtending a visual angle of 0.53 deg horizontally and 0.89 deg vertically when viewed from a distance of 33 cm. Experiments 1 and 2 used only foveal displays, where only a single space separated the letters and the entire display subtended 1.59 degrees of visual angle. Experiment 3 also used parafoveal and peripheral displays, in which 8 and 16 blank spaces separated the letters. Parafoveal displays subtended 5.3

degrees horizontally, and peripheral displays subtended 9.5 degrees horizontally. A headrest was used in Experiments 2 and 3 to fix the viewing distance at 33 cm.

Procedure

A trial began with the presentation of the warning stimulus for 50 ms (in Experiments 1 and 3) or 500 ms (in Experiment 2). Immediately after the warning, the stimulus was presented for 50 ms. The screen was then erased until a response was made.

In Experiments 1 and 3, response deadlines were imposed. After each response the RT was compared to the deadline. If the RT was less than the deadline, the screen remained blank until the deadline expired, and then the RT was presented in the center of the screen for 500 ms. The screen was again erased and remained blank for an intertrial interval of 1000 ms. If the RT was over the deadline, the RT and the message "TOO SLOW" was presented immediately after the response for 500 ms. The screen was then erased and remained blank for an intertrial interval of 1000 ms. 1000 ms after the last trial in each block, the participant was told how many responses had been made correctly and executed under the deadline. In Experiment 3, the deadline was always 1000 ms. For Blocked sessions in Experiment 1, when the deadline changed, the deadline for the following block of trials was also presented. The participant was also encouraged to take a short break. The first trial of the next block began 3000 ms after the participant pressed a key indicating readiness to begin. No feedback about response accuracy was presented until the end of a block of trials, when the participant was told how many responses had been executed correctly and how many responses had been executed under the deadline.

In Experiment 2, no deadlines were imposed. After a response, accuracy feedback was provided in the form of a 500 ms tone for incorrect responses. After feedback, a 1000 ms intertrial interval ensued, followed by the warning signal for the next trial.

Design: Experiment 1

Participants performed for 12 sessions on different days. There were six Mixed and six Blocked presentation sessions, which alternated each day, beginning with the Mixed session on the first day. The first two days were considered practice, and practice data were not included in any analysis. Each session was composed of nine subblocks of 108 trials, for a total of 972 trials per session. For Blocked sessions, the nine subblocks were grouped into three blocks of each deadline (500, 750 and 1000 ms). The three deadlines were presented in a different random order each day. Before each block, the participant was told what the deadline would be for the next three subblocks of trials. For Mixed sessions, each subblock was composed of an equal number (36) of 500-, 750- and 1000-ms deadline trials pseudo-randomly intermixed. The participant was not told what the deadline would be on any trial. For both Mixed and Blocked blocks, an equal number (27) of each display type was presented, giving 54 *same* and *different* trials in each block.

To ensure a relatively uniform distribution of display types and deadlines across the trials in a Mixed block, each subblock was further divided into 9 smaller sets of 12 trials. Within each 12-trial set, each display was presented at each deadline in a random order. This strategy insured that participants would not change their expectations toward the end of a block because of, say, a large number of 1000-ms deadline trials that occurred at the beginning of the block. The same strategy was used for Blocked trials, except that the deadline was the same for all trials within the 12-trial set. Each display type was presented three times within each set in a random order.

Design: Experiment 2

Participants performed for five sessions on different days. Each session was composed of three blocks of 360 trials, one at each level of bias (in a different random order each day). Participants were informed of the number of *same* versus *different* pairs before each block. Under "different" bias, 80% (288 of 360) of the trials were *different* pairs. Under no bias, *same* and *different* pairs were equally likely (180 trials each). Under "same" bias, 80% of the trials were *same* pairs. The first and last 36 trials of each block were not included in any analysis. Thus, there were 1440 trials in each bias condition.

Design: Experiment 3

All three display widths were presented within blocks. Participants performed three blocks of 324 trials each day across a period of four days. Each display type and width was presented equally often within a block. The first two blocks performed on the first day were not included in the analyses, for a total of 3240 trials, 1080 trials with each display width.

Results

We will now discuss the results of the experiments in terms of the mean RTs and accuracies. Because these results are not the focus of this article, we do not present the statistics of the tests that we performed. Any effects noted were significant ($p < .05$).

In Experiment 1, mean correct RTs and accuracies increased with increasing deadline in the Blocked condition (386 ms, 443 ms, and 462 ms, and .87, .94, and .95 proportion correct for the 500, 750 and 1000 ms deadlines, respectively, averaged over participants). In contrast, few differences were observed over the three deadlines for the Mixed condition (396 ms, 397 ms, and 396 ms, and .88, .88, and .87 proportion correct for the 500, 750 and 1000 ms deadlines, respectively, averaged over participants).

Only Participant 3 showed a fast-"same" effect of 19 ms, whereas Participants 1 and 2 showed no such effect. Participants 1 and 2 also showed no effect of pair-type on accuracy of responding, whereas Participant 3 did in the Mixed condition. In this condition his false "same" responses outnumbered his false "different" responses, indicating a bias to respond "same" (.21 and .16 for false "same" and "different" responses, respectively). The data for the three participants, therefore, did not show strong effects of pair-type on efficiency of performance. The simultaneous increase in mean RT and accuracy with increasing deadline when deadlines were blocked is consistent with an adjustment of thresholds upward or away from the starting point of the accumulation process.

In Experiment 2, accuracy decreased and mean correct RT for "different" responses increased as bias shifted from predominantly "different" to predominantly "same" (496 ms, 536 ms, and 603 ms, and .98, .94, and .83 proportion correct for the 20%, 50% and 80% bias conditions, respectively, averaged over participants). This pattern was reversed for "same" responses (573 ms, 545 ms, and 474 ms, and .87, .94, and .97 proportion correct for the 20%, 50% and 80% bias conditions, respectively, averaged over participants). These results are consistent with an adjustment of the "same" threshold downward or toward the starting point, and of the "different" threshold upward or away from the starting point, as the number of *same* pairs increases.

Participants 2 and 3 demonstrated an overall fast-"same" effect. The size of this effect was 44 ms for Participant 2 and 35 ms for Participant 3. There was a tendency for a fast-"same" effect for Participant 1 (7 ms). Only Participant 1 showed any main effect of pair type on accuracy; her responses to *same* pairs were more accurate (.95) than her responses to *different* pairs (.92), which indicates a bias to respond "same" and not a false-"different" effect.

In Experiment 3, accuracy decreased while mean correct RT increased with increases in the width of the displays (545 ms, 575 ms, and 591 ms, and .94, .89, and .82 proportion correct for the 1.5, 5.1 and 9.5 degree displays, respectively, averaged over participants). All participants save one showed a fast-"same" effect, which was particularly pronounced for the 1.5 degree displays. All participants save one showed a false-"different" effect at all widths, but it was greatest for the 9.5 degree displays: responses to *same* stimuli were less accurate than responses to *different* stimuli. The increase in correct mean RTs with increasing display width is consistent with a decrease in the rate of accumulation. It cannot be due to an elevation in the thresholds alone because threshold adjustments would lead to a corresponding increase in accuracies. Accuracies were decreasing, however, suggesting that the incorrect accumulation rate was increasing as the correct accumulation rate was decreasing. In the diffusion model, this would correspond to both "same" and "different" mean drift rates converging toward zero.

This pattern of effects is consistent with Krueger's (1978) noisy operator theory.

If participants elevated the "different" threshold relative to the "same" threshold to compensate for the perceptual noise in the wider displays, the "same" responses should be faster than the "different" responses when the evidence is less noisy, as in the 1.5 degree displays. When the quality of the evidence deteriorates as stimuli are moved into the periphery, perceptual noise would tend to make all pairs look different. Thus, an increase in false-"different" errors should be observed. This is consistent with the finding that while the fast-"same" effect was present for the foveal displays, false-"different" errors predominated in the peripheral displays.

Discussion

Over the three experiments, each manipulation (deadlines, bias, and stimulus width) resulted in a unique pattern of effects between mean RTs and accuracies. With increasing response deadline, both mean RTs and accuracies increased. With changes in bias from "different" to "same," mean RTs and accuracies increased for the "different" response and decreased for the "same" response. With increasing stimulus width, mean RTs increased and accuracies decreased. There was no clear fast-"same"/false-"different" pattern in these data, probably due to several factors, including that the fast-"same"/false-"different" pattern is stronger and more reliable with successive, rather than simultaneous, presentation (Proctor, 1981; Proctor & Rao, 1983) and that the participants received extensive practice responding to a small set of stimulus pairs. Therefore, we will not discuss the fast-"same"/false-"different" effect in the analyses to follow.

The stimulus presentation was very brief. This insured that the observers would make a significant number of mistakes and that eye movements would not be possible. This also implies that the process of information accumulation that the observers engaged could not be based on the physical stimulus, but rather on an internal representation of that stimulus. For the fits of the models, we made the simplifying assumption that the stimulus representation did not decay over time. For the race model, the decaying representation is fairly easy to implement, however (Smith & Van Zandt, 1999).

The race model and the diffusion model will now be fit to these data in an attempt to recover the RT distributions and accuracies through adjustments of the appropriate parameters. These fits will be contrasted to fits achieved through inappropriate adjustments of the parameters. It will be demonstrated that both models fit the RT distributions when the parameters are appropriately constrained, although the diffusion model has difficulty fitting both the RT distributions and accuracies.

Fitting the Race and Diffusion Models

In this section, we investigate the diffusion and race models by way of their fits to the entire RT distribution and the accuracy data. The experimental conditions were such that the primary change in parameters should be localized to response thresholds for Experiments 1 and 2, and to accumulation rates in Experiment 3. The general procedure used for fitting the models will first be outlined, followed by the results of the fits and interpretation of the parameters.

Procedure

To fit any model, several choices must be made, including what aspect of the data should be used to fit the model (the "data summary"), what objective function should be minimized, and which algorithm should be used for the minimization procedure. We will discuss each of these choices in turn, and then describe how these choices were applied to the data.

The Data Summary

There are a number of alternatives regarding the data summary to which the models could be fit. For example, Ratcliff (1978) routinely fits the diffusion model to RT quantiles smoothed by an ex-Gaussian distribution. The ex-Gaussian is the distribution of the sum of an exponential and a normal random variable, and it has been shown to provide excellent fits to most empirical RT densities (Hockley, 1984; Ratcliff & Murdock, 1976). To fit the diffusion model, Ratcliff generates its density function for a set of parameters and fits a second ex-Gaussian density to the diffusion density. His strategy is to vary the parameters of the diffusion so that the ex-Gaussian parameters estimated from the diffusion are as close as possible to the ex-Gaussian parameters estimated from the data. This roundabout way of fitting the diffusion has been very successful, and results in a smooth, well-behaved objective function - the sum of squared deviations between the data-derived and diffusion-derived ex-Gaussian parameters.

This procedure was attempted at first, and then abandoned, because the ex-Gaussian did not fit the densities well for several participants. Despite the past success of the ex-Gaussian distribution, this result is not particularly surprising. In Experiment 1, for example, the empirical densities at the 500 ms deadline showed very little skew and the ex-Gaussian is a positively-skewed distribution (see Figures 2 and 3). Fits of the models' density functions were also attempted directly to histogram and quantile estimates of the density function, but good fits were difficult to achieve with these estimates (see Van Zandt, 2000a). Similar problems were encountered using maximum likelihood techniques (which do not require a data summary).

For the experiments presented above, the models' distribution functions were fit to the deciles of the RT distribution (e.g., Logan, 1992). For each condition and display type, the RTs corresponding to the 0^{th}, 10^{th}, ..., 90^{th} and 100^{th} percentiles were computed by linear interpolation. An empirical estimate of the distribution function is therefore provided by $\hat{F}(t_i) = .1i, i = 0, 1, \ldots, 10$ for each quantile $t_0, t_1, t_2, \ldots, t_{10}$. Van Zandt (1999) has shown that fits of distributions to quantiles yield more accurate parameter estimates at smaller sample sizes than do fits to any density estimate, and that the accuracies of the estimates were generally comparable to those obtained using maximum likelihood. Finally, and most importantly for our purposes, fits to the deciles worked when all else failed.

<u>The Objective Function</u>

The objective function to be minimized by the fitting algorithm was the next decision that had to be made, although our choices were limited by the data summary we selected. Among possible options were the sum of squared deviations between the predicted (F) and observed ($\hat{F}$) percentiles ($\sum_i(\hat{F}(t_i) - F(t_i))^2$) and the χ^2 goodness-of-fit statistic ($\chi^2 = N \sum_i \frac{(.10-[F(t_{i+1})-F(t_i)])^2}{F(t_{i+1})-F(t_i)}$, where N is the number of responses used for the estimate $\hat{F}$). In addition, to fit both the distribution and the mean RTs and accuracies simultaneously, some function of the difference between the predicted and observed means and accuracies had to be included in the objective function. For the fits of the distribution, either the sum of squared error or the χ^2 statistic was used, and to this was added the sum of squared errors for the correct and incorrect mean RTs, plus the sum of squared errors for the "same" and "different" response probabilities weighted by the number of trials. The number of error responses was too small to fit the error RT distributions directly.

The probabilities were weighted to assure that a large part of the minimization routine was concentrated on fitting the probabilities. When the χ^2 statistic was used as the objective function for the distributions, weighting the probabilities by the number of observations put the χ^2 statistic and the squared difference of probabilities on an equivalent scale, so that the size of the χ^2 statistic did not swamp the contribution of the probabilities to the objective function.

We considered both the sums of squared errors and the χ^2 statistics for the distribution fits because, for different data sets, one or both of these functions could either be well-behaved or very unstable. That is, while the surface of the function defined by the sum of squared errors for one participant may be relatively smooth and promote rapid convergence to a global minimum and a "good fit," that same objective function for a different participant's data or for a different model may not yield a good fit. The goal of the fits was to find parameter values for each model that resulted in a small χ^2 statistic. Because using the χ^2 statistic itself as an objective function proved to be computationally difficult in some circumstances, the more simple sum of squared errors was used when possible; minimizing this function often minimized the χ^2 statistic also. This was true most often for the race model. When the sum of squared errors did not yield low χ^2 values, the χ^2 statistic itself was minimized. For the diffusion model, direct χ^2 minimization proved to be most effective, because minimizing the sums of squared errors did not generally yield low χ^2 statistics, and did not recover the observed means and accuracies.

Frequently it proved to be impossible to fit both the RT distributions and the response probabilities simultaneously. In this circumstance, the objective functions without the weighted sums of squares of mean RTs and probabilities were fit, resulting in small χ^2 values but significant failures to recover the response probabilities. This phenomenon was extensively explored and will be discussed below.

The Algorithm

The final decision concerned the algorithm used to minimize the objective function. There are many minimization routines we could have used, including STEPIT, simplex, genetic algorithms, simulated annealing, gradient descent, the secant method, and so forth, and we tried several of these. After a great deal of exploration, we devised an iterative simplex routine. A downhill simplex algorithm (Nelder & Mead, 1965; Press, Teukolsky, Vetterling & Flannery, 1992) was the workhorse of the minimization procedure. A simplex is an N-dimensional geometric figure with $N+1$ vertices; for N=2, a simplex is a triangle. In the minimization problem, N is the number of parameters of the model to be fit. Every vertex of the simplex represents a possible set of solutions for the N parameters. The objective function is evaluated at each vertex, and the simplex reflects, expands, and contracts itself around the vertex that gave the lowest value of the objective function. When a minimum is found, the $N+1$ vertices of the simplex converge, and the simplex shrinks to a single point.

For the iterative simplex, an initial set of starting values was selected, and the simplex converged to a (possibly local) minimum value of the objective function. The obtained parameters were then perturbed by some random amount to construct the vertices of a new simplex oriented around the newly-found minimum. The simplex then converged again to a (possibly new) minimum. This process was repeated from as few as 6 to as many as 40 times, and on each iteration, the standard deviation of the perturbations increased. Thus, the volume of the simplex increased with each iteration, covering a larger and larger region of the parameter space. The overall effect is not unlike simulated annealing. After simplex found a local minimum, the iterative process attempted to "shake" the function into a still lower region of the parameter space by increasing the volume of the simplex.

Summary

A great amount of effort was expended arriving at the best-fitting parameter values for each of the models. It is important to recognize that a fit is a fit, regardless of the method used to achieve it. Other routines, objective functions, or overall strategies may produce better fits or worse fits. An important consideration of the fitting method is the properties of the recovered parameters. If the errors between the model and the observed distribution function are normally distributed with equal variance, then methods of least-squares (fits that minimized the sum of squared deviations between the model and the data) assure that the parameter estimates are unbiased (on average, they equal the true parameter values) and have minimum variance. Minimizing the χ^2 and sums of squared errors are both consistent with methods of least-squares, and the quantiles of the empirical distribution are normally distributed random variables. However, we also added the sum of squared error of the mean RTs and accuracies to the χ^2 or sum of squared error, so we cannot be assured of unbiased estimates. We checked the accuracy of the recovered parameters in a simulation study, which we will discuss later, to verify that these procedures were able to recover race and diffusion parameters under a range of conditions.

