Psychonomic Bulletin & Review
2002, 9 (2), 278-291

A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data

ROGER RATCLIFF
Northwestern University, Evanston, Illinois

A brightness discrimination experiment was performed to examine how subjects decide whether a patch of pixels is "bright" or "dark," and stimulus duration, brightness, and speed versus accuracy instructions were manipulated. The diffusion model (Ratcliff, 1978) was fit to the data, and it accounted for all the dependent variables: mean correct and error response times, the shapes of response time distributions for correct and error responses, and accuracy values. Speed–accuracy manipulations affected only boundary separation (response criteria settings) in the model. Drift rate (the rate of accumulation of evidence) in the diffusion model, which represents stimulus quality, increased as a function of stimulus duration and stimulus brightness but asymptoted as stimulus duration increased from 100 to 150 msec. To address the argument that the diffusion model can fit any pattern of data, simulated patterns of plausible data are presented that the model cannot fit.

The ability to perceive faint, briefly presented visual stimuli is usually studied in detection experiments in which the dependent measure is detection threshold. In this article, I examine detection in a two-choice task, which is rarely used, and show how the diffusion model (Ratcliff, 1978) can account not only for the accuracy of performance in visual signal detection, but also for multiple aspects of response times (RTs). The model accounts for the shapes of the RT distributions for correct and error responses and for the relationships between speed and accuracy, across levels of accuracy varying from near-chance performance to near-perfect performance.

There have been only a few previous attempts to relate RT and accuracy in signal detection tasks in which accuracy varies across a wide range, and none was completely successful in providing an account of processing. For example, in an experiment by Carterette, Friedman, and Cosmides (1965), subjects judged whether a weak auditory signal in noise was present or absent. Carterette et al. examined accuracy, correct and error RTs, and their distributions and tested two sequential sampling models that were designed to account for both RTs and accuracy (LaBerge, 1962; Stone, 1960), but the models mispredicted error RTs. Sekuler (1965) used a paradigm in which two brief visual stimuli were presented sequentially and a subject's task was to detect a stripe in one of them. As in the Carterette et al. study, differences between correct and error RTs were used to evaluate and reject the then-current sequential sampling models. Kellogg (1931) performed experiments in which the stimuli were disks for which the two halves differed in brightness and the difference was varied from small to large. Subjects decided which half was brighter, and RTs and accuracy were measured. Link (1992) reviewed Kellogg's experiments and provided a fit of his *wave theory* random walk model to RT and accuracy, but not to the distributions of the RTs. Espinoza-Varas and Watson (1994) used asterisks, two-digit numbers, and tones as stimuli in a signal detection task. As with the other studies, the results were used to address qualitative predictions of various sequential sampling models and also to extend the database of results for manipulations of stimulus probability and instructions on criterion settings.

The problem in these studies was that it was not possible to simultaneously explain the joint behaviors of accuracy and RT distributions for both correct and error responses. The relations between the measures are complex. When accuracy is varied from floor (chance performance) to ceiling (near-perfect performance), RTs for both correct and error responses vary in systematic ways. When there is a large change in one of the dependent measures, change in the other variable may be small. Also, error RTs behave differently from correct RTs, often increasing as correct RTs decrease. These complexities mean that models developed to deal only with accuracy usually cannot make correct predictions about the behavior of RT (if they can make any predictions at all). Nevertheless, a full account of performance requires an account of all aspects of both variables. The difficulty of providing such an account and the

The research in this article was supported by NIMH Grants K05 MH01891 and R37 MH44640. I thank Gail McKoon for a critical reading of the article. Correspondence concerning this article should be addressed to R. Ratcliff, Department of Psychology, Northwestern University, Evanston, IL 60208 (e-mail: r-ratcliff@nwu.edu).

lack of success of the models that have been examined have led to reliance by the field only on measures of accuracy or threshold level.

Recently, the diffusion model (Ratcliff & Rouder, 1998, 2000; Ratcliff, Van Zandt, & McKoon, 1999) has enjoyed considerable success in explaining accuracy and RT measures in a number of two-choice decision tasks similar to those used in signal detection and visual perception experiments. The goal of the research described in this article was to apply the model to signal detection of briefly presented visual stimuli. A brightness discrimination task was used. Each stimulus was a 64 × 64 patch of black and white pixels on a gray background. Brightness was manipulated by varying the proportion of black to white pixels in the patch, and stimulus duration was varied from 50 to 150 msec, terminated by presentation of a mask. A subject's task was to decide whether the patch was *dark* or *bright*. Brightness and stimulus duration were varied over a range sufficient to produce accuracy rates from near 50% correct to near 100% correct.

This paradigm is similar to those that have provided the data that form the basis of Bloch's (1885) law. The law states that when a pulse of light is presented for some duration that is lower than a critical duration, detection will be at threshold (e.g., the light will be detected on 50% of trials) when intensity (I) multiplied by duration (T) is a constant—that is, $IT = k$ (Smith, 1998; Watson, 1986). The value of k varies as a function of background illumination (as well as stimulus type and other variables). The law breaks down at about 50-msec stimulus duration, and depending on stimulus properties, additional stimulus duration may or may not improve performance (e.g., Smith, 1998; Watson, 1986).

The brightness discrimination paradigm we used asks the same questions as a paradigm used by Rayner, Inhoff, Morrison, Slowiaczek, and Bertera (1981). They used a reading task with eye movement monitoring equipment. During eye fixations, they masked the fixated word and words to the right of the fixated word after several stimulus durations. They found that after about 80–100 msec of processing time (stimulus duration), reading rate was close to normal, which means that most of the information derived in an eye fixation during reading is obtained with a stimulus duration of 80–100 msec.

To anticipate the results of the brightness discrimination experiment presented in this article, the diffusion model gives a good account of accuracy, RTs, RT distributions for both correct and error responses, and the relationships among these dependent measures. The model also provides an estimate of the quality of the stimulus information that drives the decision process and of how quality varies as a function of stimulus duration.

The diffusion model is a member of the class of sequential sampling models for which it is assumed that evidence is accumulated gradually over time toward one of two alternative decision boundaries (Ratcliff, 1978, 1981, 1985, 1988; Ratcliff & Rouder, 1998, 2000; Ratcliff, Thapar, & McKoon, 2001). The model is summarized in Figure 1. The diffusion process represents the decision component of processing; all other components of processing, such as stimulus encoding and response execution, are summed into one parameter, T_{er}, which is assumed to be distributed

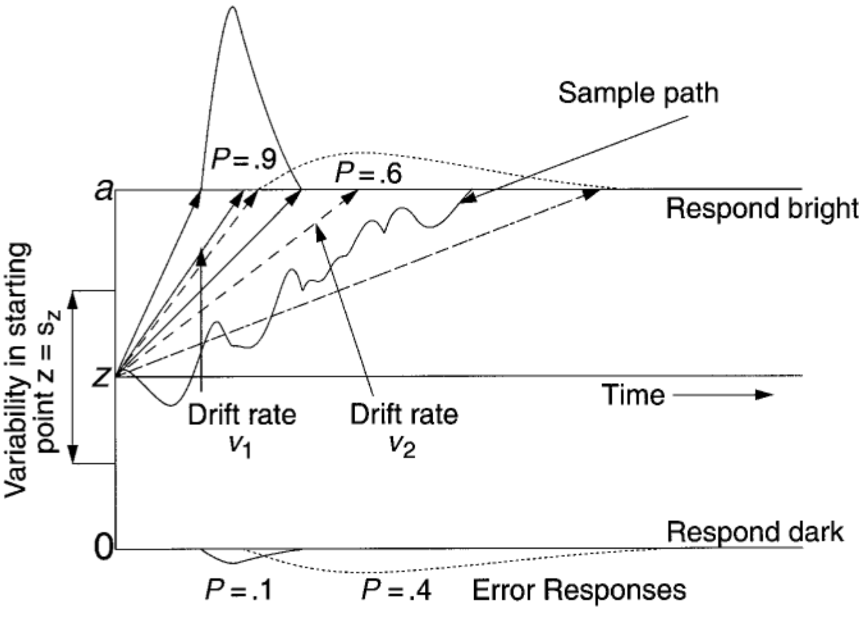

Figure 1. A summary of the diffusion model. Two drift rates, v_1 (solid lines) and v_2 (dashed lines), are shown that represent high and low rates of accumulations of evidence with correct response probabilities .9 and .6 (with associated error probabilities of .1 and .4). Schematic fast and slow processes from the two drift rates are shown, which show how equal steps in drift rate map into skewed response time distributions.