Another problem is knowing that the sum of squared error has actually been minimized. Searching complicated parameter spaces is something of an art, and there is no way to be sure that the minimum found in such a space is local or global. Therefore, caution must be exercised when attempting to make comparisons between models, especially when one seems to fit the data and the other does not. Just because a good fit has not been found for a particular model does not mean that one does not exist.

We took several steps to insure that we could make comparisons between the models. First, we computed more than one goodness-of-fit statistic. The fact that the χ^2 statistic was minimized in the fitting routine precludes its use as an objective goodness-of-fit measure, although this is fairly standard practice (e.g., Smith & Vickers, 1989). Also, the χ^2 statistic is notoriously sensitive to large sample sizes, and will tend to be "significant" even when no interesting differences exist between observed and expected values. Therefore, we also computed Kolmogorov-Smirnoff and the Root Mean Squared (RMS) statistics. The Kolmogorov-Smirnoff goodness-of-fit test is a nonparametric test of the difference between two distribution functions. Unfortunately, like the χ^2 statistic, it is not optimal for our purposes. The Kolmogorov-Smirnoff test loses power when the parameters of the theoretical distribution are unknown or estimated, and the asymptotic behavior of the statistic is unknown (Stephens, 1983). A better statistic, as we will demonstrate, is the RMS. The RMS is a statistic representing lack of fit, commonly used in structural equation modeling (Browne & Cudeck, 1992; Steiger, 1990). Also, we searched the parameter spaces extensively before each fit, so that the best possible starting values could be used. When possible (for the race model), the maximum-likelihood estimates of the parameter values also were calculated and used as starting values in the minimization.

Finally, as we mentioned above, we conducted a study in which each model was simulated, and then each model was fit back to the data from each simulation. Not only does this procedure verify the accuracy of the fitting routines, it also gives an indication of what a bad fit really is. When a model that is known to be wrong is fit to a data set, the ways that it fails can give insight into the fits of the model to a data set where the underlying process is unknown. For example, if the model consistently fails to fit the accuracies, then that same model's failure to fit the accuracies for an empirical data set might then be diagnostic of a general failure of the model.

For each participant's data, the correct RT quantiles for "same" and "different" stimuli for each condition were computed using linear interpolation as described above. For each fit, the "same" and "different" distributions, mean correct and incorrect RTs, and "same" and "different" response accuracies were fit simultaneously. To each data set the race and diffusion models were fit, and for each model, three conditions were examined. In the first condition, all parameters were free to vary across the three experimental conditions to produce the best fit possible (unconstrained fits). In the second condition, only the threshold parameters were allowed to vary, consistent with experimental conditions that would produce changes in response bias or in the total amount of information required for a decision (Experiments 1 and 2; rate-constrained fits). In the third condition, only the rate parameters were allowed to vary, consistent with experimental conditions that would produce changes in the quality or amount of information available for each response (Experiment 3; bias-constrained fits).

There are seven important parameters in the race model: the two criteria K_S and K_D, one for the "same" and the other for the "different" counter; four rates, λ_{ij}, where $i = S, D$ and $j = S, D$ for the stimulus and counter, respectively; and a residual time T_{er} encompassing those processes of encoding and response not represented in the race model. In the unconstrained fits, there are values for K_S, K_D, λ_{DS}, λ_{DD}, λ_{SD}, λ_{SS}, and T_{er} for each response deadline, bias or stimulus condition (a total of $7 \times 3 = 21$ parameters). For rate-constrained fits, where only bias is free to vary across condition, the rates and T_{er} are the same for each condition and the thresholds K_S and K_D change (a total of $5 + 2 \times 3 = 11$ parameters). For bias-constrained fits, only rates vary across condition while thresholds and T_{er} are constant (a total of $3 + 4 \times 3 = 15$ parameters). In the Mixed condition in Experiment 1, an additional constraint was imposed on the thresholds: in the rate-constrained fits, the thresholds also were held constant (a total of 7 parameters).

There are six critical parameters in the diffusion model. The early stage of the model requires the means of the drift rate for each stimulus type (ξ_D and $-\xi_S$) and their variance (η^2). The diffusion process requires two additional parameters, a and z, the upper boundary to respond "different" and the starting point of the accumulation process. Finally, there is a residual or base time component T_{er}. In the unconstrained fits, there are values for a, z, ξ_D, ξ_S, η^2, and T_{er} for each response deadline, bias or presentation condition (a total of $6 \times 3 = 18$ parameters). For rate-constrained fits, where only a and z are free to vary across condition, the same drift rate values and T_{er} are used for each condition and three pairs of values for a and z are used (a total of $4 + 2 \times 3 = 10$ parameters). For bias-constrained fits, only drift rates vary across condition while a, z, η^2 and T_{er} are constant (a total of $4 + 2 \times 3 = 10$ parameters). In the Mixed condition in Experiment 1, an additional constraint was imposed on a and z: in the rate-constrained fits, these also were held constant (a total of 6 parameters). The diffusion coefficient, s^2, is not an identifiable parameter for individual fits (although the ratio of two diffusion coefficients is), and was set to 0.1 - a value consistent with Ratcliff's (1978) early fits of the diffusion model.

For the diffusion model, all parameters were entered into the simplex simultaneously. For the race model, a grid search was performed over all pairs of threshold values, beginning from $K_S = 1$, $K_D = 1$. The iterative simplex operated at each grid point, and threshold values were chosen that yielded the smallest objective function. The data from each experiment will now be considered separately, and the fits of each model will be discussed in turn. At the end of the discussion of each experiment, the overall findings will be summarized and the two models will be compared. It will be shown that both models fit the RT distributions well, and both models performed well when the parameters were constrained. However, the diffusion model had significant difficulty in explaining both RT distributions and response accuracy.

Experiment 1

In Experiment 1, conditions were such that, if an accumulation process was responsible for the participants' performance, participants should adjust thresholds upward or away from the starting point of the process as deadline increased across blocks. The critical models in this case are those that are rate-constrained. That is, rate should not vary across experimental conditions to account for the data. The rate-constrained fits are therefore "appropriate" for this experiment, and bias-constrained fits are therefore "inappropriate." In the following sections, we present in some detail the results of the appropriate fits, and discuss more briefly the results of the inappropriate and unconstrained fits.

The Race Model

RT distributions. Unconstrained, appropriate and inappropriate fits of the race model are shown for one typical participant (Participant 1) in Figures 2 (blocked deadlines) and 3 (mixed deadlines). It is difficult to distinguish between the three fits on these plots because, overall, the fits were all reasonable and very similar.

The best-fitting parameters and χ^2 statistics for the appropriate fits are presented in Table 1.[1] The parameters are changing in sensible ways across the three deadline conditions: the accumulation rates for the "different" counter when the stimulus pair was *different* (λ_{DD}) are larger than the accumulation rates for the "different" counter when the stimulus pair was *same* (λ_{SD}), and a similar pattern holds for the "same" counters. In the blocked condition, the thresholds K_S and K_D are smallest under the 500-ms deadline, and increase as the deadline increases. In the Mixed condition, a single pair of thresholds accounts for performance under all three deadlines, consistent with the hypothesis that participants would be required to operate under the same thresholds for all deadlines if the deadline for each trial is unknown. In contrast, the parameters recovered from unconstrained and inappropriate fits did not vary in sensible ways given the conditions of the experiment and the hypothesized mechanisms of the race model (see Appendix B).

The total χ^2 statistics for all fits, summed over deadline and display type, are given in Table 2. The differences between the observed and theoretical distributions shown in Figures 2 and 3 are small, but most of the χ^2 statistics show significant differences between the observed and theoretical distributions, even for the appropriate fits (see Footnote 1). This may be because of the sensitivity of the χ^2 statistic to large sample sizes. In Table 3 are presented the Kolmogorov-Smirnoff statistics for all the fits. For unconstrained fits, none of the deviations were large enough to reach significance, despite the large χ^2 values given for these fits in Table 2. The Kolmogorov-Smirnoff statistics show that while there are some failures of the model in the Blocked conditions, there are no interesting differences at all in the Mixed condition. The deviations, while significant, are small in light of the ability of the model to capture the gross aspects of the RT distributions. Recall, however, that the power of the Kolmogorov-Smirnoff test is reduced because of estimated parameters.

One issue regarding comparisons between the different fits is whether the appropriate and inappropriate fits of the race model accounted for any more or less of the variance than the unconstrained fits. The unconstrained model had the most parameters, and so it fit the best. This was confirmed using likelihood ratio tests (Gallant, 1987). In the majority of cases, both the appropriate and inappropriate fits accounted for significantly less variance than the unconstrained fits. Unfortunately, there is no good way to determine whether the appropriate model fit more or less well than the inappropriate model. This is because, although the appropriate and inappropriate models are nested within the more general unconstrained model, they are not nested versions of each other, and neither are they independent. This means that their respective χ^2 statistics are not independent, eliminating the possibility of forming an F-ratio to test for equality. Furthermore, they have different numbers of parameters, and so the χ^2 statistics cannot be directly compared. Similar issues arise when we attempt to compare the fits of race and diffusion models. We will discuss this issue shortly, when we present the results of the RMS analysis.

Means and accuracies. The parameters recovered from the fits were used to calculate the predicted mean correct and incorrect RTs and accuracies. These means were plotted against the observed means, and the result is shown in Figure 4. The correct RTs are fit well by all three models. The accuracies were also well fit by the unconstrained and appropriate models. The inappropriate model, as shown by the open triangles, had more difficulty accounting for the accuracies.

All three models have difficulty predicting incorrect mean RTs. The predicted incorrect mean RTs were consistently slower than the observed incorrect mean RTs. Sometimes the race model is criticized for being unable to predict error RTs that are faster than correct RTs. This criticism is invalid for the models we are investigating here, because neither the thresholds nor the rates are constrained to be equal (Townsend & Ashby, 1983). (Note also the predicted error RTs less than 350 ms which are faster than any of the predicted correct RTs.) The failure of the models to accommodate fast error RTs is not, therefore, attributable to a shortcoming of the race model in general. Rather, it is possible that something else (variability in the starting point of the accumulation levels or "fast guesses") is producing fast incorrect responses that decrease the mean. There is a strong linear relationship ($r = .92$) between observed and predicted error RTs for the appropriate model, a pleasant surprise given that error RTs were not given much consideration in the fits of the model. The inappropriate model produced the error RTs shown as open triangles, which were less well correlated with the observed error RTs ($r = .73$).

The Diffusion Model

RT distributions. Unconstrained, appropriate and inappropriate fits of the diffusion model are shown for Participant 1 in Figures 5 and 6. As for the race model fits, it is difficult to distinguish between the three fits on these plots because, overall, the fits were reasonable and very similar. There are, however, notable failures of the appropriate model.

The best-fitting parameters and χ^2 statistics for the appropriate fits are presented in Table 4. The parameters are changing in sensible ways across the three deadlines in the Blocked condition: while the drift rates are constant, the boundaries a and z are smallest under the 500 ms deadline, and increase as the deadline increases. In the Mixed condition, a single pair of boundary values accounts for performance under all three deadlines. In contrast, the parameters recovered from the unconstrained and inappropriate fits did not vary in sensible ways (see Appendix B).

The total χ^2 statistics for all fits, summed over deadline and display type, are presented in Table 2. Unlike the race model fits, there are few significant differences between observed and theoretical distributions. The Kolmogorov-Smirnoff statistics (see Table 3) show that while the appropriate model failed for Participant 1, there are no large differences between the observed and predicted distributions for Participants 2 or 3. The deviations from the observed distribution for Participant 1 are quite large, and raise some concern about the viability of the appropriate diffusion model for this participant.

Likelihood ratio tests again were used to determine the relative goodness of fit of the constrained and unconstrained models. As for the race model, nearly all of the constrained fits accounted for significantly less variance than did the unconstrained fits. We will discuss the goodness of fit for the appropriate versus the inappropriate fits below, when direct comparisons are made between the race and diffusion models.

Means and accuracies. The parameters recovered from the fits to the distribution functions were used to calculate the mean correct and incorrect RTs and accuracies. These means were plotted against the observed means and the result is shown in Figure 7. As for the race model, the correct RTs are fit well by all three models. Unlike the race model, the observed incorrect RTs are not too fast; the mean predicted and observed incorrect RT is about equal. However, the predicted and observed mean incorrect RTs are not as well correlated for the diffusion model as for the race model (for the race model, $r = .84$, and for the diffusion model, $r = .68$). Unlike the race model fits, there is no significant failure of the inappropriate model for the means. The inappropriate diffusion model does as well as the appropriate diffusion model at predicting the means.

Unlike the race model, the diffusion model fails to predict the accuracy data. Whereas accuracy varied from 75% to 99%, parameters recovered for the diffusion process produced accuracy close to 100% for each participant, condition and model. Because the diffusion model has been quite successful at accounting for accuracy and RT data in the past (Ratcliff, 1978; Ratcliff et al., 1999), we explored this failure of the model at great length. Many different starting values for the parameters were used in the iterative simplex algorithm. "Step-wise" fits were attempted, in which the values of the boundary parameters were fixed to guarantee fits of the accuracies, and the drift rates and variance were then free to fit the distribution. A genetic algorithm, which does not rely on finding good starting values, was also used.[2] Both sums of squared errors and χ^2 statistics were used as objective functions. Other objective functions were tried. For all of these options, it was determined that either the distributions could be fit well at the expense of the accuracies, or the accuracies could be fit well at the expense of the distributions.

To attempt to map out the objective function and determine why the diffusion model fits were failing, we entered the accuracies and the distributions into the overall objective function separately. The distribution objective function was the χ^2 statistic, and the accuracy objective function was the sum of squared differences between the observed and predicted accuracies multiplied by their sample sizes. With equal weights on both the distribution and probability functions, simplex did not minimize on the basis of accuracy. The distributions were fit well and predicted accuracy was always very close to 100%. As the weight given to accuracy increased relative to that given to the distributions, the results remained the same, even the best-fitting parameter values were almost equal, until a critical point. At this point, accuracy was fit perfectly but the distribution fits were very poor. No relative weighting could be determined that resulted in intermediate fits to both probabilities and distributions. The fact that this occurred with several objective functions and several minimization routines suggests that it is not simply the choice of fitting techniques which caused the diffusion model to fail.

Comparing the Race and Diffusion Models

Both the race model and the diffusion model fit the RT distributions well. When the parameters were constrained, the race model accounted for RTs in conditions where the thresholds were both increasing and constant. The diffusion model also fit the RT distributions well when appropriately constrained. The race model had difficulty fitting the incorrect mean RTs, although the predicted values were highly correlated with the observed values. The diffusion model did no better. The most puzzling failure was the diffusion model's inability to predict the accuracy data and the RT distributions simultaneously. Many techniques were used to try to find parameter values for the diffusion model that would accommodate these data but none were successful.

Both models did well overall, and both models had strengths and weaknesses. It is difficult to determine from these data if one model did "better" than another. Similarly, it is difficult to discriminate between the appropriate and inappropriate fits for a particular model. Although the appropriate race model accounted for accuracies and RT distributions, the appropriate diffusion model's fits to the RT distributions were superior to those of the race model, as evidenced by the much smaller χ^2 and insignificant Kolmogorov-Smirnoff statistics. To directly compare the two models, and the appropriate and inappropriate fits, the Root Mean Squared (RMS) was calculated for each fit. To calculate the RMS, the χ^2 statistics for the RT distributions were added to three χ^2 statistics measuring the fits to the accuracies across the three deadlines (an additional six degrees of freedom). The resulting χ^2 variable was divided by the number of data points fit (60 deciles plus 6 accuracies) minus the total number of parameters (which varied according to the model fit). The square root of this statistic is the RMS.

The RMS is an ordinal "lack of fit" measure that can be used to distinguish between nested, nonnested or misspecified models (Browne & Cudeck, 1992; Steiger, 1990). The model with the smallest RMS is selected as the best fitting. It penalizes each model by the number of parameters required, because the value of the RMS increases as the number of parameters increases. If a model fits perfectly, the RMS will equal zero. On average, if the χ^2 statistic is used for the RMS, the RMS should equal 1. Notice also that, by entering the response probabilities into the overall χ^2 calculations for the race and diffusion models, the diffusion model has been severely penalized for its inability to fit the accuracies. The RMSs are shown in Table 5 for each model fit to each participant and each condition. If a model predicts an accuracy of 1.0 for a particular condition, the χ^2 statistic will be either 0 (if no errors were made) or infinite (if any errors at all were made). In this situation, the RMS is given by an asterisk, indicating a number too large to compute. This situation arose only for the race model; although the diffusion model predicted very high accuracies, it did not usually predict perfect accuracy. The race model predicted perfect accuracy when the "incorrect" accumulation rates were very small.

The critical comparisons to be made are between the appropriate and inappropriate models, and between the race model and the diffusion model. The lowest constrained RMS is in boldface font. The diffusion model, penalized for poor fits to accuracy, did not yield RMSs as low as those of the race model. For Participants 1 and 2, the appropriate race model gives the lowest RMS for both mixed and blocked conditions. For Participant 3, the inappropriate race model gave the lowest RMS. However, the differences between the two RMSs is very small (1.46 versus 1.37 for the Blocked condition and 1.28 versus 1.27 for the Mixed condition), and well within one standard error (.0951 versus .0988 for the Blocked condition and .0919 versus .0988 for the Mixed condition).[3] Therefore, it is safe to conclude that, for the race model, the appropriate fits were as good as or better than the inappropriate fits for all participants. This was also the case for the diffusion model; although the inappropriate fits yielded slightly smaller RMSs in the Blocked condition, they were very close to the RMSs for the appropriate fits. Disregarding the poor fits to the accuracies for the diffusion model, both models performed quite well in terms of accounting for the RT data by way of changes in the appropriate parameters. Most importantly, the race model performed at least as well as the diffusion model.

Experiment 2

In Experiment 2, the conditions were such that, if any accumulation process was responsible for the participants' performance, participants should adjust thresholds depending on the probability of *same* or *different* stimulus pairs. For the race model, the "same" threshold should decrease as the probability of a *same* stimulus increases, and this decrease may be offset by a corresponding increase in the "different" threshold. For the diffusion model, the starting point z should move toward the upper boundary a as the probability of a *same* stimulus increases, and this movement may be accompanied by an overall increase in a, resulting in a greater difference between a and the lower boundary 0. For both models, the changes in the threshold and boundary parameters should reverse as the probability of a *different* stimulus increases. As for Experiment 1, the appropriate models are those that are rate-constrained. Rate should not need to vary across experimental conditions to account for the data.

The Race Model

RT distributions. Unconstrained, appropriate and inappropriate fits of the race model are shown for one typical participant (Participant 1) in Figure 8. Unlike the race model fits from Experiment 1, these fits are poor for both the appropriate and inappropriate fits.

The best-fitting parameters and χ^2 statistics for the appropriate fits are presented in Table 6. The parameters are changing in sensible ways across the three bias conditions: the correct accumulation rates are larger than the incorrect accumulation rates, and the "same" thresholds are decreasing as the proportion of *same* pairs increases. At the same time, the "different" thresholds are increasing as the proportion of *same* pairs increases. The parameters recovered from unconstrained and inappropriate fits did not vary in sensible ways (see Appendix B).

For all three bias conditions combined, the χ^2 values are not as small as those obtained for the fits in Experiment 1 (see Table 2). All the combined statistics are significant at a $p < .001$ level. In Table 7 are presented the Kolmogorov-Smirnoff statistics for the differences between the observed and theoretical distributions. Although both the appropriate and the inappropriate fits were poor, as indicated by a large number of significant Kolmogorov-Smirnoff statistics, the appropriate model managed to fit more distributions than did the inappropriate model. Likelihood ratio tests showed that the constrained fits accounted for significantly less variance than the unconstrained fits.

Means and accuracies. The parameters recovered from the fits to the distribution functions were used to calculate the predicted mean correct and incorrect RTs and accuracies. These predicted means were plotted against the observed means and the result is shown in Figure 9. As in Experiment 1, the correct RTs are fit very well. Also, there is a tendency, as in Experiment 1, for the model to predict incorrect RTs that are slower than the data. Unlike Experiment 1, the model had some trouble recovering the accuracies. While the appropriate model predicted the "different" accuracies well, as shown by the open diamonds on the plot, the "same" accuracies were not fit well at all. The unconstrained and inappropriate fits could not recover the accuracy data.

The Diffusion Model

RT distributions. Unconstrained, appropriate and inappropriate fits of the diffusion model are shown for Participant 1 in Figure 10. As for the race model fits, the fits are not as good for either of the constrained models as they were in Experiment 1.

The best-fitting parameters and χ^2 statistics for the appropriate diffusion model are presented in Table 8 for all participants. With constant drift rates, the boundary a remains relatively constant across conditions and z steadily approaches a as "same" bias increases. The χ^2 statistics indicate significant differences overall for all participants (see Table 2), although the individual fits to Participant 2's data did not show any significant differences. The Kolmogorov-Smirnoff statistics in Table 7 indicate poor fits of the model for Participant 3, and for Participant 1's "same" responses under "different" bias, which can easily be seen in Figure 10. Likelihood ratio tests showed that the constrained fits accounted for significantly less variance than did the unconstrained fits.

Means and accuracies. The recovered parameters were used to calculate the predicted mean correct and incorrect RTs and accuracies. The predicted means were plotted against the observed means and the result is shown in Figure 11. The correct mean RTs are fit well by all three models except for the slowest mean RTs, where the model predicts that they should be slower than they are. Unlike previous fits, the correlation between the predicted and observed incorrect RTs for all models is low ($r = .46$). As in Experiment 1, the diffusion model fails to predict the accuracy data.

Comparing the Race and Diffusion Models

Unlike Experiment 1, both the race model and the diffusion model failed to fit the data in significant ways, even when all parameters were free to vary. One aspect of the data, not apparent in Figures 8 or 10, that seemed to contribute to this failure was the fact that the RT distributions crossed. For each participant in several conditions, one distribution from one condition (say, 20% *same*) would initially be greater than another (say, 50% *same*), but later the other distribution (50% *same*) would be greater. Some numerical investigations of the race and diffusion models suggest that the conditions under which crossovers can be predicted are fairly limited, although it appears that both can predict crossovers using reasonable parameter values. For example, for the rate-constrained race model, using rates $\lambda_{SS} = 36$ and $\lambda_{SD} = 4$, if the thresholds for one condition are $K_S = 2$ and $K_D = 6$, and under another condition $K_S = 4$ and $K_D = 2$, then the distributions for correct "same" RTs will cross. However, because the model was required to fit six distributions simultaneously, the parameters may not have been able to vary in ways that could capture these crossovers.

Participants may have changed both rates and bias across the three conditions. Such changes would have provided the flexibility needed to produce crossovers. (Generally, the unconstrained fits of the models produced the observed crossovers.) By necessity the bias conditions were blocked, and the experiment took several days to complete, which would have allowed for such broad parameter changes. Ratcliff (1985) fit the diffusion model to perceptual matching data collected under conditions of bias induced by changing instructions, and his fits also required changes in both rates and thresholds across conditions. One reason, then, why the models could not fit the data well when all parameters were free to vary is that the data are formed from a mixture of several sets of parameters used across blocks and days. Van Zandt and Ratcliff (1995) have examined the behavior of variable parameters on RT distributions and shown that the resulting data need not look at all like the distributions predicted by the process that actually generated the data. It is possible that in the present case, a diffusion or a race was indeed responsible for producing the data, but parameter variability is obscuring the fits of the models.

To compare the fits of the constrained and unconstrained models, as well as the fits of the race and diffusion models, the RMS was calculated as in Experiment 1. The RMS statistics are shown in Table 5. Although the statistics are somewhat larger for these data than for those of Experiment 1, the smallest RMSs are comparable. The appropriate race model performed best for Participants 1 and 2, and the appropriate diffusion performed best for Participant 3. The fact that these models showed the smallest lack of fit indicates that for both the race and the diffusion, the parameters need to change in sensible ways to best account for the data. The race model gave the smallest lack of fit for two of three participants, indicating again that the race model is doing as well as or better than the diffusion model for these data.

Experiment 3

In Experiment 3, the conditions were such that, if any accumulation process was responsible for the participants' performance, correct accumulation or drift rates should decrease as the elements of the stimulus pairs moved away from the fixation point. For the race model, incorrect accumulation rates should also increase as stimulus width increased. The critical models for these data are those that are bias-constrained. Bias should not need to vary across experimental conditions to account for the data.

The Race Model

RT distributions. Unconstrained, appropriate and inappropriate fits of the race model are shown for one typical participant (Participant 6) in Figure 12. All three race models fit the data well, and it is difficult to distinguish between the curves on each graph.

The best-fitting parameters and χ^2 statistics for the appropriate fits are presented in Table 9. With thresholds held constant, the correct accumulation rates decreased as stimulus width increase. The incorrect accumulation rates either changed very little or increased as stimulus width increased. Furthermore, the "different" thresholds K_D were larger than the "same" thresholds K_S. This is consistent with Krueger's (1978) noisy operator theory; because noise tends to make stimuli appear different, the "different" threshold must be elevated to prevent fast, erroneous "different" responses. As in previous fits, most of the χ^2 statistics are signaling significantly poor fits. However, the Kolmogorov-Smirnoff statistics (see Table 10) show few differences between the observed and theoretical distribution functions, with the exception of Participant 4. Most of the likelihood ratio tests indicated that the appropriate and inappropriate fits accounted for significantly less variance than did the unconstrained fits.

Means and accuracies. The parameters recovered from the fits to the distribution functions were used to calculate the predicted mean correct and incorrect RTs and accuracies. These predicted means were plotted against the observed means and the result is shown in Figure 13. As in the previous experiments, the correct RTs are fit very well by all the models. However, the incorrect mean RTs are not predicted by any of the models; there is no correlation between the predicted and observed incorrect mean RTs ($r = .00$). As in Experiment 1, but not Experiment 2, the accuracies are fit well by the unconstrained and appropriate models. The inappropriate model failed to predict accuracies well at all, producing accuracies all along the range from 0.0 to 1.0.

The Diffusion Model

RT distributions. Unconstrained, appropriate and inappropriate fits of the diffusion model are shown for Participant 6 in Figure 14. Each model fits well, and it is difficult to discriminate between them. The best-fitting parameters and χ^2 statistics for the appropriate fits are presented in Table 11. With the upper boundary and starting point held constant, the drift rate parameters for both the *different* and *same* stimulus pairs are decreasing with increasing stimulus width. However, the χ^2 statistics indicate significant failures of the model for almost every fit, as well as overall (see Table 2) The Kolmogorov-Smirnoff statistics on the other hand show that a few, but not all of the fits significantly failed. Likelihood ratio tests showed that for Participants 2 and 5, the appropriate and inappropriate fits accounted for significantly *more* variance than did the unconstrained fits, but the fits for all other participants were worse when the models were constrained.

Means and accuracies. The parameters recovered from the fits to the distribution functions were used to calculate the predicted mean correct and incorrect RTs and accuracies. These predicted means were plotted against the observed means and the result is shown in Figure 15. The correct RTs are fit well by all models including the rate-constrained model. As for the race model, there is little correlation between the predicted and observed incorrect RTs for all models ($r = .25$). As in all previous fits, the diffusion model fails to predict the accuracy data.

Comparing the Race and Diffusion Models

In the fits of the race and diffusion models to the data from Experiments 1 and 2, when parameters were appropriately constrained, fits were better than when they were inappropriately constrained. This was not the case for all participants in Experiment 3. The RMS statistic was calculated for each model and these are shown in Table 5. The RMSs were smaller overall for the race model than the diffusion model for each participant. While the appropriate race model performed best for Participants 1, 2 and 3, the inappropriate race model performed best for Participants 4, 5 and 6. However, the inappropriate race model could not predict the accuracy data, whereas the appropriate model did quite well in this respect. This is indicated by the rather large RMSs for the inappropriate model for Participants 4, 5, and 6, compared to the smaller RMSs for the appropriate model for Participants 1, 2, and 3.

For the diffusion model, the fits actually seemed to improve under the inappropriate model, as demonstrated by smaller χ^2 statistics, fewer significant deviations as indicated by the Kolmogorov-Smirnoff statistics, and smaller RMSs. As in all previous fits, the diffusion model failed to predict the accuracy data.

Summary

In three experiments, conditions were varied such that, in Experiments 1 and 2, the models should be fit well through changes in bias parameters alone, whereas in Experiment 3, the models should be fit well through changes in the rate parameters alone. In Experiment 1, the rate-constrained race model fit the data from all three participants in Mixed and Blocked conditions well. In Experiment 2, none of the models achieved satisfactory fits, suggesting that either the models were inappropriate, or that changes in strategies across blocks of trials resulted in a mixture of processes with different parameter values. Despite the poor fits, the models that allowed for changes in bias gave smaller lack-of-fit measures than the models that allowed for changes in rates. For Experiment 3, the data from half of the participants were fit best by the appropriate race model and those from the other half were fit best by the inappropriate race model. However, the inappropriate model failed to predict the accuracy data. The diffusion models did not give lack-of-fit measures as small as those given by the race models, except for Participant 3 in Experiment 2.

The purpose of these fits was to demonstrate that the race model can fit RT data from a choice-RT task as well as the diffusion model can. Overall, with the exception of the data from Experiment 2, the two models fit the RT distributions well through variations in parameters that were generally appropriate to the experimental conditions. However, only the race model was able to fit the RTs from three conditions in each experiment simultaneously with the accuracies in those conditions.

To determine more precisely the quality of the fits for the different models, it is important to know what the fits would look like when the models are correct or incorrect. For instance, what appears to be a lack of fit of one of the models in a particular condition may in fact be an expected deviation for that model. If, say, fits of a race model to data derived from a race fail to accommodate error RTs due to small sample sizes or to chance, then the failure of the race model to fit error RTs in the data from these experiments should be discounted. Furthermore, it is important to determine, by fitting the diffusion model to data produced by a diffusion, if the inability of the diffusion model to fit the accuracy data is due to the specific procedures used to fit the model or, even worse, an error in the complex routines that calculate probability and the distribution functions. To answer these questions, the four models examined above (rate- and bias-constrained race and diffusion models) were simulated, and observations were collected in three "conditions" designed to produce changes in mean RT and accuracy comparable to those obtained in the present experiments. The empirical distributions were estimated in exactly the same way as for the experimental data, and the four models were fit back to the data from the four simulations using exactly the same procedures as described above.

Simulation Studies

Four questions have been raised that will be answered in this section. The first involves the accuracy of the fitting routines. Because the objective function did not have any of the nice properties that guarantee unbiased and minimum variance parameter estimates, we needed to verify that we could in fact recover accurate parameters. The second involves the diffusion model's surprising inability to account for the accuracies in the three experiments. By fitting the diffusion model to race and diffusion data, we can verify both that the diffusion model can fit RTs and accuracies simultaneously when it is the correct model, and also that when it is fit to the wrong model, its failure to fit RTs and accuracies simultaneously is diagnostic of a general failure of the model. The third question deals with the goodness-of-fit statistics we have used. Of the χ^2, Kolmogorov-Smirnoff and RMS, we need to determine which, if any, accurately reflect good and bad fits. We would like for our goodness-of-fit statistic to diagnose a good fit when the right model is fit to the data, and a bad fit when the wrong model is fit to the data. Finally, we need to know the extent to which the race model is able to fit data from a diffusion process and vice versa. The extent to which we can distinguish between them determines how hard we should try to verify that one or the other is true.

Two race and two diffusion models were simulated in three conditions. The bias parameters of one race and one diffusion model increased across conditions, and the rate parameters of the other race and diffusion models decreased across conditions. The parameter values used for the bias- and rate-constrained simulations are shown in Table 12. For each condition, 800 trials were performed, resulting in sample sizes that were comparable to those obtained from each participant in Experiment 1. These four simulations should be viewed as data from four different experiments. Two of the experiments (1 and 3) required shifts in thresholds, and the other two (2 and 4) required shifts in rates across conditions.

For the rate-constrained models, accuracy and RT increased as "condition" increased. This is comparable to the effects observed with increasing response deadline in Experiment 1. For the bias-constrained models, accuracy decreased with increasing RTs as condition increased. This is similar to the effects observed in Experiment 3 as stimulus width increased. The bias-constrained diffusion model had RTs that are much longer and variable than any observed in the experiments and any observed for the other models (see Figure 19). This is a result of the parameters that were chosen for this model. This difference is important because it will provide a challenge for the other models to fit.

Unconstrained, rate-constrained and bias-constrained fits of the race model and diffusion model to each simulation are presented in Figures 16-19. Each figure shows the fits of the six models to one simulation. The rate-constrained and bias-constrained race simulations are presented in Figures 16 and 17, respectively, and the rate-constrained and bias-constrained diffusion simulations are presented in Figures 18 and 19, respectively. Panel *a* of each figure shows the fits of the race models to the "different" and "same" responses, and Panel *b* shows the fits of the diffusion models to the "different" and "same" responses. Fits of the appropriate and inappropriate models are shown as solid and dashed lines, respectively. The unconstrained fits are shown as dotted lines.