uniformly (for simplicity) with range s_t. The diffusion process itself starts at a point z and terminates when the process reaches a boundary at 0 or at a. The process is driven by the drift rate v, which is a function of stimulus quality. If v is zero, the average drift of the process is toward neither boundary. A positive value of drift drives the process toward the positive boundary, and a negative value of drift drives it toward the negative boundary. In the brightness discrimination task, a stimulus perceived as very bright would have a large value of drift toward the *bright* boundary, and a stimulus perceived as very dark would have a large value of drift toward the *dark* boundary. Variability across time in the diffusion process (standard deviation, s) means that the process can sometimes terminate at the wrong boundary by mistake, producing an error. Even with zero drift, variability ensures that the process will eventually hit one or the other of the boundaries, so the process terminates with a probability of 1.

Drift rate (v) and starting point (z) are assumed to vary across different trials with the same stimulus (Laming, 1968; Ratcliff, 1978, 1981; Ratcliff et al., 1999). This allows the model to account for the patterns of error and correct RTs that have been observed across a number of experiments: fast errors when accuracy is high, slow errors when accuracy is low, and a crossover pattern so that errors are slow when accuracy is low but fast when accuracy is high (Ratcliff, Gomez, & McKoon, 2002; Ratcliff & Rouder, 2000; Ratcliff et al., 1999; Smith & Vickers, 1988). Without the assumptions of across-trial variability, the model could not produce the correct patterns of error and correct RTs. It is reasonable to assume variability in these parameters because subjects cannot encode information exactly the same way across (ostensibly) equivalent trials, nor can they accurately reset the zero point of the information accumulation process from trial to trial. Specifically, drift rate is assumed to be normally distributed with standard deviation η, and starting point is assumed to be uniformly distributed with range s_z. The uniform assumption for starting point replaces the normal assumption used by Ratcliff and Rouder (1998, 2000) and Ratcliff et al. (1999) because the normal assumption allows the starting point to be placed outside the decision boundaries with some small probability (see Ratcliff & Tuerlinckx, in press).

The diffusion model is a good candidate for application to the data from the brightness paradigm because Smith (1998) recently showed that a different member of the diffusion model family (the Ornstein Uhlenbeck, or OU, process) could explain Bloch's law and, more impressively, could explain the different ways the law broke down. Although the OU model did not deal with RTs or RT distributions, it has all the machinery needed to produce RT predictions in simple detection paradigms (Smith, 1995).

The diffusion model itself has given good accounts of data in visual signal detection of categorical stimuli—for example, letters (Ratcliff & Rouder, 2000). On each trial of Ratcliff and Rouder's (2000) experiments, one of two letters was displayed for one of five stimulus durations, then masked. Accuracy varied from near chance to near ceiling. Two variations of the diffusion model were tested. For one, it was assumed that the information from the stimulus that drives drift rate follows the availability of perceptual information: When the stimulus is on, drift rate has some nonzero value, and when the stimulus turns off, drift rate returns to zero. For the second model, it was assumed that the perceptual information is integrated (summed) over the time for which the stimulus is presented and that this constant value drives the drift rate. Only this latter version of the diffusion model, the version with constant drift rate, was consistent with the experimental data. This version of the model gave an accurate account of accuracy rates, RTs, and RT distributions across all values of accuracy. The question for the experiment described below was whether the diffusion model could explain visual signal detection with the noncategorical, brightness patch stimuli.

The goal for modeling is to predict the distributions of correct and error RTs and accuracy for each experimental condition. To show the distributions and accuracy values for all the conditions at once, we introduce quantile-probability functions. These plot the quantile points of the RT distributions (e.g., the .1, .3, .5, .7, and .9 quantile points) against the probability of a response. Thus, for each value of response probability, the five quantile RTs are plotted in a vertical line. Horizontal or inverted-U-shaped lines can be drawn to connect each of the quantiles (e.g., the .1 quantiles, the .3 quantiles, and so on) across conditions or to connect the theoretical predictions for the quantiles. Usually, about five quantile points are sufficient to capture the shapes of the distributions. Because a quantile-probability plot shows the RT distribution for each level of accuracy, it provides in one display the multiple aspects of the data that need to be explained.

EXPERIMENT

The goal of this experiment was to generate data that provide a strong test of the diffusion model by varying stimulus duration and stimulus brightness to produce a range of values of accuracy. The experiment was also designed to examine processing of brief visual stimuli as a function of duration and brightness. For application of the diffusion model, it is assumed that the effects of brightness and stimulus duration are qualitatively the same: Both limit the amount of available stimulus information. Stimulus information translates into a single quantity, drift rate, to drive the decision process in the diffusion model.

In addition to the stimulus brightness and duration manipulations, speed–accuracy instructions were varied across alternating blocks of trials. For a speed block, the subjects were instructed to respond as quickly as they possibly could, and for an accuracy block, they were instructed to be as accurate as possible. This manipulation places additional strong constraints on the diffusion model. If processing of stimulus information is the same in the speed and the accuracy conditions in every way except a change in decision criteria, the model should be able to capture the data with a change only in the distance between the two decision boundaries and no change in any other parameter.

Method

Subjects. One Northwestern undergraduate participated in nine sessions, 1 in eight sessions, and 1 in seven sessions (plus each participated in one practice session). They were paid $8 for each 45-min session.

Stimuli. The stimuli were 64 × 64 squares of black and white pixels on a gray background (the whole display was 320 × 200 pixels). *Brightness* of the square was manipulated by varying the probability that a pixel was white. Four checkerboard patterns, each 64 × 64 pixels, were used to mask each stimulus; presented sequentially, they were a checkerboard with 2 × 2 black and white squares, a checkerboard the same as the first but with the black and white squares reversed, a checkerboard with 3 × 3 black and white squares, and its reverse. The checkerboards were designed by trial and error to mask both smaller and larger random features of a stimulus that might have remained visible through only one or two of the masks. The smaller checkerboard seemed to eliminate the smaller random patterns in a stimulus, and the larger checkerboard seemed to eliminate the larger random patterns.

Apparatus. The stimuli were presented on a Pentium II class machine, and the responses were collected on the keyboard.

Procedure. There were six levels of brightness, achieved with six values for the probability of a pixel's being white: .350, .425, .475, .525, .575, and .650. These were crossed with three stimulus durations: 50, 100, and 150 msec.

Each trial began with a + sign fixation point presented on a gray background for 250 msec. Then the stimulus was displayed, followed by the four checkerboard masks displayed for 17 msec each. Then a gray background was presented until a response was made. In accuracy blocks, if a response was correct, there was a 500-msec pause and then the next trial; if a response was incorrect, the word ERROR was displayed for 300 msec, then erased, and then there was a 500-msec pause before the next trial. In speed blocks, there was no accuracy feedback. If a response was shorter than 250 msec, a message TOO FAST was displayed; if a response was longer than 700 msec, TOO SLOW was displayed. Then, there was a 500-msec pause before the next trial.