From Figures 16 and 17, it appears that both the appropriate rate-constrained and inappropriate bias-constrained race models fit the data from the rate-constrained race simulations well. The diffusion models did not fit as well, although some of the fits are very close. For the bias-constrained race simulations, the appropriate bias-constrained race model fit whereas the inappropriate rate-constrained race model did not. When the diffusion models were fit to the bias-constrained race simulations, the inappropriate rate-constrained model fit well, but not the appropriate bias-constrained model.

From Figures 18 and 19, it appears that for the rate-constrained diffusion simulations only the appropriate rate-constrained diffusion model fit well, although the inappropriate bias-constrained race model did not do too badly. For the bias-constrained diffusion simulations, the appropriate bias-constrained and unconstrained race model fits are poor. The shape of the fitted race model is an exponential, in comparison to the nonmonotonic and unimodal density of the diffusion model. Both the appropriate and the inappropriate diffusion models fit well.

<u>Do the fitting routines accurately recover the parameters of the model?</u>

The best-fitting parameters recovered for the models are given in Tables 13 and 14 for the race and diffusion model fits, respectively. The parameters recovered for the appropriate models are not exactly equal to those used for the simulations, but they are quite close. This variation from the true parameter values occurs because the model is trying to fit random fluctuations in the data and succeeds. Because the models have only been fit to a single simulation, we cannot determine the degree or extent of bias in the recovered parameters. However, it is clear that the routines we used can find best-fitting parameters that are very close to the true parameters for the race and diffusion models.

<u>Can the diffusion model account for RTs and accuracies simultaneously?</u>

The parameters given in Tables 13 and 14 were used to compute the predicted accuracies, correct and incorrect RTs for the race and diffusion models. These predictions were plotted against the observed values, and the results are shown in Figure 20. The figure shows the predictions generated for the race and diffusion model fits in the left and right panels, respectively. Each fit is identified by whether or not the model is correct (e.g., race fit to race) or incorrect (e.g., race fit to diffusion), and appropriate (e.g., rate-constrained fit to rate-constrained) or inappropriate (e.g., rate-constrained fit to bias-constrained). So, for example, the points marked by circles (Incorrect, Inappropriate) on the left panel show the predictions of the rate-constrained and bias-constrained race models for the bias-constrained and rate-constrained diffusion data, respectively. The open circles in the right panel mark the corresponding predictions of the diffusion model for the race data.

The critical plot is in the upper right. The most important thing to note from this figure is the fits of the correct, appropriate diffusion model to the diffusion data (diamonds). The diffusion fits recovered the accuracy simultaneously with the RT distributions. Furthermore, the incorrect fits of the diffusion models (circles and triangles) to the race data produced predicted accuracies that were near 1.0 for every fit of the diffusion model to race data. The inappropriate diffusion model (squares) also had trouble accounting for accuracy. Therefore, we can conclude that the diffusion model is able to account for RTs and accuracies simultaneously, and that its failure to do so for our experiments is indicative of a more general failure of the diffusion model for these matching data.

The recovered parameters were also used to compute the predicted correct and incorrect mean RTs, and these are plotted against the observed correct and incorrect mean RTs. For the race and diffusion model fits in the center left and right panels, the correct mean RTs were recovered well for all simulations. However, both the race and diffusion fits tended to underestimate slow incorrect mean RTs from the diffusion simulation, regardless of whether or not they were fit by the diffusion model. This finding suggests two things. First, the fitting procedure cannot necessarily recover the incorrect mean RTs even when the model is exactly right. Second, because the models consistently *overestimated* the incorrect RTs from the experimental data, the fast incorrect RTs observed in the experiments may be due to a different process, such as fast guessing, rather than a misspecified accumulator model.

Which goodness-of-fit statistics accurately diagnose the correct model?

The overall χ^2 statistics for each fit are given in Table 2, and the Kolmogorov-Smirnoff statistics are given in Table 15. The χ^2 statistics generally mirror the results obtained by looking at Figures 16-19. According to the χ^2 statistics, when a correct and appropriate model was fit to the simulated data, it fit well. However, other fits were also good. The unconstrained race and diffusion models seemed to fit all four simulations well. The inappropriate bias-constrained race model fit the rate-constrained race simulation (1), and the inappropriate rate-constrained race model fit the bias-constrained race simulation (2). The incorrect bias-constrained race model also fit the bias-constrained diffusion simulation (4). The rate-constrained diffusion model fit all four simulations well, and the inappropriate (and incorrect) bias-constrained diffusion model fit the rate-constrained race well. Likelihood ratio tests were performed for these fits as for the experimental data. In general, only when the correct model was fit to the simulated data was there no significant decrement in fit for the constrained models. The other fits (with the exception of the rate-constrained diffusion to the bias-constrained race simulation) showed significantly less variance accounted for by the constrained fits.

The χ^2 statistics do not generally distinguish between correct and incorrect fits, although all significant χ^2 statistics were associated with incorrect or inappropriate models (a perfect "correct rejection" rate). Examining the Kolmogorov-Smirnoff statistics next leads to very similar conclusions as for the χ^2 statistics. However, the Kolmogorov-Smirnoff statistics show that the bias-constrained diffusion model had significant problems fitting the rate-constrained race simulation. Also, the rate-constrained diffusion model had slight difficulty fitting the rate-constrained race simulation.

The χ^2 and Kolmogorov-Smirnoff goodness-of-fit statistics are consistent, but ambiguous. The rate-constrained race model is well-fit by three of the four models (not considering the unconstrained fits); the bias-constrained race model is well-fit by two of the models; the rate-constrained diffusion simulation is well-fit only by the rate-constrained diffusion model; and the bias-constrained diffusion model is well-fit by three of the models. Fortunately, the RMS statistic (Table 5) clears the situation up a bit. For the race simulations (1 and 2) it is clear that the best-fitting model for each is the appropriately-constrained race model. Similarly, the best-fitting model for the rate-constrained diffusion simulation (3) is the rate-constrained diffusion model. The bias-constrained diffusion simulation (4) is fit best by the bias-constrained diffusion model, although the bias-constrained race model also has a very low (one standard error's difference) RMS. Examining the density functions produced by the race model, however, it is clear that the fits are poor; the bias-constrained race model is an exponential density that fits the tails quite well but cannot accommodate the leading edge of the curve.

In sum, the χ^2 and Kolmogorov-Smirnoff statistics were very similar, although the Kolmogorov-Smirnoff detected some differences that the χ^2 did not. (Note that the Kolmogorov-Smirnoff test does not suffer from lack of power in this case because the parameters of the models are known.) Usually the χ^2 statistic tends to reject models that are correct, especially when the sample sizes are large, instead of failing to reject inappropriate models. Van Zandt (2000a) examined the behavior of the χ^2 statistic for a large number of simulated fits of different models, including the diffusion and the race. She showed that, for fits to the distributions from a single experimental condition, the diffusion model yielded significant ($p < .05$) χ^2 statistics between 7% and 11% of the time (for sample sizes between 500 and 1000), and the race model yielded significant χ^2 statistics between 45% and 49% of the time. Our results are not as bad because the models were constrained by three experimental conditions. This allowed for far more accurate parameter estimation (especially for the race model) than can be obtained from a single condition.

The best diagnostic of fit is the RMS statistic. The smallest RMS statistic for each fit was given by the correct fit. In the one situation where the RMS statistic could be argued to be ambiguous, the alternative model was clearly wrong. The RMS statistic took into account the simultaneous fit to accuracy and RT, and so the poor recovery of accuracy data by inappropriate models helped distinguish between the fits.

To what extent can the race model mimic the diffusion model and vice versa?

As we have noted, there is some potential ambiguity between the race and diffusion models. When all parameters are free to vary, the race model can mimic the diffusion model and vice versa. Examining the density functions, it appears that the rate- and bias-constrained diffusion models can mimic the rate- and bias-constrained race model (Figures 16b and 17b), and the bias-constrained race model can mimic the rate-constrained diffusion model (Figure 18a). However, we have demonstrated that we can disambiguate between these models by considering a number of sources of information simultaneously. First, the models were fit to a number of experimental conditions, and the parameters of the models were constrained by those conditions. Second, a number of different goodness-of-fit statistics were used to discriminate among the fits. Third, the models were expected to fit both the RT and accuracy data simultaneously, and goodness of fit was penalized to the extent that the models failed to do this. The information provided by these criteria, together with an application of the Principle of Correspondent Change, uniquely identifies the correct model for each simulation.

General Discussion

The race model and the diffusion model were compared across three experiments. In the experiments, participants performed a perceptual matching task under a number of conditions designed to influence particular parameters of each model. In Experiment 1, "same" and "different" responses were made under three deadlines, which should have induced participants to change response thresholds. In one condition, the different deadlines were presented within the same block of trials, which should have prevented participants from changing response thresholds. In Experiment 2, the probabilities of *same* and *different* stimulus pairs varied across blocks of trials, which also should have induced participants to change response thresholds. In Experiment 3, stimulus pairs were presented at one of three angles of separation, which should have changed the rates at which information could accumulate toward alternative responses.

Both the race and the diffusion models fit the RT data for Experiment 1. Most importantly, the fits of the models to the data when the rates were constrained across the three deadline conditions were good, and were at least as good as the fits of the model when the rates were allowed to vary. When deadlines were intermixed within blocks of trials, a single set of parameters successfully fit all six RT distributions across the three deadlines. When the fits to accuracy and distributions were considered together, the (appropriate) rate-constrained race model fit the data best. The diffusion model, because it could not predict the accuracy data, did not fit the data as well as the race model.

Both the race and the diffusion models fit the RT data well for Experiment 3. Most importantly, for most participants, the fits of the models to the RT distributions were better when the rates varied across stimulus conditions than when response thresholds varied. When the fits to accuracy and distributions were considered together, half of the participants' data were fit better when rates were constrained and thresholds varied. This was because the fits were penalized by the models' inability to fit accuracy. Overall, the race model fit the data better than the diffusion model, again because of the diffusion models' problems with the accuracy data.

Neither model fit the RT data well for Experiment 2. Part of the problem with these data may be the fact that the RT distributions (for the same responses) crossed in different conditions. Neither model seemed able to predict crossed RT distributions with the constraints imposed on the parameters. One possible reason for the models' failure with these data is the fact that the different bias conditions were presented in different blocks of trials. This could have allowed the participants to engage different strategies under different conditions, resulting in a mixture of processes with different parameters across conditions. Such a mixture also could result solely as a consequence of the differing numbers of same and different trial repetitions across the blocks, since the repetition of events (such as the complete repetition of a same pair or the repetition of a "same" response) likely contributes to changes in accumulation rates across trials (e.g., Hommel, 1998) – a process similar to priming. Consistent with this notion, fits were poor when either the threshold or rate parameters were held constant, but improved when both were free to vary. Despite the poorer fits of the constrained models, the appropriate models showed smaller lack-of-fit statistics than the inappropriate models.

To get a better idea of the ways that the models fail, and to determine how flexible the models are, a simulation study was conducted. The two models were simulated under conditions that mimicked those of Experiments 1 and 3. The threshold parameters varied for conditions in Experiment 1, and the rate parameters varied for conditions in Experiment 3. The race models, when fit to the simulated race data, recovered the original parameters of the simulation fairly accurately. The rate-constrained race model only fit the rate-constrained race simulation data. The bias-constrained race model fit not only the race simulation data but also seemed to fit the bias-constrained diffusion simulation. However, even though the goodness-of-fit statistics were small, close inspection of the densities showed that the race model failed to fit the diffusion simulation.

The appropriate diffusion models fit the diffusion simulation RTs well and fairly accurate parameters were recovered. The rate-constrained diffusion model fit even the race simulation RTs, while the bias-constrained diffusion model fit only the bias-constrained diffusion simulation RTs. Therefore, the models are quite flexible; e.g., the diffusion model can account for RTs simulated by a race model if the parameters of the diffusion model are allowed to vary. Similarly, the race model can account for RTs simulated by a diffusion model if the parameters are allowed to vary. However, when the fits are attempted over a number of experimental conditions, the extent of the models' flexibility decreases. Accounting for both RT and accuracy was also problematic for misspecified models.

A striking feature of the fits to the data from the experiments was the inability of the diffusion model to account for both RT and accuracy. The misspecified models in the simulation studies also failed to recover the accuracy. The major purpose of the simulation studies was to determine that accurate parameters could be recovered for these models and that the fits of the models could reproduce the simulated RTs and accuracies. Therefore, the inability of the diffusion model to recover the accuracies observed in the experiments was not due to any of the routines or algorithms we chose to fit the models. This suggests that the diffusion model may not be appropriate for our matching data.

The purpose of the present study was to demonstrate that the race model should be given equal consideration as the diffusion model. By showing that the race model can fit RT and accuracy data at least as well as the diffusion model, and that the race model can do so by appropriate changes in bias and rate parameters, we have accomplished this goal.

Is there a "correct" model?

An obvious question that could be asked at this point is whether these results indicate that the race model or the diffusion model is more appropriate for these data. That is, is one model true and the other false? Both models are very powerful and can accommodate a wide range of data. Both models fit the data from several experimental conditions with appropriate changes in parameters. It could be argued that the mechanisms underlying both models are used in different circumstances. The diffusion model with upper and lower response boundaries is mathematically equivalent to a race between two correlated diffusions with single response boundaries. One possibility is that observers can switch between strategies of monitoring the difference between two counters (a random walk or diffusion process), or monitoring the absolute levels of each counter (a race process), depending on the task requirements or capacity available to perform the task.

One reason why this question cannot be answered at present is that both the models fit the RT data to a statistically acceptable level (in different conditions), and trying to distinguish between them on the basis of goodness of fit puts too heavy an emphasis on statistical analyses in the model-selection process. As clearly discussed by Roberts and Pashler (2000), the fact that a model can be made to fit a set of data is not strong evidence for that model. Demonstrating that a model fits data does not show what the model cannot do, says nothing about whether the variability of the data would allow the model to be ruled out, and does not show whether the model could have fit a different pattern of data. Furthermore, the fact that one model can be fit to the data does not exclude the possibility that other, quite different models could also be fit to the same data (Van Zandt & Ratcliff, 1995). One way to distinguish between models is by setting up experimental conditions in which one model predicts a different pattern of results than another and collecting data. Several theoretical findings (Dzhafarov, 1993; Marley & Colonius, 1992) suggest, however, that it may be difficult or impossible to find experimental conditions to distinguish between the race and random walk classes of models.

Therefore, it may be inappropriate to ask which model is correct. Both models form frameworks within which hypotheses about response selection mechanisms can be devised and tested. In this capacity, both models allow for clarity of certain ideas and predictions that verbal theorizing alone cannot. If either of the models is as powerful and wide-reaching as the other, then the reasons for selecting one over the other must come down to ease of use. On this basis, the race model is clearly superior to the diffusion process. The expressions for fitting and simulating the race model are far more tractable than those of the diffusion model, even if the drift parameter of the diffusion is constant instead of a normally-distributed random variable. It was much easier to fit and simulate the race models than it was to fit and simulate the diffusion models.

It may be possible to distinguish between the two types of models on the basis of other data, such as response confidence judgments (Smith & Vickers, 1988; Van Zandt, 2000b; Vickers, 1979; Vickers, Caudrey & Willson, 1971), or double responses, where the observer attempts to "undo" the already executed response by executing the other response (St. James & Eriksen, 1992). For the race model, both response confidence and the execution of a double response are assumed to depend on the relative difference between the levels of activation on the two counters. If this difference is very small, response confidence will be small. Further, if the system does not shut down completely after one threshold is exceeded, the likelihood of both counters exceeding their thresholds at nearly the same time will be high when the relative difference between them is small, resulting in a double response. The random walk models cannot account for such data without modification, because the relative difference between the perfectly correlated counters is always the same on every response.

The race model has been criticized on several grounds, one being how it handles error RTs and the other in the behavior of the model as thresholds are increased. It is true that if the rates are equal for the two stimulus types (*same* and *different*, that is, $\lambda_{DD} = \lambda_{SS}$ and $\lambda_{DS} = \lambda_{SD}$, then error response times must always be slower than correct response times (Smith & Vickers, 1988; Townsend & Ashby, 1983). Because these conditions did not hold for the fits of the model presented here, this criticism is unwarranted (Townsend & Ashby, 1983). Furthermore, as shown in most of the plots of observed versus predicted mean RT, there are many predicted and observed incorrect mean RTs that are less than the predicted and observed correct mean RTs.