In each session, there were five blocks of accuracy trials alternating with five blocks of speed trials, with 144 trials per block presented in random order. There was a total of 40 trials for each brightness, duration, and speed versus accuracy condition in each session.

In accuracy blocks, the subjects were instructed to respond accurately. In the speed blocks, the subjects were instructed to respond quickly, using the TOO SLOW message as a guide to when they were responding too slowly.

Results

Trials with responses shorter than 250 msec and longer than 2,500 msec were eliminated from analyses (less than 0.9% of the data), and the data in each condition are based on about 880–920 observations. There was a strong bias toward a *bright* response, especially at the shortest stimulus duration for 2 of the 3 subjects. Figure 2 shows accuracy as a function of brightness and duration for the speed and accuracy conditions averaged over the 3 subjects. If there was no bias, the three functions would pass through the cross hairs in the middle of the figure. The bias probably occurred because the neutral gray background was perceived as dark, as compared with the white in the stimulus. However, the 3rd subject seemed less affected and showed a small *bright* bias at short stimulus durations and a small *dark* bias at longer durations. Because the subjects differed in their bias and 1 showed little bias, this suggests that the bias is acting just as in a standard signal detection analysis.

The manipulation of the brightness and duration variables led to values for the probability of a *bright* response ranging from .05 to .95. The RTs were shorter and their range was smaller in the speed condition relative to the accuracy condition. In the speed condition, RTs varied from 390 to 470 msec, and in the accuracy condition, responses varied from 450 to 610 msec. The differences in response probability between the speed and the accuracy conditions showed a steeper psychometric function for the accuracy condition than for the speed condition, and differences between the two were as large as 7% when accuracy was in the middle of its range between ceiling and floor—that is, between .7 and .8.

To display the joint RT and accuracy data, we collapsed the 18 conditions into 5 conditions by grouping conditions with the same values of accuracy (within .04). This was pos-

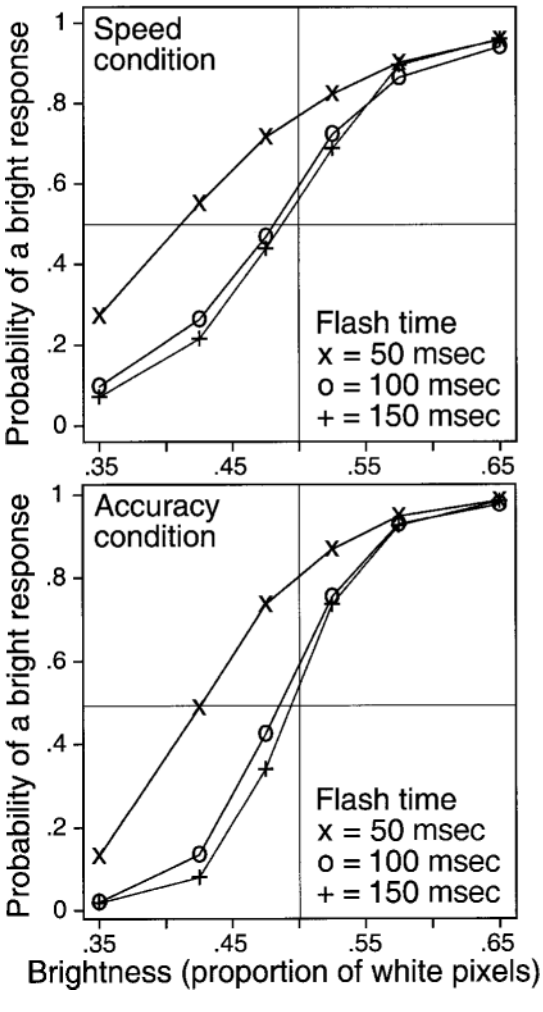

Figure 2. Probability of a *bright* response as a function of brightness of the stimulus for three flash times and the speed and accuracy conditions. The data are averaged over subjects.

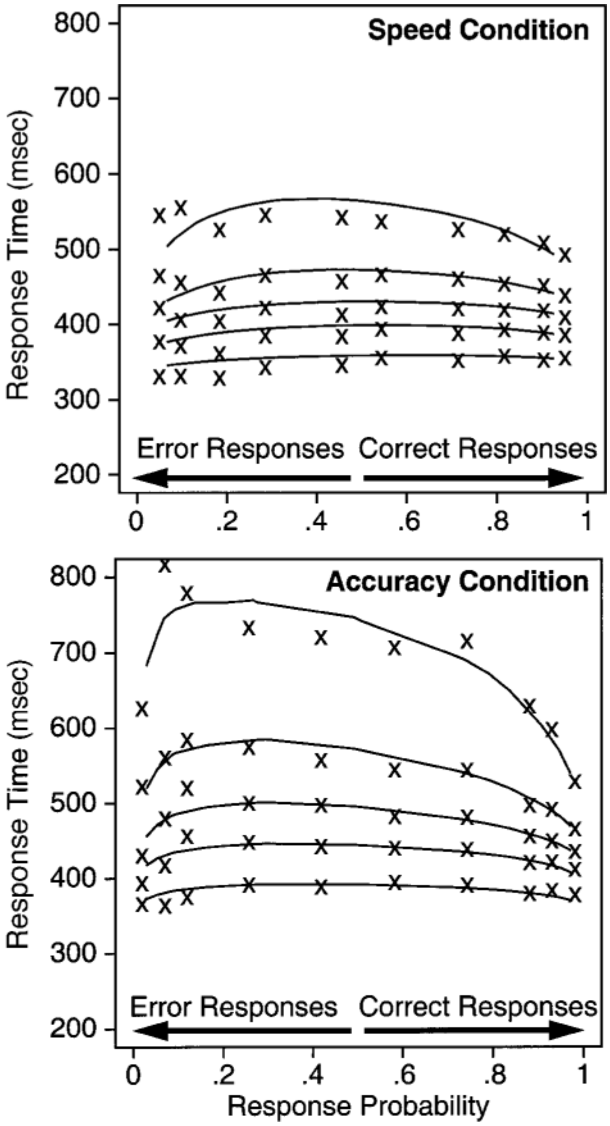

Figure 3. Quantile-probability functions for the speed condition (top panel) and the accuracy condition (bottom panel) for average data in the experiment. The continuous lines show the fit of the diffusion model and the *x*s are the experimental data. For the 18 experimental conditions, conditions with similar accuracy and response times were grouped to produce the 5 conditions presented in the panels. Also, *bright* responses to bright stimuli were grouped with *dark* responses to dark stimuli, and *bright* responses to dark stimuli were grouped with *dark* responses to bright stimuli.

sible because there were no systematic differences in RTs for conditions with the same accuracy values. The data are shown in quantile-probability functions in Figure 3, separately for the speed and the accuracy conditions. The *x*s are the experimental data, and the solid lines are fits of the diffusion model, which will be discussed later. On the right of the *x*-axis, the probability of a response ranges from .5 to 1.0; responses with these probabilities are correct responses (*bright* responses to bright stimuli and *dark* responses to dark stimuli). On the left, responses with probabilities 0 to .5 are error responses (*bright* responses to dark stimuli and *dark* responses to bright stimuli). The RT quantiles were averaged across subjects (cf. Ratcliff, 1979) in order to display them.

The vertical spread of the five quantile points for each condition gives a picture of the shape of the RT distribution. With speed instructions, the quantiles show little change in distribution shape across the nine brightness × duration conditions. The distributions are all skewed to the right, as is shown by the larger spread between the .7 and .9 quantile points than between any of the other pairs of adjacent quantiles. For the more difficult conditions (those with a response probability of around .5), the tails of the distributions spread somewhat, relative to the easier conditions (which is what makes the mean RTs larger for the difficult conditions), but the effect is not large. The leading edge of the distributions (the .1 quantile) stays constant at about 330 msec.

With accuracy instructions, the differences between the distributions for the difficult versus easier conditions are much more pronounced. The .9 quantile point ranges from about 540 msec for responses with a probability of .95 to about 720 msec for responses with a probability of .5, a difference of about 180 msec. In contrast, there is almost no change in the .1 quantile across conditions. This pattern shows that the RT distribution spreads in the tail as accuracy goes from ceiling to floor.

With accuracy instructions, subjects produce error responses slower than correct responses for all but the conditions in which accuracy is highest; in the highest accuracy conditions, the errors are only a little slower than the correct responses. With speed instructions, errors are a little slower than the correct responses across all conditions, but essentially they show the same pattern as with accuracy instructions.

Figure 3 shows the value of quantile-probability functions in providing a compact summary of all the dependent variables. The quantiles show the shapes of the RT distributions at all the levels of accuracy, and they show how the shape changes as a function of accuracy. Comparison of the right to the left halves of the plots allows comparison of RTs for correct and error responses and shows how the relationship between them changes as a function of accuracy. Comparison of the plots for accuracy instructions to the plots for the speed instructions shows the squashing of the distributions because of faster responses with speed instructions.

Quantile-probability functions are parametric plots for sequential sampling models such as the diffusion model. The shapes of the lines that connect the quantile values across accuracy levels are determined by only a few parameters. In Figure 3, there are five quantiles for each of 5 conditions (grouped from 18 conditions), and for the diffusion model, the shapes of the five lines that connect the quantiles across conditions are completely determined by only three parameters, a, η, and s_z (drift rates determine the positions along the functions, and T_{er} determines the vertical location of the function). Fitting the five lines means fitting all the aspects of the data that they summarize, a strong challenge for a model.

FITTING THE DIFFUSION MODEL

As was mentioned in the introduction, Ratcliff and Rouder (2000) tested two versions of the diffusion model, one with drift rate reflecting the on and then off again availability of stimulus information, and the other with drift rate determined by a constant value, the integration of stimulus information over time until mask. The data in Figure 3 show the inverted-U-shaped pattern to be consistent with the constant drift rate model, and so this version of the model was used here.

As was noted in the results section, there was a bias toward *bright* responses. There are two ways to model bias in the diffusion model. One involves the bias in the criterion that separates positive from negative drift rates, which is analogous to the criterion in signal detection theory. The point between *bright* and *dark* responses that we thought would correspond to a drift rate of 0 was 50% white pixels. If subjects set their zero point differently, this would be a bias in their drift criterion. The other way of modeling bias is to move the starting point of the diffusion process nearer the boundary toward which the responses are biased. Both these ways of handling bias have been needed in fitting data (Ratcliff, 1985; Ratcliff et al., 1999). The two sources of bias have different empirical signatures. If the leading edge of the RT distribution, the .1 quantile RT, is shorter in a biased condition than in an unbiased condition, the data are modeled by moving the starting point of the diffusion process nearer the response boundary corresponding to the biased response. If the .1 quantile RT is about the same for the biased and the unbiased conditions, the data are modeled by shifts in the drift criterion.

The data in this experiment show little change in the .1 quantile RTs across conditions, which indicates that drift bias, not a bias in starting point, is needed. The subjects in the experiment set their drift criterion at a point brighter than 50% white pixels. The drift bias assumption is reasonable because subjects in a bright/dark discrimination task have to determine where on the bright–dark dimension to set the point that separates *bright* and *dark* responses. In real life, this point could vary over a large range, from pitch black to a car headlight in the face.

Bias in drift rate was added to the model with the parameter *vc*, defined as the offset of the subjects' drift criterion from the unbiased criterion (50% white pixels). For a stimulus with drift rate toward the *bright* boundary, *vc* was subtracted from the drift rate the stimulus would otherwise have, and for a stimulus with drift rate toward the *dark* boundary, *vc* was added to the drift rate it would otherwise have. A different value of *vc* was used for each of the three stimulus durations, because the biases were different at each stimulus duration.

The diffusion model was fit to the data by minimizing a chi-square value with a general SIMPLEX minimization routine (Nelder & Mead, 1965) that adjusts the parameters of the model to find the parameters that give the minimum chi-square value. The data entered into the minimization routine for each experimental condition were the RTs for each of the five quantiles for correct and error responses and the accuracy values. The quantile RTs were fed into the diffusion model, and for each quantile, the cumulative probability of a response by that point in time was generated from the model. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between each quantile. For the chi-square computation, these are the expected values, to be compared with the observed proportions of responses between the quantiles (multiplied by the number of observations). The observed proportions of responses for each quantile are the proportions of the distribution between successive quantiles (i.e., the proportions between 0, .1, .3, .5, .7, .9, and 1.0 are .1, .2, .2, .2, .2, and .1) multiplied by the probability correct for correct response distributions or the probability of error for error response distributions (in both cases, multiplied by the number of observations). Summing over (Observed − Expected)^2/Expected for all conditions gives a single chi-square value to be minimized.

Research on fitting the diffusion model to data (Ratcliff & Tuerlinckx, in press) found two problems with this chi-square fitting method. First, when long or short outlier RTs were added to simulated data, the method could not accurately recover the parameter values that were used to generate the data. Second, variability in the fastest RTs across conditions (e.g., the .1 quantile) also led to poor recovery of parameter values and poor fits to the data.

Short outliers can be trimmed out by examining the time at which accuracy begins to rise above chance (e.g., Swensson, 1972), but this does not eliminate the problem with variability in the fastest responses. Because the chi-square method has to produce probability density below the shortest quantile in the data set (in order to compute an expected value below the .1 quantile), any variability above what the model predicts in the shortest quantile distorts the fits. This is because the chi-square computation has the expected value on the denominator, so that a very small expected value will produce a very large value of chi-square and so dominate the fitting process. In practice, variability in the shortest quantile produces large misfits in the longest quantiles (e.g., up to 400-msec misses for the .9 quantile).

To address these problems, the two possibilities were explicitly modeled, which added two parameters to the model (Ratcliff & Tuerlinckx, in press). One parameter (p_o), was added to represent the probability of a contaminant in each condition of the experiment. The contaminant was assumed to come from a uniform distribution that had maximum and minimum values corresponding to the maximum and minimum RTs in the condition. The value of the probability parameter, p_o, was the same across all experimental conditions. There may be better ways of estimating the range or distribution of contaminants, but the small proportion usually estimated (less than 5%), the ease of implementation, the fact that this adds only one parameter to the model, and the ability to recover parameter values better than do methods without this assumption (Ratcliff & Tuerlinckx, in press) all indicate that this was a reasonable approximation.

The second added parameter represented variability in T_{er} across trials. It is quite reasonable to assume that the components of processing other than the decision component vary across trials, but this assumption has not usually been needed in fitting the model to data. However, it turns out that for the data presented here, variability in T_{er} is needed to deal with variability in the fastest responses—that is, in the .1 quantile. For ease of implementation, the distribution of T_{er} was assumed to be uniform, with a range of s_t. Because the standard deviation in the distribution of T_{er} is typically less than one quarter the standard deviation in the decision process, the combination of the two (convolution) alters neither the distribution nor the standard deviation in the distribution predicted from the decision process. For example, if the standard deviation in T_{er} is 25 msec and the standard deviation in the decision process is 100 msec, the combination (square root of the sum of squares) is 103 msec. Variability in T_{er} stretches out the leading edge of the RT distribution (e.g., the .1 quantile RT), stretching the difference between the .1 and .3 quantiles (typically, by less than 10% of s_t). In fitting data, variability in T_{er} allows larger variability in the leading edge of the distribution from condition to condition and prevents T_{er} from being determined completely by the shortest quantile. When Ratcliff and Tuerlinckx (in press) simulated data with variability in T_{er}, the chi-square method recovered parameter values accurately.