It is also true that as the threshold increases for a particular counter, more and more exponential deviates are added to the total response time. Therefore, the finishing time for the counter tends to be normally distributed. However, the observed RT is the minimum of two random variables that are, as both thresholds become large, normally distributed. The minimum of two normals is not normally distributed, even in the limit. The criticism that the race model must predict normally distributed RTs therefore is also unwarranted. But, the minimum distribution of two normal random variables does not necessarily predict the extent of skewness observed in the data. The skewness of the distribution depends on the degree of overlap between the two distributions. If the two normal distributions are separated by several standard deviations, the skew will be small and the minimum distribution will take the shape of the normal distribution with the smallest mean. Therefore, the race model may have difficulty in the limit with skew. It may also be appropriate to assume that thresholds do not become very large, because of capacity limitations or time pressure when speed is an issue. Under this assumption, the normality problem will not arise.

What happened to the diffusion model?

An unexpected problem arose with the fits of the diffusion model. While the diffusion model fit the RT distributions very well, probably better than the race model, it could not simultaneously account for accuracy. This is in contrast to other good fits of the diffusion model to both RTs and accuracy (Ratcliff, 1978, 1981, 1985; Ratcliff et al., 1999) . However, in other applications of the diffusion model, rarely have any goodness-of-fit criteria been applied to the RT fits. It is possible that in these other applications, if the χ^2 or Kolmogorov-Smirnoff statistics were computed, they would show significant failures of the diffusion model to fit the RTs. Fits by eye, showing that the predicted and observed curves are generally consistent with each other, do not guarantee that significant differences do not exist between the curves. Therefore, this project is one of the first times that such measurements have been taken of the diffusion model (cf. Ratcliff et al., 1999). It is also possible that the diffusion model simply has trouble with perceptual matching data. The conditions of the task may be such that the race process was the best way to characterize task performance, perhaps because of the simultaneous presentation of the elements of the pair or the fact that only two elements were compared instead of a more complicated letter string. Empirical work is required to determine if this is the case.

One handicap faced by the diffusion model is that it had fewer parameters than did the race model. The drift coefficient was held constant at $s^2 = 0.1$ because, for a single distribution, it is unobservable. The drift coefficient scales all other parameters so that for any s^2, a set of parameters can be found that leave the distribution unchanged. However, the ratio of two drift coefficients is observable, and so we might have allowed the ratio of coefficients to vary across conditions. Because of the added complexity that this would have implied for our fits, we did not attempt this.

The fact that no parameters could be found to fit the diffusion model to both RTs and accuracies should not necessarily be taken as an indication of its failure. Excellent fits may very well be possible with parameters that were not found in these analyses. However, the fact that such parameters were difficult to find may be attributed to the complexity of the diffusion model. The diffusion equations are quite complicated, and unstable for various time values. Allowing the drift rate to vary makes the problem much worse, because each prediction then requires a numerical integration of the already complicated expressions over all possible drift rates. The race model does not suffer from this problem, although rates and thresholds for the model could, if theoretical concerns warranted, be allowed to vary as well.

It should be noted that the drift variance arises from a theoretically motivated source, perceptual encoding, and therefore the diffusion model is more complete than the race model. It provides an explanation for how information is transformed into evidence toward alternative responses. The race model presented here does not specify how information arrives from earlier stages. However, a similar mechanism can be constructed for the race model, resulting in variable accumulation rates (Van Zandt, 2000b). For instance, suppose that noise in the processing system tends to make stimulus pairs register at least some difference, even when the members of the pair are identical. The amount of difference to which a pair gives rise is a random variable that depends on the pair type. If there is some total amount of capacity that can be allocated to each counter, the number of perceived mismatches would determine how much of that capacity is allocated to the "different" counter. The rest would be allocated to the "same" counter. On each trial, the amount of capacity allocated to each counter is a random variable determined by the number of perceived mismatches. Although this mechanism could have been exploited in the present project, it was not necessary to do so. The race model already fit the data reasonably well, and the additional flexibility provided by varying rates (or thresholds) would only have made the fits better.

Conclusions

The relative merits of random walk and race model representations have been debated. However, the results of Marley and Colonius (1992) have suggested that random walk models may have equivalent representations as independent race models. Race models have not been studied as extensively as the diffusion models. This study demonstrates that the race model is a viable model of response selection in a choice RT task, perceptual matching.

Race models can predict RT and accuracy (and also response confidence and double response behavior; St. James & Eriksen, 1992; Vickers, 1979). The diffusion model can also predict RT and accuracy, although it had problems doing so in the present study. Both the race and diffusion models are most useful as a framework within which subtle aspects of performance can be investigated. The calculations required to determine the predicted mean RTs, accuracy and the RT distributions from a random walk or diffusion model are considerably more complicated than those required for the race model. Eventually data may be found that the random walk or diffusion representation can describe but the race representation cannot. Until then, the greater tractability of the race model and its now-demonstrated ability to produce the observed patterns of accuracy and RT, suggests that the race model may be a more useful representation of the response selection process than the diffusion model, at least in certain situations.

References

Ashby, F. G. (1983). Testing the assumptions of exponential, additive reaction time models. *Memory and Cognition*, *10*, 125–134.

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

Bamber, D. (1969). Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception and Psychophysics*, *6*, 169–174.

Browne, M. W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *2*, 230–258.

Colonius, H. (1995). The instance theory of automaticity: Why the Weibull? *Psychological Review*, *102*, 744–750.

Cox, D. R. & Isham, V. (1980). *Point processes*. New York: Chapman and Hall.

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

Dzhafarov, E. N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology*, *36*, 235–268.

Dzhafarov, E. N. (1993). Grice-representability of response time distribution families. *Psychometrika*, *58*, 281–314.

Eriksen, C. W. & Schultz, D. W. (1977). Retinal locus and acuity in visual information processing. *Bulletin of the Psychonomic Society*, *9*, 81–84.

Farrell, B. (1985). "Same"-"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, *98*, 419– 456.

Feller, W. (1968). *An introduction to probability theory and its applications*, volume 1. John Wiley: New York.

Gallant, A. R. (1987). *Nonlinear statistical models.* New York: Wiley.

Grigelionis, B. (1963). On the convergence of sums of random step processes to a Poisson process. *Theory of Probability and Its Applications*, *8*, 177–182.

Heath, R. A. & Willcox, C. H. (1990). A stochastic model for inter-keypress times in a typing task. *Acta Psychologica*, *75*, 13–39.

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

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *6*, 598–615.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*, 183–216.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*, 188–196.

Khintchine, A. Y. (1960). *Mathematical methods in the theory of queuing.* London: Griffin.

Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review*, *85*, 278–304.

Krueger, L. E. (1985). Effect of intermixed foveal and parafoveal presentation on same-different judgments: Evidence for a criterion-inertia model. *Perception and Psychophysics*, *37*, 266–271.

Krueger, L. E. & Allen, P. A. (1987). Same-different judgments of foveal and parafoveal letter pairs by older adults. *Perception and Psychophysics*, *38*, 188–193.

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

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

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

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

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 883–914.

Logan, G. D. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review*, *102*, 751–756.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Marley, A. A. J. & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology*, *36*, 1–20.

Mordkoff, J. T. & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 520–538.

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

Newell, A. & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

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

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2 ed.). New York: Cambridge University Press.

Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychological Review*, *88*, 291–326.

Proctor, R. W. (1986). Response bias, criteria settings, and the fast-same phenomenon: A reply to Ratcliff. *Psychological Review*, *93*, 473–477.

Proctor, R. W. & Rao, V. K. (1983). Evidence that the same-different disparity in letter matching is not attributable to response bias. *Perception and Psychophysics*, *34*, 72–76.

Proctor, R. W., Van Zandt, T., & Watson, H. (1990). Effects of background symmetry on same-different pattern matching: A compromise criteria account. *Perception and Psychophysics*, *48*, 543–550.

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

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

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

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

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

Robert, S. & Pashler, H. (2000). How persuasive is a good fit? *Psychological Review*, *107*, xxx–xxx.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, *7*, 191–218.

Schwarz, W. (1990). Stochastic accumulation of information in discrete time: Comparing exact results and Wald approximations. *Journal of Mathematical Psychology*, *34*, 229–236.

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

Smith, P. L. & Van Zandt, T. (1999). Time-dependent Poisson counter models of response latency in simple judgment. Manuscript submitted for publication.

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

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

St. James, J. D. & Eriksen, C. W. (1992). Response competition produces a "fast same effect" in same-different judgments. In J. Pomerantz & G. Lockhead (Eds.), *The perception of structure* (pp. 157–168). Washington, D.C.: American Psychological Association.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Stephens, M. A. (1983). Kolmogorov-Smirnov statistics. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences*, volume 4 (pp. 393–396). New York: Wiley Press.

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

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

Van Zandt, T. (2000a). How to fit a reaction time distribution. *Psychonomic Bulletin and Review*, *7*, xxx–xxx.

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

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

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*, 37–5.

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

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

Ward, R. & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 664–672.

Appendix A

The Poisson race model

For a given stimulus, let T_S and T_D to be random variables representing the finishing times for the "same" and "different" counters, respectively. If T_S and T_D have probability density functions (pdfs) f_S and f_D, then the event that, say, a "same" response was selected at some time t has the pdf

$$f(t|\text{``same''}) = \frac{1}{P(\text{``same''})} f_S(t)[1 - F_D(t)]$$

where $F_D(t)$ represents the cumulative distribution function (cdf) of T_D and

$$P(\text{``same''}) = \int_{t=0}^{\infty} f_S(u)[1 - F_D(u)]du.$$

If it is assumed that the "same" and "different" counters receive unit amounts of information at exponentially distributed interarrival times, then the random variables T_S and T_D are gamma distributed; the shape parameter for each is the value of the response criteria K_S and K_D, and the rate parameters for each depend on the stimulus presented. Let λ_{SS} and λ_{SD} represent the accumulation rates on the "same" and "different" counters, respectively, when a *same* stimulus pair is presented. The race model pdf for a "same" response given a *same* stimulus pair is then (Townsend & Ashby, 1983, p. 274)

$$f(t|\text{``same''}) = \frac{1}{P(\text{``same''})} \frac{\lambda_{SS}(\lambda_{SS}t)^{K_S-1}}{(K_S-1)!} e^{-\lambda_{SS}t} \sum_{j=0}^{K_D-1} \frac{(\lambda_{SD}t)^j}{j!} e^{-\lambda_{SD}t},$$

which, when integrated, yields the cdf

$$\begin{aligned} F(t|\text{``same''}) &= 1 - \frac{e^{-t(\lambda_{SS}+\lambda_{SD})}}{P} \left(\frac{\lambda_{SS}}{\lambda_{SS}+\lambda_{SD}}\right)^{K_S} \\ &\quad \times \sum_{i=0}^{K_D-1} \begin{pmatrix} K_S+i-1 \\ i \end{pmatrix} \left(\frac{\lambda_{SD}}{\lambda_{SS}+\lambda_{SD}}\right)^{i} \sum_{j=0}^{K_S+i-1} \frac{[t(\lambda_{SS}+\lambda_{SD})]^j}{j!}. \end{aligned}$$

The probability that the "same" response was given is

$$P(\text{``same''}) = \sum_{i=0}^{K_D-1} \begin{pmatrix} K_S+i-1 \\ i \end{pmatrix} \left(\frac{\lambda_{SS}}{\lambda_{SS}+\lambda_{SD}}\right)^{K_S} \left(\frac{\lambda_{SD}}{\lambda_{SS}+\lambda_{SD}}\right)^{i}.$$

Simple closed-form expressions exist for the race model pdf, probabilities, cdf, and all means and higher moments. To obtain expressions for the densities, distributions and probabilities for a "different" response, the S and D subscripts in the above expressions should be reversed. Notice that in this study we have considered only integer values for thresholds K_S and K_D. This was done strictly for convenience, and there is no reason why future investigations of this model should be so limited. Dzhafarov (1993) has noted that the "units" represented by the thresholds are dimensionless entities, and so the issue of real versus integer-valued thresholds is an empty one.

The diffusion model

For a particular stimulus type, a drift rate ξ is selected at random from a normal distribution with mean $\mu = -\xi_D$ or ξ_S and variance η^2. The activation level begins at z and drifts toward either boundary 0 (for "different") or a (for "same"). The drift variance parameter s^2, fit in other applications using the diffusion model (Ratcliff, 1978), is not generally observable and was set equal to 0.1. The finishing time pdf for the time that the activation reaches 0 can be expressed in two ways. The most well-known expression is

$$f(t|\text{“different”}, \xi) = \frac{1}{P}\frac{\pi s^2}{a^2} e^{-\frac{z\xi}{s^2}} \sum_{k=1}^{\infty} k \sin\left(\frac{\pi z k}{a}\right) e^{-\frac{1}{2}t(\xi^2/s^2 + \pi^2 k^2 s^2/a^2)},$$

where P is the probability of absorption at 0, or a "different" response:

$$P = \frac{e^{-2\frac{\xi z}{s^2}} - e^{-2\frac{\xi a}{s^2}}}{e^{-2\frac{\xi z}{s^2}} - 1}.$$

The cdf is

$$\begin{aligned} F(t|\text{“different”}, \xi) &= 1 - \frac{1}{P}\frac{\pi s^2}{a^2} e^{-z\xi/s^2} \\ &\times \sum_{k=1}^{\infty} \frac{2k \sin\left(\frac{k\pi z}{a}\right) e^{-\frac{1}{2}t(\xi^2/s^2+\pi^2 k^2 s^2/a^2)}}{\xi^2/s^2 + \pi^2 k^2 s^2/a^2}. \end{aligned}$$

(The expressions for the probability of and finishing times for absorption at boundary a are found by replacing ξ with $-\xi$ and z with $a - z$ in the equations above.) The marginal pdf is found by integrating over all values of ξ:

$$f(t|\text{“different”}) = \int_{-\infty}^{\infty} f(t|\text{“different”}, \xi) \phi\left(\frac{\xi - \xi_D}{\eta}\right) d\xi,$$

where $\phi(\xi)$ is the standard normal pdf. The marginal finishing time pdf has no simple closed-form expression and must be found by numerical means. Unfortunately, this form of the pdf is unstable for small t. As t goes to zero, the infinite sum of sine functions begins to fluctuate from negative infinity to positive infinity. This behavior can be troublesome for numerical integration.

An alternative form of the joint pdf is given by Feller (1968):

$$f(t|\text{“different”}, \xi) = \frac{1}{P}\frac{e^{-\frac{(t\xi+2z)\xi}{2s^2}}}{\sqrt{2\pi s^2 t^3}}\sum_{k=-\infty}^{\infty}(z+2ka)e^{-\frac{(z+2ka)^2}{2ts^2}}.$$

As t goes to zero, this expression goes to zero. Unfortunately, as t gets very large, this expression becomes unstable whereas the expression above converges quite nicely. However, this expression can be evaluated very quickly, because the infinite sum converges with only a very few terms. We used this expression in the numerical routines, but switched to the alternative when it became unstable. It can be integrated to give

$$\begin{aligned} F(t|\text{“different”}, \xi) &= \frac{-e^{-z\xi/s^2}}{2P}\sum_{k=-\infty}^{\infty}\left(\operatorname{sgn}(k)\ e^{|\xi|(z+2ka)/s^2}\{\operatorname{erf}\left(\frac{z+2ka+|\xi|t}{\sqrt{2ts^2}}\right)-1\}+\right. \\ &\qquad \left.\operatorname{sgn}(k)\ e^{-|\xi|(z+2ka)/s^2}\{\operatorname{erf}\left(\frac{z+2ka-|\xi|t}{\sqrt{2ts^2}}\right)-1\}\right), \end{aligned}$$

where $\operatorname{erf}(x)$ is the error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du,$$

and the sign function

$$\operatorname{sgn}(k) = \begin{cases} -1 & \text{if } k < 0 \\ 1 & \text{if } k \geq 0 \end{cases}.$$

Appendix B

Table B1. Unconstrained and inappropriate (bias-constrained) parameter values and χ^2 statistics of the race model for each participant (P) in each condition of Experiment 1. Deadlines (Dead) are given in ms. Significance levels for unconstrained and inappropriate fits are calculated with 13 and 15 degrees of freedom, respectively.