The downside of these two additions to the diffusion model is to increase the standard deviations in the recovered parameter values, because of the two added parameters. However, the model fits the data with these additions, whereas without them there were significant misses.

The diffusion model, as fitted to the data presented here, has 19 parameters that are free to vary. The data to be fitted have 36 conditions (speed and accuracy crossed with six levels of brightness crossed with three stimulus durations), and in each condition, there are 12 RT bins from five correct quantiles and five error quantiles. This means there are $(12 - 1) * 36$ degrees of freedom in the data. With 19 parameters to be fit, there are 377 degrees of freedom. However, the model is much more constrained than might be expected with this number of free parameters. First, the nine drift rates and the three drift bias parameters determine position along the x-axis of the quantile-probability function. They do not influence the shapes of the functions. Second, the value of T_{er} only locates the vertical positions of the quantile-probability functions; it also does not influence their shape. Third, the estimate of the proportion of contaminant responses (p_o) is close to zero and has no effect on the predicted quantile probability functions relative to the case in which it is zero. Fourth, although variability in T_{er} (s_t) allows the .1 quantile RTs to be more variable and more in line with the observed variability, its only effect is a less than 7-msec increase in the difference between the .1 and the .3 quantile RTs. Fifth, the shape of the quantile-probability functions is determined only by the parameters η (standard deviation in drift across trials), s_z (range of starting point across trials), and the values of boundary separation (a), one value for the speed conditions and one for the accuracy conditions. All the parameters except a were held constant across the speed and accuracy data. The starting point z was set to $a/2$. Thus, the shape and location of the quantile-probability functions are determined only by four parameters, with only one of them varying between the speed and the accuracy conditions.

FITS OF THE MODEL TO DATA

Tables 1 and 2 show the parameter values of the fits of the model to the experimental data. First, the fits to the group average data, combining the data from all 3 subjects, are presented; next, the means of the parameter values from fits of the model to the data from the individual subjects are presented; and third, the parameter values for the fits to each individual subject are presented. The fits to the group data are shown in Figure 3 by solid lines.

Typically, in applications of the diffusion model (Ratcliff & Rouder, 1998, Experiment 1; Ratcliff et al., 2001, Experiment 2), differences in performance between speed and accuracy conditions are fully accounted for by a change in the boundary separation parameter a. However, occasionally some subjects might "hold up" their responses in accuracy conditions, leading to a slowdown in all the responses, which could be modeled by an increase in T_{er}. To check for this, we fitted the diffusion model to the group data with one value of T_{er} for the speed condition and a sep-

Table 1
Parameters for Fits of the Diffusion Model

Fit	a (speed)	a (accuracy)	T_{er}	η	s_z	p_o	s_t
Fits to average data	.0700	.0998	.3650	.2532	.0556	.0003	.1133
Means over single subjects	.0674	.0986	.3538	.1803	.0355	.0033	.1187
Fit to Subject 1	.0669	.0950	.3619	.1975	.0454	.0000	.1106
Fit to Subject 2	.0628	.0910	.3284	.1923	.0399	.0000	.1175
Fit to Subject 3	.0724	.1099	.3711	.1512	.0211	.0098	.1281

Note—Chi-square values for the fits for the 3 subjects are $\chi^2(11{,}941) = 754$, $\chi^2(10{,}956) = 2{,}535$, $\chi^2(9{,}706) = 2{,}089$, $df = 377$.

Table 2
Drift Rates for Fits of the Diffusion Model

Brightness	Flash Time (msec)	Drift Rate: Fit to Average Data	Mean Over the Single Subjects	Subject 1	Subject 2	Subject 3
.350 and .650	50	.5301	.4307	.4672	.3609	.4640
.425 and .575	50	.2595	.2064	.2234	.1363	.2595
.475 and .525	50	.0632	.0615	.0591	.0488	.0767
.350 and .650	100	.6083	.4866	.5414	.4127	.5057
.425 and .575	100	.3878	.3049	.3229	.2638	.3281
.475 and .525	100	.1398	.0713	.1130	.0978	.0031
.350 and .650	150	.6117	.4960	.5066	.4320	.5495
.425 and .575	150	.4119	.3208	.3527	.2842	.3256
.475 and .525	150	.1492	.1255	.1345	.1091	.1329
Bias	50	.2661	.1958	.2335	.2844	.0695
(subtract from	100	.0675	.0576	.0541	.1104	.0084
above drift rates)	150	.0273	.0233	.0411	.1035	−.0745

arate value for the accuracy condition, as well as two values of boundary separation a. The two best fitting values of T_{er} were 333 and 343 msec, close enough that we report fits of the model using only one value of T_{er}.

The only parameter that changes between the speed and the accuracy conditions for all the data sets is boundary separation a. The shapes of the quantile probability functions across accuracy values (the five solid lines in each panel of Figure 3) are determined entirely by the parameters a, η, and s_z. The vertical location of the function is determined by T_{er}. Drift rates for the different conditions determine only the accuracy values along the horizontal axis of the quantile probability functions, not the shapes of the functions or their vertical placement. The estimated range (s_t) of times for nondecision components of processing (T_{er}) was between 110 and 130 msec (note that the standard deviation in a uniform distribution is the range divided by the square root of 12, so if the range is 110 msec, the standard deviation is $110/\sqrt{12} = 32$ msec). Mean values of T_{er} across subjects were between 330 and 370 msec. The estimated proportion of contaminants (p_o) was less than 1%.

As might be expected, drift rate, representing the quality of the information extracted from the stimulus, increases with brightness and stimulus duration. The bias in drift rate (Table 2) was the same for the speed and the accuracy conditions (see also Figure 2). Drift rates are plotted in Figure 4 for the different brightness conditions as a function of stimulus duration. The functions increase toward asymptote with different asymptotes for different values of brightness. The drift rates are close to asymptote by 100 msec of stimulus processing time, which means that 100 msec is enough time to extract all the information available from the stimulus (as in Rayner et al., 1981).

Figure 4 also contains a plot of drift rate against stimulus duration for a letter identification experiment (Experiment 1, averaging over easy and hard conditions in the experiment; Ratcliff & Rouder, 2000). The function shows the same qualitative shape as that for our brightness discrimination functions (although we would not want to equate the processes underlying the two tasks). This shows a converging result from a experiment with similar design, but with different stimuli.

It is important to note that the diffusion model extracts from the whole set of data (correct and error RTs and accuracy) a single quantity, drift rate, for each condition. Drift rate is a measure of the quality of stimulus information that drives the decision process.

In modeling extraction of information from a stimulus, it is usually assumed that information is gradually extracted over time (see Busey & Loftus, 1994; Smith, 1995). This suggests that drift rate in the diffusion model should not, at the beginning of the decision process, abruptly change from a value of zero to some fixed high value. More realistically, drift rate might be expected to increase from zero over the first 20–50 msec of processing. In fitting the diffusion model to data, the abrupt change from zero to a fixed value has always been assumed, but here we address the

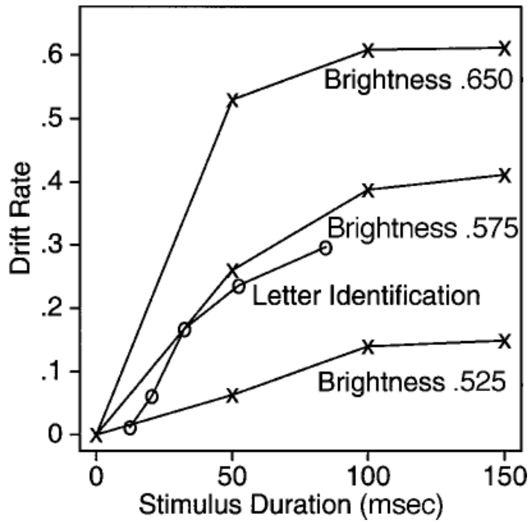

Figure 4. Drift rate from the fits of the diffusion model as a function of stimulus duration and brightness (*x*s) and drift rate as a function of stimulus duration in a letter identification experiment (*o*s) from Ratcliff and Rouder (2000, Experiment 1).

question of what happens to fits of the model if a gradual increase in drift rate is assumed instead.

To answer this question, simulated data were generated from the diffusion model, with a linear increase in the drift rate over the first 50 msec of processing from zero to a constant value after 50 msec (see Busey & Loftus, 1994). Accuracy values and quantile RTs were generated under the assumptions that there were no contaminant RTs and no variability in T_{er} (i.e., $s_t = 0$). Data were generated for four experimental conditions, with drift rates of .3, .2, .1, and 0, with 100,000 observations per condition. The other parameters of the model for the simulations were $a = .12$, $T_{er} = 300$ msec, $\eta = .08$, and $s_z = .02$. The model was then fit to the simulated data with the assumption that drift rate was constant.

Although the simulated data were generated with a gradual increase in drift rate over 50 msec, fitting the model under the assumption of an immediate abrupt increase in drift rate led to accurate recovery of the parameter values for a, η, and the drift rates (within 0.5%, 10%, and 3%, respectively). However, variability in starting point (s_z) was increased from .02 to .066, T_{er} was increased by 28 msec, and variability in T_{er}, s_t, was estimated to be 55 msec when the value used to generate the simulated data was zero. The simulated quantile RTs were within 25 msec of the predictions from the fitted model, the accuracy values were within 2%, and there were no systematic deviations of the simulated data from the fits.

In sum, the effect of fitting a decision process with a linear rise in drift rate with a model without a gradual rise was to produce, first, an increased value of T_{er}; second, an estimated value of variability in T_{er} greater than zero ($s_t >$ 0); and third, increased variability in the starting point (z) of the diffusion process. Thus, the effect of a gradual rise in drift rate cannot be distinguished from increased variability in the starting point, z, and increased variability in the nondecisional component of response time, T_{er}.

HOW GOOD IS THE FIT?

The issues of goodness of fit and model freedom are usually brought up in casual conversations (or by reviewers) with the comment "with 8 or 16 parameters you can fit any pattern of results." Slamecka (1991) voiced this view about the global memory models: "If the models have anything, they have resilience, or to put it more precisely, their inventors have resilience, and I suspect that after some skillful patching of assumptions and/or fine-tuning of some parameters, these veterans will lumber down the runway and lift off again" (pp. 303–304). For many experimental manipulations in the memory domain, it is true that the global memory models are capable of fitting almost any pattern of data, but it is also the case that the models make strong predictions that are testable (and the data that Slamecka addressed raised problems for the models and could not be fit by them; in part, this led to the development of new models).

Roberts and Pashler (2000) have presented a comprehensive discussion of constraints on models. They asked whether a good fit alone gives any reason to increase belief in a theory. One of the main pieces of information that is needed, they argued, is information about what kinds of patterns the theory cannot accommodate. If a theory can fit almost any pattern of results and if the patterns of results that are obtained experimentally are significantly more restricted than the range of possible predictions of the model, the theory can predict patterns of data that do not occur. Roberts and Pashler argued that a theory like this is not supported by any fit to any individual set of data. They argued that what is needed to support a theory is the existence of patterns of results that neither occur experimentally nor can be fit by the theory.

Researchers working with stochastic models for RT and accuracy have known that their models are inflexible—that is, that there are many possible patterns of data the models cannot fit—but there have been few attempts to systematically demonstrate their lack of flexibility. To begin to remedy this for the diffusion model, in this section I present a number of patterns of results that the diffusion model cannot fit. I generated fake data by hand (often by altering predictions of the model) and then attempted to fit the diffusion model to them. With a quantile-probability plot, it is possible to display simultaneously all aspects of the data and show how the diffusion model fails.

The first thing to note is that the diffusion model can almost always, for any single experimental condition, fit the condition's accuracy value and two mean RTs, one for correct and one for error responses. It can do this with many different combinations of parameter values. One aspect of the data that does provide at least some constraint is the shape of RT distributions. The model cannot fit RT distributions that are not positively skewed by about the right amount. For example (as I demonstrate below), the diffusion model cannot fit normal or close to normal RT distributions or highly skewed distributions. The distribution shape predicted from the model is right skewed with a tail that is approximately exponential at the extreme right (e.g., Ratcliff et al., 1999, Figures 12 and 13, show flat hazard functions in the right tail, which means that the density function is exponential in the right tail; see also Luce, 1986). However, although distribution shape does constrain the model, it does not constrain it sufficiently to provide unique estimates of parameter values for a single condition.

Because a single experimental condition can be fit with many combinations of parameter values, in order to uniquely determine the parameters that underlie performance on a particular task, experimental manipulations are needed that constrain the parameters of the model. This can be done by varying the difficulty of the task from easy (high accuracy and fast responses) to difficult (low accuracy and slow responses) in an experimental design in which items are presented in random order so that a subject cannot know before an item is presented what condition it is in. In this situation, subjects cannot alter their response criteria or the nondecision components of processing time, so these parameters of the model are fixed. Then, the only parameter of the model that is free across experimental conditions is drift rate, representing the amount of stimulus information.

As was noted earlier, across-trial variability in drift rate and starting point allows the model to produce empirically observed patterns of error versus correct RTs. These include errors always faster than correct responses (variability in starting point dominates), errors always slower than correct responses (variability in drift rate dominates), or a crossover in which errors in high-accuracy conditions (e.g., accuracy greater than 90%) are faster than correct responses, whereas errors in lower accuracy conditions are slower than correct responses. With variability in drift