Model	P	Dead	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
Unconstrained										
		500	10	9	66.38	33.28	58.15	37.98	.262	13.18
	1	750	8	6	44.07	12.15	32.36	17.79	.305	19.65
		1000	10	8	43.07	12.43	35.85	18.16	.289	31.30[2]
		500	15	19	82.90	66.95	101.57	50.87	.183	28.38[2]
Blocked	2	750	8	9	52.47	22.37	60.06	18.63	.274	18.33
		1000	7	7	49.75	15.08	47.90	16.15	.285	12.65
		500	11	5	75.05	19.98	35.59	49.11	.254	13.77
	3	750	5	3	40.65	8.93	25.44	18.64	.309	19.09
		1000	5	3	35.72	8.04	22.14	15.28	.309	5.58
		500	9	6	65.09	22.39	42.33	33.85	.286	20.23
	1	750	8	5	60.65	17.19	35.67	30.20	.292	34.27[2]
		1000	11	7	71.57	24.80	44.01	41.76	.271	13.68
		500	25	22	127.47	74.65	111.75	81.27	.177	12.28
Mixed	2	750	20	19	105.26	61.17	98.81	61.96	.183	26.66[1]
		1000	20	19	105.85	61.91	101.08	66.40	.187	20.78
		500	7	3	49.78	11.90	23.45	28.86	.278	25.95[1]
	3	750	8	4	53.05	15.98	29.61	33.93	.265	25.66[1]
		1000	9	4	57.61	15.69	29.15	38.58	.261	23.79[1]

Table B1 continued

Model	P	Dead	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
Inappropriate										
		500			81.91	40.94	68.38	49.66		14.59
	1	750	14	12	56.94	26.19	49.06	24.97	.241	20.44
		1000			50.29	6.26	44.40	23.44		28.64^1
		500			99.33	0.00	107.35	0.67		39.06^3
Blocked	2	750	16	18	69.07	43.59	78.59	35.90	.196	27.44^1
		1000			69.22	44.00	77.28	13.25		27.59^1
		500			65.62	21.18	40.49	39.45		23.55
	3	750	9	5	54.09	14.63	30.63	31.49	.265	29.56^1
		1000			49.12	12.89	27.95	26.59		18.24
		500			62.94	22.27	40.42	32.06		18.24
	1	750	9	6	62.99	20.33	39.69	30.46	.281	30.77^2
		1000			62.84	21.76	40.39	34.05		13.94
		500			88.31	50.11	82.30	50.80		15.35
Mixed	2	750	14	13	89.69	40.10	81.88	44.75	.216	21.15
		1000			87.01	44.03	81.57	50.45		19.55
		500			56.03	15.14	27.47	35.55		24.66
	3	750	9	4	57.23	15.33	27.89	37.56	.259	23.65
		1000			56.51	15.34	28.80	37.35		21.92

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table B2. Unconstrained and inappropriate (bias-constrained) parameter values and χ^2 statistics of the diffusion model for each participant (P) in each condition of Experiment 1. Significance levels for unconstrained and inappropriate fits are calculated with 14 and 16 degrees of freedom, respectively.

Model	p	Dead	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
Unconstrained									
		500	1.749	.883	3.601	3.502	.118	.161	12.95
	1	750	1.957	1.059	3.101	2.617	.010	.140	16.81
		1000	1.605	.816	2.405	2.401	.026	.180	25.83[1]
		500	1.730	.862	3.707	3.634	.010	.126	21.46
Blocked	2	750	1.669	.838	3.581	3.588	.429	.188	12.48
		1000	1.612	.806	3.387	3.315	.348	.184	8.41
		500	1.509	.873	3.748	2.919	.134	.160	11.87
	3	750	1.301	.747	3.925	3.121	.666	.231	19.44
		1000	1.158	.648	2.963	2.450	.388	.219	3.82
		500	1.543	.846	3.758	3.070	.229	.191	11.98
	1	750	1.504	.792	3.357	2.948	.010	.181	26.49[1]
		1000	1.690	.889	3.566	3.218	.118	.169	12.68
		500	1.545	.812	3.887	3.503	.010	.161	15.68
Mixed	2	750	1.551	.804	3.787	3.496	.232	.158	16.79
		1000	1.566	.819	3.714	3.429	.011	.153	21.23
		500	1.582	.917	3.696	2.930	.346	.158	18.45
	3	750	1.202	.697	3.274	2.624	.256	.189	21.92
		1000	1.511	.849	3.336	2.869	.205	.149	17.33

Table B2 continued

Model	P	Dead	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
Inappropriate									
		500			4.005	3.685			13.01
	1	750	2.017	1.047	3.130	2.893	.177	.141	22.37
		1000			2.829	2.698			27.23[1]
		500			3.302	3.231			83.35[3]
Blocked	2	750	1.022	.504	2.281	2.367	.001	.205	86.08[3]
		1000			2.284	2.298			85.77[3]
		500			3.706	2.958			19.17
	3	750	1.325	.758	3.232	2.490	.178	.185	26.17
		1000			3.028	2.317			13.37
		500			3.005	2.534			7.99
	1	750	1.051	.567	2.984	2.487	.010	.229	21.33[1]
		1000			2.997	2.550			15.25
		500			4.005	3.859			17.89
Mixed	2	750	1.916	.976	3.990	3.819	.001	.123	21.39
		1000			3.940	3.838			19.45
		500			3.012	2.295			23.14
	3	750	.995	.588	3.047	2.338	.193	.209	21.13
		1000			3.018	2.384			20.53

[1]$p < .05$, [2]$p < .01$, [3]$p < .001$

Table B3. Unconstrained and inappropriate (bias-constrained) parameter values and χ^2 statistics of the race model for each participant (P) in each condition of Experiment 2. Significance levels for unconstrained and inappropriate fits are calculated with 13 and 15 degrees of freedom, respectively.

Model	P	Bias	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
Unconstrained										
		20	9	8	60.09	0.00	37.40	20.65	.307	25.47^1
	1	50	5	2	32.08	2.36	11.04	9.01	.357	35.62^3
		80	20	2	85.06	2.96	12.66	48.31	.288	15.28
		20	3	2	29.51	1.68	12.50	9.31	.338	21.41
	2	50	10	2	57.87	2.33	14.83	6.47	.312	19.51
		80	14	2	63.24	4.15	20.26	11.88	.294	41.08^3
		20	2	3	12.66	0.00	11.24	1.64	.427	60.25^3
	3	50	3	2	11.00	0.61	7.83	0.08	.402	138.83^3
		80	3	2	6.69	1.10	7.71	0.04	.328	61.24^3
Inappropriate										
		20			70.86	2.89	10.12	32.47		86.01^3
	1	50	13	3	54.54	3.39	10.83	24.40	.274	59.19^3
		80			51.42	5.51	18.76	7.71		39.23^3
		20			56.40	0.01	8.67	25.13		71.86^3
	2	50	8	2	42.15	2.34	11.99	18.46	.296	69.76^3
		80			35.30	4.19	20.67	3.87		60.76^3
		20			16.63	0.10	3.86	0.00		172.45^3
	3	50	3	1	11.06	0.16	3.15	3.32	.405	192.28^3
		80			7.74	0.31	5.32	0.01		82.26^3

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table B4. Unconstrained and inappropriate (bias-constrained) parameter values and χ^2 statistics of the diffusion model for each participant (P) in each condition of Experiment 2. Significance levels for unconstrained and inappropriate fits are calculated with 14 and 16 degrees of freedom, respectively.

Model	S	Bias	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
Unconstrained									
		20	1.613	.896	3.892	2.551	.476	.225	20.03
	1	50	1.499	.928	3.222	1.936	.447	.215	23.26
		80	1.293	.989	3.289	1.403	.010	.219	8.17
		20	1.475	.820	3.955	2.836	.754	.224	12.73
	2	50	1.420	.964	3.687	2.091	.471	.218	15.33
		80	.855	.681	2.926	1.496	.342	.275	23.30
		20	1.713	.792	3.246	2.791	1.058	.316	20.32
	3	50	1.765	.981	3.112	2.679	1.215	.319	33.11^2
		80	1.799	1.094	2.004	1.859	.633	.173	24.67^1
Inappropriate									
		20			3.879	1.713			98.42^3
	1	50	1.938	1.273	3.348	1.692	.210	.124	46.44^3
		80			3.266	2.170			57.01^3
		20			3.231	1.571			59.03^3
	2	50	1.097	.709	2.673	1.692	.267	.213	54.33^3
		80			2.402	2.190			88.84^3
		20			2.204	1.154			116.27^3
	3	50	1.216	.740	1.785	1.192	.391	.235	121.74^3
		80			1.471	1.461			70.73^3

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table B5. Unconstrained and inappropriate (rate-constrained) parameter values and χ^2 statistics of the race model for each participant (P) in each condition of Experiment 3. Significance levels for unconstrained and inappropriate fits are calculated with 13 and 16 degrees of freedom, respectively.

Model	P	Width	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
Unconstrained										
		1.5	10	4	44.27	5.39	21.94	19.23	.313	37.86[3]
	1	5.3	7	3	34.62	3.84	13.07	14.76	.344	15.21
		9.5	4	2	22.02	2.46	9.66	9.59	.402	20.89
		1.5	9	3	47.62	3.22	22.14	26.04	.383	15.53
	2	5.3	7	2	43.95	1.71	10.67	16.06	.423	14.39
		9.5	4	2	23.90	2.38	10.75	9.32	.446	17.10
		1.5	8	3	44.11	2.74	18.74	22.61	.393	21.44
	3	5.3	5	2	26.03	1.74	9.96	13.21	.422	51.38[3]
		9.5	4	2	21.20	2.01	9.02	9.69	.434	21.83
		1.5	4	4	23.44	6.07	18.90	6.58	.361	17.04
	4	5.3	3	8	20.62	3.38	35.55	5.17	.384	32.09[2]
		9.5	2	6	12.88	19.49	26.57	5.49	.413	25.39[1]
		1.5	7	2	32.56	2.05	12.51	11.33	.416	26.21[1]
	5	5.3	9	4	33.00	6.80	15.35	19.14	.373	53.49[3]
		9.5	7	3	28.58	5.57	13.36	16.83	.424	24.45
		1.5	7	3	37.83	4.54	17.94	15.15	.347	17.53
	6	5.3	5	4	24.80	9.63	19.14	11.92	.352	24.02[1]
		9.5	8	9	29.30	24.92	30.64	21.75	.291	15.08

Table B5 continued

Model	P	Width	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
Rate constrained										
		1.5	5	1						55.61^3
	1	5.3	4	2	0.10	28.83	1.34	21.97	.372	30.03^1
		9.5	5	2						23.47
		1.5	6	1						29.45^1
	2	5.3	6	2	35.86	2.72	4.23	26.80	.417	27.20^1
		9.5	7	2						38.11^2
		1.5	4	2						50.15^3
	3	5.3	5	3	25.21	2.47	16.84	4.23	.418	69.53^3
		9.5	5	3						31.69^1
		1.5	5	5						45.94^3
	4	5.3	5	7	27.55	5.83	18.57	18.88	.350	57.76^3
		9.5	5	7						42.59^3
		1.5	9	3						34.80^2
	5	5.3	9	4	36.46	3.81	15.96	19.69	.386	60.45^3
		9.5	10	4						29.37^1
		1.5	6	3						27.49^1
	6	5.3	6	4	19.14	20.39	12.17	23.96	.337	30.83^1
		9.5	5	5						18.66

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table B6. Unconstrained and inappropriate (bias-constrained) parameter values and χ^2 statistics of the diffusion model for each participant (P) in each condition of Experiment 3. Significance levels for unconstrained and inappropriate fits are calculated with 14 and 16 degrees of freedom, respectively.

Model	P	Width	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
Unconstrained									
		1.5	1.869	1.100	3.076	2.482	.331	.175	29.18[2]
	1	5.3	1.658	.992	2.983	1.988	.349	.206	11.44
		9.5	1.732	1.018	3.031	2.170	.481	.234	21.64
		1.5	1.961	1.135	3.051	2.649	.080	.196	8.64
	2	5.3	1.476	.953	3.477	1.873	.435	.302	11.85
		9.5	2.068	1.090	2.727	2.461	.309	.195	23.79[1]
		1.5	1.719	.928	3.513	3.478	.712	.301	30.51[2]
	3	5.3	2.436	1.256	2.808	2.840	.375	.155	69.02[3]
		9.5	1.844	1.025	2.808	2.319	.441	.244	18.58
		1.5	1.904	.968	2.858	2.497	.387	.179	16.42
	4	5.3	1.281	.472	2.112	2.721	.383	.299	30.83[2]
		9.5	1.962	.858	2.923	2.980	.444	.238	21.83[1]
		1.5	1.750	1.069	2.924	2.209	.392	.251	21.78
	5	5.3	1.595	.957	2.506	1.779	.229	.247	42.42[3]
		9.5	1.932	1.074	2.587	2.254	.267	.232	18.10
		1.5	2.017	1.110	3.235	2.849	.430	.178	15.05
	6	5.3	1.862	.986	2.866	2.511	.357	.188	18.58
		9.5	2.111	1.057	2.858	2.718	.243	.160	13.89

Table B6 continued

Model	S	Width	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
Rate Constrained									
		1.5	1.905	1.122					37.67^2
	1	5.3	2.019	1.138	2.830	2.213	.162	.133	21.19
		9.5	2.161	1.225					33.66^2
		1.5	1.592	1.005					27.45^1
	2	5.3	1.761	1.040	3.330	2.379	.388	.262	29.79^1
		9.5	1.881	1.110					29.70^1
		1.5	1.180	.689					47.04^3
	3	5.3	1.324	.765	2.280	1.828	.001	.273	45.75^3
		9.5	1.365	.764					19.11
		1.5	1.547	.752					33.63^2
	4	5.3	1.660	.750	2.318	2.299	.001	.200	38.40^2
		9.5	1.699	.761					39.36^3
		1.5	1.231	.768					23.04
	5	5.3	1.289	.765	2.257	1.678	.001	.290	36.84^2
		9.5	1.357	.804					24.53
		1.5	1.637	.895					31.03^1
	6	5.3	1.726	.913	2.595	2.309	.070	.177	27.60^1
		9.5	1.747	.900					18.29

$^1p < .05$, $^2p < .01$, $^3p < .001$

Author Notes

Portions of this work were presented at the 23^{rd} annual meeting of the Society for Mathematical Psychology, University of Toronto, 1990. The project was made possible with grants from NIMH #MH-44640 and NSF #SBR-9702291. The authors wish to thank In Jae Myung for advice on model comparison statistics, and the reviewers, F. Gregory Ashby, Gordon Logan, and Philip Smith, for many helpful comments that greatly improved this paper, as well as Lester Krueger for comments on an earlier draft.

Footnotes

[1]The degrees of freedom for the χ^2 statistics were calculated in the following way. For the overall fits shown in Table 2, the degrees of freedom were the total number of bins used in the χ^2 calculation minus the total number of parameters. Each distribution was described using 11 quantiles, which yielded 10 bins. Six distributions entered into the overall fits, for a total of 60 bins. So, for example, there were 60-11=49 degrees of freedom for the appropriate fits of the race model to the distributions from the Blocked condition. For fits to each condition (in which the two "same" and "different" distributions were fit), the total degrees of freedom was divided by three and conservatively rounded downward.

[2]The algorithm was written by D. L. Carroll, University of Illinois, and is available through the World Wide Web at *http://www.staff.uiuc.edu/~carroll/ga.html.*

[3]The standard error of the RMS statistic is given by $\sqrt{1 - \frac{2}{v}\left(\frac{\Gamma(v/2+1/2)}{\Gamma(v/2)}\right)^2}$, where v is the number of degrees of freedom (number of data points fit minus the number of parameters).

Table 1. Parameter values and χ^2 statistics of the appropriate race model for each participant (P) in each condition of Experiment 1. Deadlines (Dead) are given in ms. Significance levels for rate-constrained are calculated using 16 degrees of freedom for the blocked condition and 17 degrees of freedom for the mixed condition.

	P	Dead	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
		500	5	5						21.58
	1	750	8	8	43.77	17.91	42.67	18.49	.299	41.78[3]
		1000	10	9						60.54[3]
		500	4	5						139.36[3]
Blocked	2	750	7	8	45.63	20.35	52.25	16.05	.275	20.44
		1000	7	8						22.16
		500	6	3						30.73[1]
	3	750	7	4	48.11	11.53	25.79	25.81	.282	33.39[2]
		1000	8	4						27.12[1]
		500	8	5						19.76
	1	750	8	5	61.18	18.05	36.46	31.05	.293	39.67[2]
		1000	8	5						18.06
		500	15	14						17.18
Mixed	2	750	15	14	92.93	50.00	86.38	52.21	.213	30.47[1]
		1000	15	14						25.44
		500	7	3						28.44[1]
	3	750	7	3	50.16	11.93	23.90	30.20	.277	25.20
		1000	7	3						30.35

[1]$p < .05$, [2]$p < .01$, [3]$p < .001$

Table 2. Overall χ^2 goodness-of-fit statistics for each model fit to each participant's data in Experiments 1, 2, and 3, and also for each model fit to each simulation. Simulations 1 and 2 are rate- and bias-constrained race models, respectively, and simulations 3 and 4 are rate- and bias-constrained diffusion models, respectively.