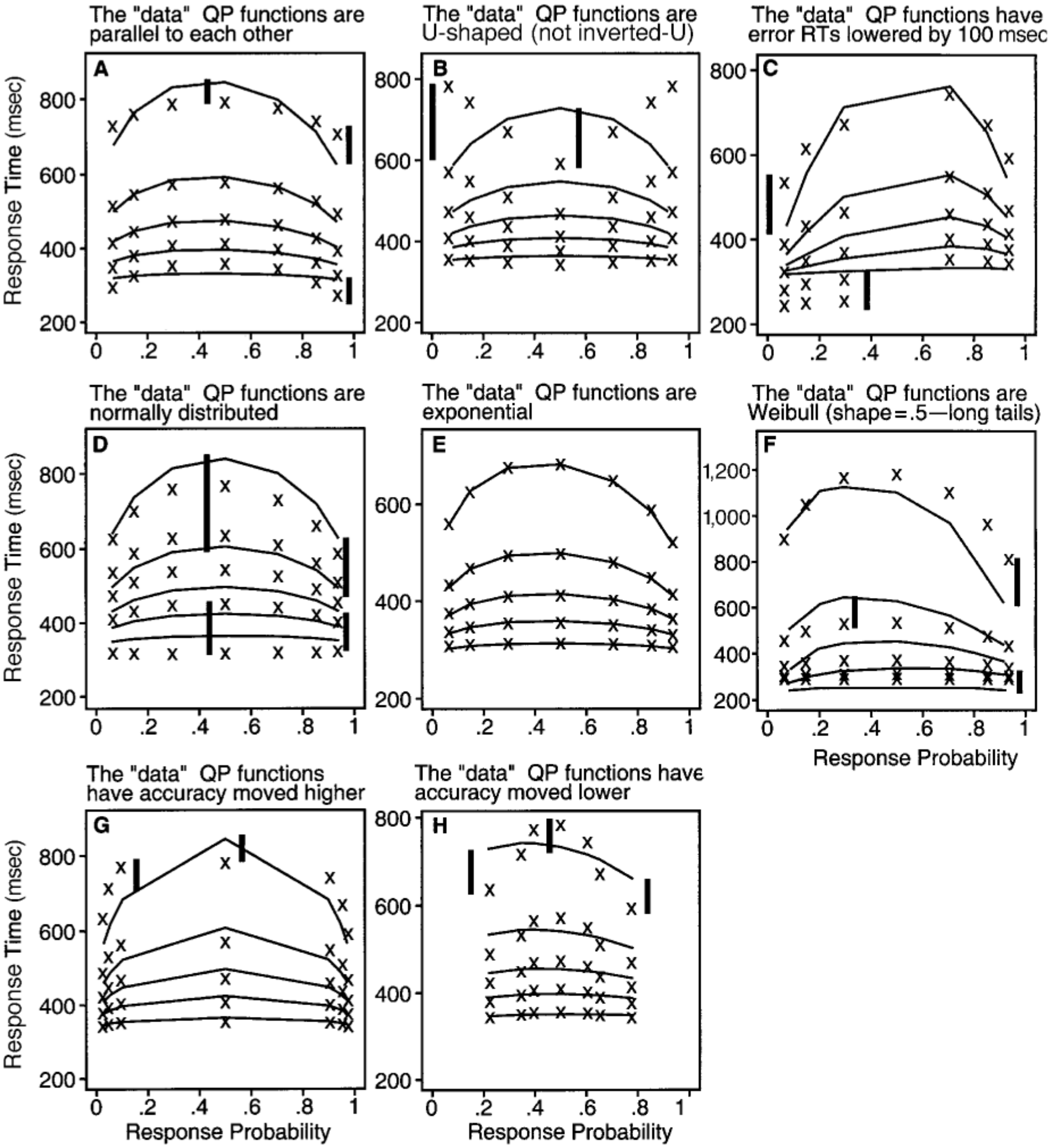

Figure 5. Eight patterns of made-up data (the *x*s) and the fits of the diffusion model (the solid lines). Only drift rate changes across conditions (i.e., along the quantile-probability functions). The heavy dark bars show points of major mispredictions between fake data and fits of the diffusion model.

rate and starting point across trials, the diffusion model cannot produce a pattern of results in which errors in highly accurate conditions are slower than correct responses and errors in less accurate conditions are faster than correct responses (so long as all the other parameters are fixed across conditions).

Figure 5 displays eight quantile-probability functions, seven of which are never observed in real data. These are the data I generated by hand, and they are represented by the *x*s in the figure. Four experimental conditions are represented, such as might be used in a letter identification task in which subjects are asked to identify which of two letters was displayed (Ratcliff & Rouder, 2000), and there are four levels of stimulus presentation time. On each quantile-probability plot, the three points to the right of .5 are correct responses, the three points to the left are error responses (correct and error responses for zero drift are coincident and are represented by the middle point in the figures).

The diffusion model was fit to the data in the eight panels in Figure 5. For the quantile-probability functions in each panel, the only parameter allowed to vary between conditions was drift rate. In all except one of the fake data cases, the model failed to fit the data well. The largest or most important misses between the fits and the data are highlighted with dark bars.

For the data in panel A, the differences in the quantile RTs are the same for each experimental condition—that is, each vertical line of *x*s. For example, if the .9 quantile is 50 msec longer than the .7 quantile in the most accurate experimental condition, it is also 50 msec longer in the next most accurate condition and in all the other conditions. This pattern would come about if the changes in mean RTs across conditions were the result of a shift in the whole RT distribution, with the shape of the distribution remaining constant. The diffusion model cannot fit this pattern; the model requires that changes in mean RTs across conditions be accompanied by changes in distribution shape. As drift rate decreases and RT slows, the RT distribution must skew, which means that the .9 quantile RTs slow a lot and the .1 quantile RTs slow a little. Data like the fake data in panel A have been collected in judgment-of-recency tasks (Hacker, 1980; Muter, 1979), and the model cannot fit the data from these tasks with only changes in drift rate across conditions.

The fake data in panel B show U-shaped quantile-probability functions with accurate correct responses very slow and inaccurate correct responses very fast. With only drift rate changing as a function of condition, the diffusion model cannot produce this pattern. The best-fitting functions show the inverted-U-shaped functions that the model predicts.

In panel C, for each experimental condition, the RT for error responses is shorter than the RT for correct responses, with much of the RT difference being the result of a shift in the distribution. For example, correct responses at the .1 quantile in the most accurate condition are over 100 msec longer than the corresponding .1 quantiles for error RTs. The diffusion model cannot fit these data because it requires .1 quantile RTs to be about equal across all conditions for both correct and error responses. For the same reason, the model could not fit data for which the error RT distribution was slowed relative to the correct RT distribution.

Panels D, E, and F show the quantile-probability functions for data with normally distributed RT distributions, moderately skewed RT distributions, and highly skewed RT distributions, respectively. The fact that the model fits the quantile-probability functions in panels D and F so poorly shows its inflexibility with respect to distribution shape. For the normally distributed data in panel D, the poor fit comes from the model's requirement for right skewed distributions, reflected in the wider separations between the upper quantiles of the quantile-probability functions than between the lower quantiles. Exponential RT distributions (panel E) are well fit by the model. The tail of the distribution generated by the diffusion model is roughly exponential, but the initial rise is not. However, when responses are quite fast, as they are with the parameter values with which these fake data were generated, the rise in the distribution generated by the model is quite abrupt and peaks at about the .1 quantile RT. For this reason, the model approximates the abrupt rise in the exponential as displayed with only five quantiles.

For the data in panel F, the distributions have highly skewed tails. They were generated from a Weibull distribution with a shape parameter of .5. The diffusion model cannot fit such distributions, and attempts to fit the data produce distributions with much shorter tails. The quantile-probability functions show this with wider separations between the .7 and .9 quantiles in the fake data than in the best-fitting quantiles from the model and with narrower separations between the .1 and .3 quantiles in the fake data than in the best-fitting quantiles.

For panels G and H, the accuracy values for each experimental condition were increased and decreased, respectively, relative to the values generated from the model, keeping the same RTs. Increasing the accuracy values shifts the quantile RTs outward from the .5 probability point, which leads to a wide flat quantile-probability function. When the accuracy values are decreased, the quantile RTs are shifted inward, leading to a narrow peaked quantile-probability function. The diffusion model cannot fit either pattern of fake data. The model cannot produce the flat quantiles over the wide range of accuracy values with a large drop in RT at the extreme in panel E. Also, it is not able to fit large changes in RT over the moderate range of accuracy values shown in panel F.

The experiment presented in this article included a speed–accuracy manipulation. On separate blocks of trials, the subjects were instructed either to respond quickly or to respond accurately. The assumption made in fitting the diffusion model was that only boundary separation (a) changed between speed and accuracy blocks of trials. The experimental data were well fit by the diffusion model with this assumption.

Figure 6 shows examples of the fit of the model to plausible patterns of fake data under the assumption that only

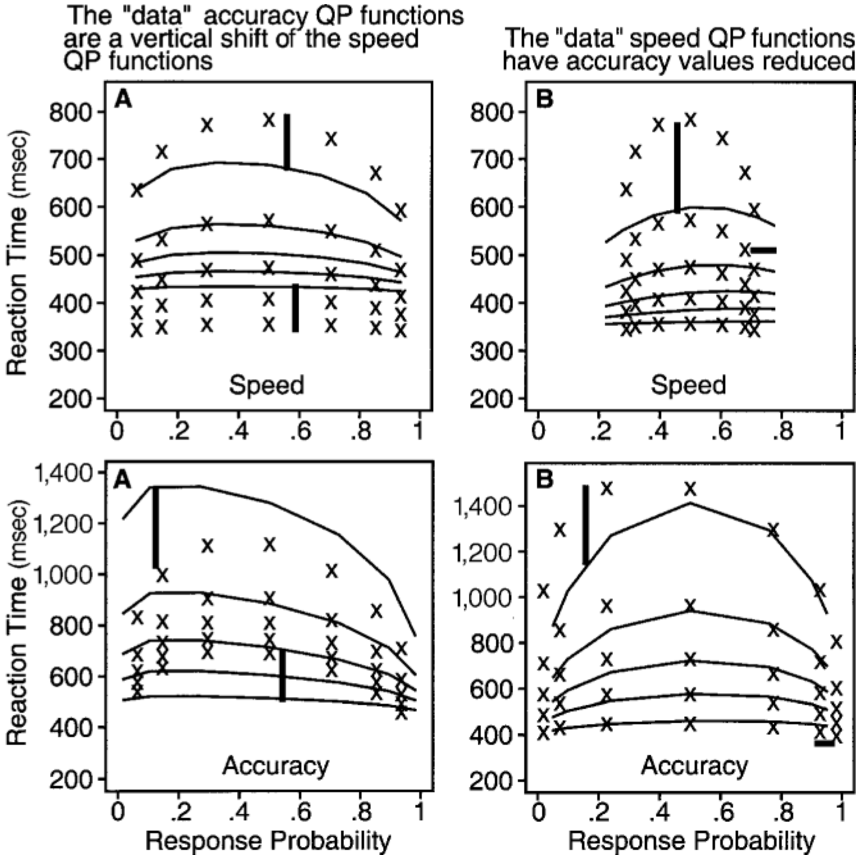

Figure 6. Two patterns of made-up data to be fit with a speed–accuracy manipulation. In the diffusion model, only boundary separation is allowed to change between speed and accuracy conditions. The heavy dark bars show points of major mispredictions between fake data and fits of the diffusion model.

boundary separation changes between speed and accuracy conditions. In the A panels, the data for the speed condition were produced from the diffusion model, and the data for the accuracy condition were fake; in the B panels, the data for the accuracy condition were produced from the diffusion model, and the data for the speed condition were fake.

For the A panels, the data for the speed and the accuracy conditions are identical, except that a constant time was added to each quantile RT in the speed condition to produce the accuracy condition. The diffusion model fits show a smaller change between the speed and the accuracy conditions in the .1 quantile RTs than the fake data do and a much larger change in the .9 quantile RTs than the fake data do.

In the fake data for the B panels, the speed conditions have very low accuracy values, and the accuracy conditions have much higher accuracy values. The model cannot produce large changes in the .7 and .9 quantile RTs in the speed conditions, along with low values of accuracy.

Except for the data in Figure 5, panel E, these patterns of data are all patterns that the diffusion model cannot fit, and with the one exception noted, they are patterns that do not appear in real empirical data within the scope of application of the diffusion model. If one of these patterns did appear in real data, the diffusion model would fail. Such a pattern might have a simple explanation. For example, subjects who are extremely concerned about accuracy might hold up their responses to do a double check of their response. Setting aside such possibilities, there are currently few experiments that have produced qualitative patterns of data like the ones presented above that are completely inconsistent with the predictions of the diffusion model.

It is important to understand that the course of development for the diffusion model was not one in which the basic structure of the model was repeatedly adjusted and redesigned until it was able to fit data. Rather, the basic structure of the model is fixed by its mathematics. The mathematics determine first, inverted-U-shaped quantile-probability functions; second, right skewed RT distributions; and third, distributions that skew rather than shift as drift rate is varied. Additions of variability in parameters across trials, as was discussed earlier, allow the model to produce fast or slow errors, but these do not affect the overall shape of the quantile-probability function, the overall relationship between correct RTs and accuracy, or the shape of the RT distributions.

Figures 5 and 6 show some of the patterns of results that the diffusion model cannot fit, demonstrating that the model is much more constrained than the "with 8 parameters you can fit anything" view. These demonstrations address the concerns discussed by Roberts and Pashler (2000) and help to indicate why people working with models of the sequential sampling class believe they are strongly constrained.

DISCUSSION

The experiment presented here combined manipulations of stimulus brightness and duration in a two-choice brightness discrimination paradigm. Two-choice paradigms using RT measures have had limited use in brightness and many other perceptual discrimination paradigms, because the two dependent variables, RT and accuracy, are difficult to model simultaneously. The two measures have different scales, and their values change by different amounts at different points on their respective scales. Interpretations of such data are extremely difficult in the absence of a model that relates the two variables. This is because an interpretation based on mean RT may lead to one set of conclusions (e.g., serial scanning), whereas an interpretation based on the other dependent variable may lead to a different conclusion (e.g., a signal detection analysis). The diffusion model has been successful in other similar situations, and it is successful with this paradigm, providing an account for the behavior of both dependent variables and extracting a single measure of stimulus information, drift rate, that drives the decision process.

The model also accounts for the effects of speed versus accuracy criteria settings with a single parameter. Performance under speed versus accuracy stress is different in terms of accuracy, RTs for correct and error responses, and the behaviors of RT distributions, yet all the differences are explained by a change in the single parameter, boundary separation (a).

There are two main possibilities about how drift rate might represent stimulus duration. One is that drift rate increases when the stimulus is presented, increasing at a rate determined by the brightness, and then decreases back toward zero when the stimulus is masked. The other is that drift rate is constant at a value reflecting the product of brightness and stimulus duration. The first possibility predicts slow errors, becoming dramatically slower as accuracy increases and error rate decreases (see Ratcliff & Rouder, 2000, Figure 3). This was not the pattern obtained. Instead, the error RTs slowed a little, and then sped up as accuracy increased from 50% to over 90%. This is consistent with the constant drift hypothesis (see Ratcliff & Rouder, 2000). Drift rate increased to asymptote as a function of stimulus duration, and the asymptotic value of drift rate was determined by brightness (the proportion of white vs. dark pixels).

The results in this article show that the diffusion model is capable of explaining data from a choice RT paradigm that is more perceptual than other paradigms to which the model has been applied (see also Smith, 1995). The application of the model shows how choice RT methods can be used to address perceptual questions, how the data can be interpreted in terms of the model, and how the variables driving performance can be extracted from the joint RT- and accuracy-dependent variables.

REFERENCES

Bloch, A. M. (1885). Expériences sur la vision. *Comptes Rendus de la Societé de Biologie*, **37**, 493-495.

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

Carterette, E. C., Friedman, M. P., & Cosmides, R. (1965). Reaction-time distributions in the detection of weak signals in noise. *Journal of the Acoustical Society of America*, **38**, 531-542.

Espinoza-Varas, B., & Watson, C. S. (1994). Effects of decision criterion on response latencies of binary decisions. *Perception & Psychophysics*, **55**, 190-203.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **6**, 651-675.

Kellogg, W. N. (1931). The time of judgement in psychometric measures. *American Journal of Psychology*, **43**, 65-86.

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

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

Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Erlbaum.

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

Muter, P. A. (1979). Response latencies in discriminations of recency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **5**, 160-169.

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

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

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

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

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

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

Ratcliff, R., Gomez, P., & McKoon, G. (2002). *Diffusion model account of lexical decision*. Manuscript submitted for publication.

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

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

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

Ratcliff, R., & Tuerlinckx, F. (in press). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter. *Psychonomic Bulletin & Review*.

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

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 167-179.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, **107**, 358-367.

Sekuler, R. W. (1965). Signal detection, choice response times, and visual backward masking. *Canadian Journal of Psychology*, **19**, 118-132.

Slamecka, N. J. (1991). The analysis of recognition. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 297-306). Hillsdale, NJ: Erlbaum.

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

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

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

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

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

Watson, A. B. (1986). Temporal sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 6-1 to 6-43). New York: Wiley.

(Manuscript received July 24, 2000;
revision accepted for publication June 20, 2001.)