		Race			Diffusion		
Experiment	Participant	Unc	App	Inapp	Unc	App	Inapp
1	1(B)	64.13^2	123.90^3	63.67^1	55.59	216.80^3	62.61
	1(M)	68.18^2	77.49^1	62.95^1	51.15	42.59	44.57
	2(B)	59.36^1	181.96^3	94.09^3	42.35	65.27	255.20^3
	2(M)	59.72^1	73.09^1	56.05	53.70	67.26	58.73
	3(B)	38.44	91.24^3	71.35^2	35.13	78.57^2	58.71
	3(M)	75.40^3	83.99^2	70.23^2	57.70	67.56	64.80
2	1	76.37^3	158.07^3	184.43^3	51.46	124.04^3	201.87^3
	2	82.00^3	141.82^3	202.38^3	51.36	72.51^1	202.00^3
	3	260.32^3	534.77^3	446.99^3	78.10^3	282.68^3	308.74^3
3	1	73.96^3	88.61^3	109.11^3	62.26^1	90.16^3	92.52^3
	2	47.02	78.97^2	94.76^3	44.28	85.62^2	86.94^3
	3	94.65^3	106.26^3	152.37^3	118.11^3	203.06^3	111.90^3
	4	74.52^3	161.26^3	146.29^3	69.08^2	188.38^3	111.39^3
	5	104.15^3	117.27^3	124.62^3	82.30^3	121.79^3	83.41^2
	6	56.63^1	82.41^3	76.98^2	47.52^1	93.48^3	76.92^2
Simulations	1	33.99	39.72	44.96	45.99	66.79	64.76
	2	36.94	41.55	286.21^3	37.44	134.97^3	53.19
	3	46.49	229.17^3	180.15^3	31.89	36.68	187.04^3
	4	38.94	56.77	387.16^3	43.14	50.03	58.85

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 3. Kolmogorov-Smirnoff statistics and significance levels for the unconstrained (Unc), appropriate (App) , and inappropriate (Inapp) race and diffusion model fits for each participant (P) in Experiment 1. Significance levels are calculated from the sample sizes (N) and the critical values $D_{.05} = 1.36/\sqrt{N}$ and $D_{.01} = 1.63/\sqrt{N}$.

					Race			Diffusion		
Condition	P	Dead	Pair	N	Unc	App	Inapp	Unc	App	Inapp
		500	Diff	723	.0205	.0299	.0241	.0194	.0937[2]	.0230
			Same	709	.0197	.0184	.0214	.0177	.0526[1]	.0232
	1	750	Diff	780	.0235	.0373	.0206	.0237	.0463	.0249
			Same	773	.0262	.0409	.0315	.0252	.0552[1]	.0317
		1000	Diff	798	.0231	.0533[1]	.0332	.0333	.0303	.0274
			Same	790	.0247	.0382	.0326	.0256	.0526[1]	.0272
		500	Diff	734	.0247	.0617[2]	.0313	.0332	.0215	.0590[1]
			Same	730	.0284	.0612[2]	.0326	.0311	.0420	.0586[1]
Blocked	2	750	Diff	796	.0190	.0315	.0295	.0276	.0158	.0635[2]
			Same	794	.0236	.0413	.0327	.0245	.0282	.0684[2]
		1000	Diff	803	.0213	.0259	.0371	.0187	.0230	.0678[2]
			Same	795	.0230	.0284	.0275	.0182	.0368	.0539[1]
		500	Diff	655	.0240	.0603[1]	.0251	.0241	.0525	.0208
			Same	669	.0202	.0444	.0336	.0197	.0323	.0298
	3	750	Diff	715	.0273	.0517[1]	.0372	.0231	.0391	.0355
			Same	713	.0317	.0448	.0308	.0270	.0244	.0355
		1000	Diff	711	.0236	.0331	.0372	.0179	.0388	.0334
			Same	726	.0259	.0459	.0247	.0191	.0331	.0210

[1]$p < .05$, [2]$p < .01$, [3]$p < .001$

Table 3 continued

					Race			Diffusion		
Condition	P	Dead	Pair	N	Unc	App	Inapp	Unc	App	Inapp
		500	Diff	711	.0257	.0289	.0266	.0231	.0281	.0321
			Same	720	.0259	.0298	.0367	.0288	.0344	.0259
	1	750	Diff	727	.0332	.0339	.0294	.0400	.0327	.0348
			Same	709	.0294	.0371	.0368	.0397	.0293	.0269
		1000	Diff	712	.0243	.0316	.0241	.0329	.0265	.0344
			Same	696	.0227	.0211	.0187	.0244	.0259	.0176
		500	Diff	741	.0234	.0406	.0271	.0259	.0425	.0283
			Same	761	.0236	.0356	.0261	.0336	.0406	.0281
Mixed	2	750	Diff	757	.0304	.0333	.0324	.0322	.0431	.0303
			Same	767	.0255	.0314	.0301	.0290	.0318	.0303
		1000	Diff	755	.0250	.0285	.0295	.0272	.0260	.0272
			Same	754	.0235	.0241	.0252	.0361	.0357	.0274
		500	Diff	613	.0276	.0261	.0412	.0348	.0221	.0236
			Same	685	.0387	.0361	.0249	.0246	.0437	.0310
	3	750	Diff	618	.0302	.0322	.0323	.0310	.0362	.0242
			Same	671	.0211	.0250	.0260	.0241	.0257	.0289
		1000	Diff	611	.0277	.0246	.0268	.0506	.0212	.0196
			Same	675	.0315	.0480	.0312	.0266	.0411	.0336

$^1p < .05$, $^2p < .01$

Table 4. Parameter values and χ^2 statistics for the fit of the appropriate diffusion model to each participant's (P) data from Experiment 1. Significance levels are calculated with 16 degrees of freedom for the blocked condition and 17 degrees of freedom for the mixed condition.

	P	Dead	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
		500	1.366	.651					97.79^3
	1	750	1.759	.833	2.674	2.941	.145	.168	66.13^3
		1000	1.926	.929					52.88^3
		500	1.479	.722					36.58^2
Blocked	2	750	1.900	.943	3.398	3.439	.193	.145	16.33
		1000	1.923	.950					12.36
		500	1.278	.730					30.93^1
	3	750	1.470	.827	3.281	2.665	.188	.173	22.39
		1000	1.558	.877					25.25
		500	1.535	.842					13.41
	1	750	1.535	.842	3.711	3.038	.240	.191	19.68
		1000	1.535	.842					9.50
		500	1.522	.799					22.83
Mixed	2	750	1.522	.799	3.873	3.514	.200	.165	19.13
		1000	1.522	.799					25.30
		500	1.498	.880					24.89
	3	750	1.498	.880	3.750	2.928	.377	.170	22.71
		1000	1.498	.880					19.96

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 5. Root mean squared statistics for each model fit to each participant's data in Experiments 1, 2, and 3, and also for each model fit to each simulation. An asterisk indicates a number too large to be computed.

		Race			Diffusion		
Experiment	Participant	Unc	App	Inapp	Unc	App	Inapp
1	1(Blocked)	1.22	**1.68**	28.19	38.66	36.15	36.14
	1(Mixed)	1.23	**1.18**	1.19	223.82	199.31	203.06
	2(Blocked)	1.21	**2.02**	*	29.45	27.12	26.79
	2(Mixed)	1.15	**1.15**	1.35	107.84	97.80	99.73
	3(Blocked)	0.92	1.46	**1.37**	164.83	151.78	151.74
	3(Mixed)	1.39	1.28	**1.27**	258.24	232.14	239.00
2	1	*	**2.77**	58.50	56.48	53.32	53.11
	2	59.05	**3.74**	91.40	25.22	7.71	42.73
	3	*	6.93	*	3.96	**3.24**	11.52
3	1	1.59	**2.41**	*	195.47	125.23	202.76
	2	1.05	**1.57**	7.08	187.47	172.61	173.58
	3	1.47	**1.47**	13.04	262.96	32.82	238.64
	4	92.52	464.20	**19.71**	447.92	412.90	407.63
	5	1.59	38.16	**4.04**	196.28	181.80	178.86
	6	1.16	104.27	**11.03**	426.18	396.11	391.45
Simulations	1	0.95	**0.89**	279.61	51.36	47.25	47.24
	2	13.39	**1.02**	78.99	325.83	301.71	295.65
	3	1.02	3.35	2.01	0.83	**0.87**	3.69
	4	0.93	1.06	*	0.95	**0.95**	1.38

Table 6. Parameter values and χ^2 statistics of the appropriate race model for each participant (P) in each bias condition of Experiment 2. Significance levels are calculated with 16 degrees of freedom for each level of bias.

P	Bias	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
	20	10	4						68.57^3
1	50	13	3	59.02	2.99	13.06	23.85	.289	70.57^3
	80	14	2						18.93
	20	8	4						54.92^3
2	50	11	3	55.58	4.67	19.54	21.05	.292	39.98^3
	80	13	2						46.91^3
	20	4	3						176.27^3
3	50	5	3	15.43	0.50	8.21	4.95	.336	192.85^3
	80	7	2						165.65^3

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 7. Kolmogorov-Smirnoff statistics and significance levels for the unconstrained, aprropriate rate-constrained (App), and inappropriate bias-constrained (Inapp) race and diffusion model fits for each participant (P) in Experiment 2. Significance levels are calculated from the sample sizes (N) and the critical values $D_{.05} = 1.36/\sqrt{N}$ and $D_{.01} = 1.63/\sqrt{N}$.

					Race			Diffusion	
P	Bias	Pair	N	Unc	App	Inapp	Unc	App	Inapp
	20	Diff	1126	.0268	.0406[1]	.0284	.0404	.0296	.0281
		Same	266	.0234	.1656[2]	.1924[2]	.0685	.1486[2]	.1687[2]
1	50	Diff	683	.0322	.0470	.0578 [1]	.0301	.0420	.0575[1]
		Same	669	.0371	.0810[2]	.0648[2]	.0444	.0153	.0461
	80	Diff	244	.0373	.0588	.0814	.0366	.0589	.0760
		Same	1116	.0298	.0233	.0423[1]	.0157	.0325	.0609[2]
	20	Diff	1128	.0407[1]	.0398	.0576[2]	.0275	.0295	.0441[1]
		Same	231	.0434	.1123[2]	.2795[2]	.0385	.0805	.1315[2]
2	50	Diff	667	.0361	.0581[1]	.0807[2]	.0466	.0394	.0616[1]
		Same	670	.0280	.0390	.1274[2]	.0194	.0361	.0349
	80	Diff	225	.0312	.0760	.1199[2]	.0431	.0891	.0995[1]
		Same	1115	.0548[2]	.0569[2]	.1110[2]	.0375	.0284	.0748[2]
	20	Diff	1121	.0493[2]	.1107[2]	.0576[2]	.0287	.0916[2]	.0513[2]
		Same	260	.0925[1]	.1491[2]	.2795[2]	.0466	.0868[1]	.2066[2]
3	50	Diff	689	.0804[2]	.0753[2]	.0807[2]	.0508	.0876[2]	.0633[2]
		Same	676	.0417	.1007[2]	.1274[2]	.0693[2]	.0588[1]	.1061[2]
	80	Diff	254	.1290[2]	.2110[2]	.1199[2]	.0730	.0799	.1431[2]
		Same	1108	.0441[1]	.0341	.1110[2]	.0450[1]	.0833[2]	.0431[1]

[1] $p < .05$, [2] $p < .01$

Table 8. Parameter values and χ^2 statistics of the appropriate diffusion model for each participant (P) in each bias condition of Experiment 2. Significance levels are calculated with 16 degrees of freedom for each level of bias.

P	Bias	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
	20	1.354	.800					67.82^3
1	50	1.491	.967	3.288	1.728	.313	.212	35.47^2
	80	1.381	1.014					20.75
	20	1.148	.675					24.66
2	50	1.244	.853	3.677	2.159	.626	.248	24.00
	80	1.268	.976					23.85
	20	1.859	.836					70.69^3
3	50	1.890	.985	2.793	2.621	.968	.272	71.60^3
	80	1.818	1.144					140.39^3

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 9. Parameter values and χ^2 statistics of the appropriate race model for each participant (P) for each visual angle of Experiment 3. Significance levels are calculated with 15 degrees of freedom for each visual angle.

P	Width	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
	1.5			40.72	6.33	23.19	13.80		38.84^3
1	5.3	9	4	39.76	5.74	16.04	18.38	.320	17.00
	9.5			34.37	5.57	13.87	19.92		32.78^2
	1.5			50.30	5.23	28.77	23.01		16.72
2	5.3	10	4	47.82	6.04	17.99	20.72	.374	24.10
	9.5			42.22	4.84	15.64	22.39		38.14^3
	1.5			32.20	1.66	15.03	14.04		28.53^1
3	5.3	5	2	25.74	1.76	9.78	13.19	.420	51.68^3
	9.5			24.83	1.83	8.07	12.11		26.05^1
	1.5			32.95	4.47	36.59	1.40		44.97^3
4	5.3	7	9	32.51	3.02	29.39	13.82	.313	66.99^3
	9.5			28.73	22.34	26.25	16.93		49.30^3
	1.5			38.91	5.77	20.66	5.14		30.00^1
5	5.3	10	4	36.92	6.63	15.44	21.52	.374	55.35^3
	9.5			33.92	7.06	14.17	22.47		31.91^2
	1.5			38.91	11.86	29.89	4.00		28.10^1
6	5.3	9	6	36.29	13.12	23.05	20.78	.302	30.94^2
	9.5			35.71	16.57	19.96	25.16		23.37

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 10. Kolmogorov-Smirnoff statistics and significance levels for the unconstrained, appropriate bias-constrained (App), and inappropriate rate-constrained (Inapp) race and diffusion model fits for each participant (P) in Experiment 3. Significance levels are calculated from the sample sizes (N) and the critical values $D_{.05} = 1.36/\sqrt{N}$ and $D_{.01} = 1.63/\sqrt{N}$.

					Race			Diffusion	
P	Width	Pair	N	Unc	App	Inapp	Unc	App	Inapp
	1.5	Diff	506	0.0515	0.0512	0.0824[2]	.0464	.0315	.0609[1]
		Same	507	0.0408	0.0368	0.0611[1]	.0394	.0392	.0550
1	5.3	Diff	504	0.0384	0.0382	0.0624[1]	.0348	.0468	.0469
		Same	463	0.0331	0.0413	0.0478	.0292	.0335	.0391
	9.5	Diff	491	0.0340	0.0463	0.0372	.0415	.0362	.0489
		Same	452	0.0515	0.0353	0.0349	.0578	.0711[1]	.0378
	1.5	Diff	528	0.0236	0.0251	0.0313	.0471	.0421	.0341
		Same	515	0.0326	0.0441	0.0621[1]	.0312	.0428	.0504
2	5.3	Diff	518	0.0454	0.0533	0.0506	.0382	.0588	.0646[1]
		Same	492	0.0405	0.0305	0.0365	.0721	.0341	.0409
	9.5	Diff	498	0.0424	0.0661[1]	0.0602	.0874	.0646[1]	.0477
		Same	462	0.0287	0.0466	0.0489	.0289	.0312	.0253
	1.5	Diff	531	0.0336	0.0372	0.0861[2]	.0567	.0559	.0671[1]
		Same	496	0.0381	0.0610	0.0635[1]	.0789[2]	.0901[2]	.0596
3	5.3	Diff	507	0.0734[2]	0.0714[1]	0.0853[2]	.0722[1]	.0558	.0689[1]
		Same	433	0.0362	0.0366	0.0868[2]	.0381	.0610	.0386
	9.5	Diff	499	0.0470	0.0552	0.0410	.0495	.0541	.0343
		Same	427	0.0258	0.0329	0.0696[1]	.0235	.0532	.0336

Table 10 continued

				Race			Diffusion		
S	Width	Pair	N	Unc	App	Inapp	Unc	App	Inapp
	1.5	Diff	526	0.0288	0.0345	0.0537	.0282	.0538	.0421
		Same	495	0.0248	0.0604	0.0702^1	.0264	.0529	.0551
4	5.3	Diff	503	0.0560	0.0752^2	0.0797^2	.0469	$.0742^2$	.0368
		Same	474	0.0359	0.0848^2	0.1112^2	.0378	$.0980^2$	$.0745^1$
	9.5	Diff	445	0.0425	0.0519	0.0586	.0414	.0408	.0392
		Same	357	0.0527	0.0709	0.0578	.0455	$.1027^2$	.0606
	1.5	Diff	495	0.0373	0.0347	0.0456	.0517	.0358	.0430
		Same	516	0.0452	0.0552	0.0488	.0484	$.0700^1$	.0468
5	5.3	Diff	466	0.0589	0.0621	0.0456	.0482	.0411	.0406
		Same	473	0.0303	0.0283	0.0451	.0383	.0354	.0356
	9.5	Diff	444	0.0367	0.0408	0.0516	.0599	.0371	.0385
		Same	459	0.0320	0.0516	0.0466	.0263	.0510	.0317
	1.5	Diff	505	0.0410	0.0502	0.0482	.0582	.0406	$.0646^1$
		Same	501	0.0279	0.0408	0.0332	.0363	.0572	.0381
6	5.3	Diff	451	0.0390	0.0572	0.0773^2	.0295	.0495	.0471
		Same	464	0.0303	0.0346	0.0270	.0365	.0332	.0249
	9.5	Diff	389	0.0289	0.0339	0.0566	.0301	.0413	.0330
		Same	362	0.0299	0.0648	0.0420	.0352	$.0751^1$	.0427

$^1p < .05$, $^2p < .01$

Table 11. Parameter values and χ^2 statistics of the appropriate diffusion model for each participant (P) for each visual angle of Experiment 3. Significance levels are calculated with 16 degrees of freedom for each visual angle.

P	Width	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
	1.5			3.394	2.628			39.60^3
1	5.3	1.716	1.040	3.326	2.170	.475	.224	16.27
	9.5			3.026	2.008			34.29^2
	1.5			3.108	2.747			21.67
2	5.3	1.956	1.127	3.027	2.242	.172	.206	32.77^2
	9.5			2.817	2.120			31.18^1
	1.5			3.577	3.619			96.40^3
3	5.3	1.605	.866	3.143	3.143	.829	.323	67.09^3
	9.5			3.095	2.813			39.57^3
	1.5			2.665	2.507			39.90^3
4	5.3	1.785	.879	2.644	2.226	.245	.194	89.92^3
	9.5			2.575	2.173			58.50^3
	1.5			2.787	2.215			38.52^2
5	5.3	1.696	1.018	2.746	1.982	.284	.257	53.71^3
	9.5			2.576	1.888			29.56^1
	1.5			3.132	2.816			35.59^2
6	5.3	1.947	1.062	3.031	2.528	.346	.184	23.35
	9.5			3.070	2.441			34.54^2

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 12. Parameter values used to simulate the rate-constrained and bias-constrained race and diffusion models.

	Condition	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}
	1	5	5					
1. Race/Rate	2	7	7	45.63	16.05	52.25	20.35	.275
	3	9	9					
	1			33.00	9.00	37.00	12.00	
2. Race/Bias	2	8	8	31.00	15.00	30.00	16.00	.275
	3			29.00	22.00	26.00	17.00	

		a	z	ξ_D	ξ_S	η		T_{er}
	1	.300	.150					
3. Diffusion/Rate	2	.500	.250	.500	.520	.200		.150
	3	.700	.350					
	1			.500	.520			
4. Diffusion/Bias	2	.700	.350	.350	.370	.200		.150
	3			.200	.220			

Table 13. Parameter values and χ^2 statistics of the unconstrained, rate- and bias-constrained race model fits to each condition for each simulation. Significance levels for unconstrained fits are calculated with 13 degrees of freedom for each condition. For rate-constrained fits there are 16 degrees of freedom for each condition and for bias-constrained fits there are 15 degrees of freedom for each condition.

Model	Condition	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
Unconstrained									
	1	6	6	51.49	21.58	58.37	22.49	.268	17.52
1. Race/Rate	2	7	7	45.33	16.47	51.62	20.72	.275	12.75
	3	11	11	50.74	18.19	55.89	24.91	.252	4.72
	1	10	11	36.16	16.53	44.26	6.11	.241	11.35
2. Race/Bias	2	8	8	30.65	14.64	29.69	15.86	.275	18.51
	3	5	6	19.85	17.52	21.83	10.71	.309	7.08
	1	1	1	5.30	1.25	5.73	1.44	.186	14.22
3. Diffusion/Rate	2	1	1	2.60	0.31	2.78	0.27	.246	12.90
	3	1	1	1.92	0.10	1.91	0.13	.320	19.37
	1	1	1	1.80	0.13	1.83	0.11	.310	14.26
4. Diffusion/Bias	2	1	1	1.44	0.23	1.45	0.21	.343	13.14
	3	1	1	0.86	0.35	1.01	0.32	.322	11.54

Table 13 continued

Model	Condition	K_D	K_S	λ_{DD}	λ_{DS}	λ_{SS}	λ_{SD}	T_{er}	χ^2
Rate Constrained									
	1	5	5						19.13
1. Race/Rate	2	7	7	46.26	15.87	52.36	20.30	.277	13.22
	3	9	9						7.37
	1	6	6						112.62^3
2. Race/Bias	2	7	7	32.67	4.86	30.33	15.03	.306	36.68^2
	3	8	8						136.91^3
	1	1	1						13.77
3. Diffusion/Rate	2	2	2	4.91	0.60	4.95	0.86	.181	54.10^3
	3	3	3						161.30^3
	1	2	2						156.48^3
4. Diffusion/Bias	2	2	2	2.60	0.00	2.66	0.00	.185	72.82^3
	3	2	2						157.86^3
Bias Constrained									
	1			76.95	34.55	85.61	9.18		21.38
1. Race/Rate	2	11	11	58.09	24.95	64.42	31.63	.239	14.16
	3			47.38	21.62	51.99	26.12		9.42
	1			30.06	10.13	33.81	10.97		12.85
2. Race/Bias	2	7	7	27.73	12.92	26.85	14.09	.286	19.54
	3			26.27	19.35	23.42	14.90		9.16
	1			10.08	3.32	10.60	3.82		77.67^3
3. Diffusion/Rate	2	2	2	4.47	0.98	4.71	0.88	.160	58.45^3
	3			3.04	0.44	3.03	0.57		44.03^3
	1			1.84	0.13	1.87	0.11		18.65
4. Diffusion/Bias	2	1	1	1.39	0.23	1.40	0.20	.319	26.31^1
	3			0.87	0.35	1.01	0.32		11.81

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 14. Parameter values and χ^2 statistics of the unconstrained, rate- and bias-constrained diffusion model for each simulation in each condition. Significance levels for unconstrained fits are calculated with 14 degrees of freedom for each condition and for rate- and bias-constrained fits there are 16 degrees of freedom for each condition.

Model	Condition	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
Unconstrained								
	1	.968	.494	2.957	3.072	.068	.213	17.51
1. Race/Rate	2	1.378	.689	2.782	3.045	.018	.179	10.77
	3	1.664	.910	3.427	3.072	.454	.200	17.71
	1	2.203	1.143	2.579	2.560	.040	.072	12.25
2. Race/Bias	2	2.338	1.158	2.510	2.521	.032	.064	18.33
	3	2.143	1.032	2.688	2.639	.324	.132	6.86
	1	.298	.145	.480	.632	.308	.151	8.27
3. Diffusion/Rate	2	.486	.239	.442	.512	.183	.158	12.96
	3	.616	.299	.398	.438	.046	.175	10.66
	1	.623	.309	.395	.412	.074	.163	10.14
4. Diffusion/Bias	2	.643	.319	.334	.344	.173	.185	17.74
	3	.635	.318	.132	.181	.010	.161	15.26

Table 14 continued

Model	Condition	a	z	ξ_D	ξ_S	η	T_{er}	χ^2
Rate Constrained								
	1	.863	.432					27.58^1
1. Race/Rate	2	1.089	.547	2.657	2.901	.154	.220	17.23
	3	1.305	.653					21.98
	1	1.362	.711					19.03
2. Race/Bias	2	1.471	.735	2.175	2.136	.010	.188	20.52
	3	1.501	.719					13.64
	1	.285	.141					11.36
3. Diffuse/Rate	2	.489	.240	.444	.508	.105	.154	13.38
	3	.673	.324					11.94
	1	.526	.262					10.03
4. Diffuse/Bias	2	.577	.285	.037	.116	.188	.191	19.31
	3	.624	.314					29.51^1
Bias Constrained								
	1			3.750	3.704			32.39^2
1. Race/Rate	2	1.501	.779	3.071	3.092	.109	.172	19.04
	3			2.641	2.638			13.33
	1			3.255	2.849			48.98^3
2. Race/Bias	2	1.995	1.100	3.153	2.510	.427	.171	49.38^3
	3			3.237	2.356			36.61^2
	1			1.374	1.493			62.71^3
3. Diffusion/Rate	2	.596	.292	.574	.657	.329	.108	28.66^1
	3			.007	.022			95.67^3
	1			.435	.446			13.20
4. Diffusion/Bias	2	.644	.320	.322	.325	.117	.164	20.33
	3			.137	.198			16.50

$^1p < .05$, $^2p < .01$, $^3p < .001$

Table 15. Kolmogorov-Smirnoff statistics and significance levels for the unconstrained, rate-, and bias-constrained race and diffusion model fits for each simulation. Significance levels are calculated from the sample sizes (N) and the critical values $D_{.05} = 1.36/\sqrt{N}$ and $D_{.01} = 1.63/\sqrt{N}$.

				Race			Diffusion		
	Condition	Pair	N	Unc	Rate	Bias	Unc	Rate	Bias
1. Race/Rate	1	Diff	750	.0253	.0334	.0257	.0219	.0500[1]	.0726[2]
		Same	744	.0369	.0312	.0422	.0325	.0334	.0829[2]
	2	Diff	778	.0142	.0202	.0187	.0626[2]	.0233	.0588[2]
		Same	759	.0184	.0217	.0214	.0445	.0256	.0469
	3	Diff	789	.0192	.0256	.0304	.0564[1]	.0326	.0546[1]
		Same	774	.0111	.0201	.0259	.0644[2]	.0521[1]	.0622[2]
2. Race/Bias	1	Diff	793	.0320	.1673[2]	.0354	.0425	.0369	.0680[2]
		Same	784	.0216	.0339	.0167	.0388	.0263	.0414
	2	Diff	744	.0242	.0545[1]	.0299	.0455	.0268	.0636[2]
		Same	709	.0174	.0407	.0219	.0583[1]	.0172	.0623[2]
	3	Diff	559	.0329	.1737[2]	.0297	.0637[1]	.0227	.0535
		Same	635	.0188	.0511	.0156	.0407	.0305	.0674[2]
3. Diffusion/Rate	1	Diff	647	.0325	.0722[2]	.0742[2]	.0214	.0269	.0594[1]
		Same	639	.0423	.0887[2]	.0739[2]	.0308	.0372	.0637[2]
	2	Diff	715	.0517[1]	.0571[1]	.0473	.0207	.0209	.0372
		Same	728	.0409	.0633[2]	.0433	.0235	.0246	.0507[1]
	3	Diff	759	.0433	.0921[2]	.0507[1]	.0165	.0226	.0854[2]
		Same	750	.0392	.0710[2]	.0662[2]	.0228	.0230	.0873[2]

Table 15 continued

	Condition	Pair	N	Race Unc	Race Rate	Race Bias	Diffusion Unc	Diffusion Rate	Diffusion Bias
	1	Diff	748	.0415	.1381[2]	.0495	.0170	.0226	.0205
		Same	754	.0292	.1393[2]	.0385	.0220	.0171	.0244
4. Diffusion/Bias	2	Diff	688	.0407	.0715[2]	.0319	.0250	.0191	.0210
		Same	700	.0445	.0634[2]	.0430	.0188	.0237	.0226
	3	Diff	570	.0264	.1001[2]	.0330	.0286	.0399	.0334
		Same	606	.0281	.0736[2]	.0264	.0324	.0337	.0307

[1] $p < .05$, [2] $p < .01$

Figure Captions

1. The diffusion (top panel) and race (bottom panel) models. For the diffusion model, drift rates are randomly sampled from one of two possible distributions depending on the stimulus presented. In the race model, events are recorded on two independent counters with rates determined by the stimulus presented. Both models accumulate information until a threshold is exceeded (in the race model) or a boundary is crossed (in the diffusion model).

2. Predicted densites of the unconstrained, appropriate and inappropriate race models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for Participant 1 for each deadline in the Blocked condition of Experiment 1. The deadline (500, 750 or 1000 ms) is noted on each panel. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

3. Predicted densites of the unconstrained, appropriate and inappropriate race models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for Participant 1 for each deadline in the Mixed condition of Experiment 1. The deadline (500, 750 or 1000 ms) is noted on each panel. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

4. Predicted accuracy (top panel), correct mean RTs (middle panel) and incorrect mean RTs (lower panel) of the unconstrained, appropriate, and inappropriate race models (dots, open squares and diamonds, and open triangles, respectively) versus the observed accuracy, correct mean RTs and incorrect mean RTs for each participant, deadline and condition of Experiment 1.

5. Predicted densities of the unconstrained, appropriate and inappropriate diffusion models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for Participant 1 for each deadline in the Blocked condition of Experiment 1. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

6. Predicted densities of the unconstrained, appropriate and inappropriate diffusion models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for Participant 1 for each deadline in the Mixed condition of Experiment 1. Correct “different” RTs are presented on the left, and correct “same” RTs are presented on the right.

7. Predicted accuracy (top panel), correct mean RTs (middle panel) and incorrect mean RTs (lower panel) of the unconstrained, appropriate and inappropriate diffusion models (dots, open squares and diamonds, and open triangles, respectively) versus the observed accuracy, correct mean RTs and incorrect mean RTs for each participant, deadline and condition of Experiment 1.

8. Predicted densities of the unconstrained, appropriate and inappropriate race models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for Participant 1 for each bias condition in Experiment 2. Correct “different” RTs are presented on the left, and correct “same” RTs are presented on the right.

9. Predicted accuracy (top panel), correct mean RTs (middle panel) and incorrect mean RTs (lower panel) of the unconstrained, appropriate and inappropriate race models (dots, open squares and diamonds, and open triangles, respectively) versus the observed accuracy, correct mean RTs and incorrect mean RTs for each participant and bias condition of Experiment 2.

10. Predicted densities of the unconstrained, appropriate and inappropriate diffusion models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for Participant 1 for each bias condition in Experiment 2. Correct “different” RTs are presented on the left, and correct “same” RTs are presented on the right.

11. Predicted accuracy (top panel), correct mean RTs (middle panel) and incorrect mean RTs (lower panel) of the unconstrained, appropriate and inappropriate diffusion models (dots, open squares and diamonds, and open triangles, respectively) versus the observed accuracy, correct mean RTs and incorrect mean RTs for each participant, deadline and condition of Experiment 2.

12. Predicted densities of the unconstrained, appropriate and inappropriate race models

(dotted, solid, and dashed line, respectively) and the observed quantiles (open symbols) for Participant 6 for each stimulus condition in Experiment 3. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

13. Predicted accuracy (top panel), correct mean RTs (middle panel) and incorrect mean RTs (lower panel) of the unconstrained, appropriate and inappropriate race models (dots, open squares and diamonds, and open triangles, respectively) versus the observed accuracy, correct mean RTs and incorrect mean RTs for each participant and stimulus condition of Experiment 3.

14. Predicted densities of the unconstrained, appropriate and inappropriate diffusion models (dotted, solid, and dashed line, respectively) and the observed quantiles (open symbols) for Participant 6 for each stimulus condition in Experiment 3. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

15. Predicted accuracy (top panel), correct mean RTs (middle panel) and incorrect mean RTs (lower panel) of the unconstrained, appropriate and inappropriate diffusion models (dots, open squares and diamonds, and open triangles, respectively) versus the observed accuracy, correct mean RTs and incorrect mean RTs for each participant and stimulus condition of Experiment 3.

16. Predicted densites of the unconstrained, rate-, and bias-constrained race (panel a) and diffusion (panel b) models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for the simulation of the rate-constrained race model. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

17. Predicted densities of the unconstrained, rate-, and bias-constrained race (panel a) and diffusion (panel b) models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for the simulation of the bias-constrained race model. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

18. Predicted densities of the unconstrained, rate-, and bias-constrained race (panel a) and

diffusion (panel b) models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for the simulation of the rate-constrained diffusion model. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

19. Predicted densities of the unconstrained, rate-, and bias-constrained race (panel a) and diffusion (panel b) models (dotted, solid, and dashed line, respectively) and the observed densities (open symbols) for the simulation of the bias-constrained diffusion model. Correct "different" RTs are presented on the left, and correct "same" RTs are presented on the right.

20. Predicted accuracy (top panels), correct mean RTs (middle panels) and incorrect mean RTs (lower panels) of the unconstrained, rate-, and bias-constrained race and diffusion models versus the observed accuracy, correct mean RTs and incorrect mean RTs for each simulation. The race model data are shown in the right panels, and the diffusion model data are shown in the right panels. Fits of the race model to the race data, or the diffusion model to the diffusion data, are plotted as open squares, diamonds, and solid dots, and fits of the diffusion model to the race data, or the race model to the diffusion data, are plotted as open circles, triangles and dots. Fits of a rate- or bias-constrained model to rate- or bias-constrained data are appropriate (open diamonds and triangles), but fits of a rate- or bias-constrained model to bias- or rate-constrained data are inappropriate (open squares and circles).