Interpreting the parameters of the diffusion model: An empirical validation

ANDREAS VOSS
Albert-Ludwigs-Universität, Freiburg, Germany

KLAUS ROTHERMUND
Friedrich-Schiller-Universität, Jena, Germany

and

JOCHEN VOSS
University of Warwick, Coventry, England

The diffusion model (Ratcliff, 1978) allows for the statistical separation of different components of a speeded binary decision process (decision threshold, bias, information uptake, and motor response). These components are represented by different parameters of the model. Two experiments were conducted to test the interpretational validity of the parameters. Using a color discrimination task, we investigated whether experimental manipulations of specific aspects of the decision process had specific effects on the corresponding parameters in a diffusion model data analysis (see Ratcliff, 2002; Ratcliff & Rouder, 1998; Ratcliff, Thapar, & McKoon, 2001, 2003). In support of the model, we found that (1) decision thresholds were higher when we induced accuracy motivation, (2) drift rates (i.e., information uptake) were lower when stimuli were harder to discriminate, (3) the motor components were increased when a more difficult form of response was required, and (4) the process was biased toward *rewarded* responses.

Ratcliff (1978; Ratcliff & Tuerlinckx, 2002; Ratcliff, Van Zandt, & McKoon, 1999) has proposed the diffusion model as a form of data analysis for speeded binary decisions. The diffusion model assumes that binary decisions are based on a continuous process that fluctuates between two possible outcomes (Figure 1). As soon as the process reaches a critical upper or lower value, a decision is made, and the corresponding response is executed. One main advantage of the diffusion model is that different components of the decision process (rate of information uptake, bias, conservatism, and motor component) are represented by different parameters of the model. Theoretically, distinct aspects of the decision process can be separated statistically. In a sense, the model allows inferences regarding hidden cognitive processes. Another advantage of the diffusion model data analysis is the high degree of information utilization. In contrast to conventional forms of data analysis, the diffusion model incorporates response times (RTs) for correct responses and errors, as well as the ratio of correct and erroneous responses.

The diffusion model thus provides a powerful statistical tool that allows a very fine-grained analysis of the cognitive processes underlying simple response tasks. Previous applications of the diffusion model have included, for example, the analysis of retrieval processes (Ratcliff, 1978, 1988) and the identification of the factors underlying age-related slowing (Ratcliff, Spieler, & McKoon, 2000; Ratcliff, Thapar, & McKoon, 2001, 2003). These examples reveal that the diffusion model has a wide range of possible applications and can be used to decide between competing theoretical interpretations of experimental findings. However, although interpretations of the different parameters of the diffusion model seem fairly straightforward, it should be kept in mind that these interpretations are based mainly on theoretical reflections and presuppose the validity of the diffusion model as an adequate representation of the underlying cognitive processes.[1] The main aim of the present article, therefore, is to provide an *empirical* validation of the theoretical interpretation of the parameters of the diffusion model. To this end, we investigated the effects of different experimental manipulations on the parameter set in a diffusion model data analysis. The experimental manipulations were intended to specifically affect the different processing components of the model (rate of information uptake, setting of response criteria, duration of the motor response and bias; see Ratcliff, 2002; Ratcliff & Rouder, 1998; Ratcliff et al., 2001). Analyzing whether experimental manipulations of the different processing components that are specified in the model map uniquely onto the corresponding parameters of the model (and not
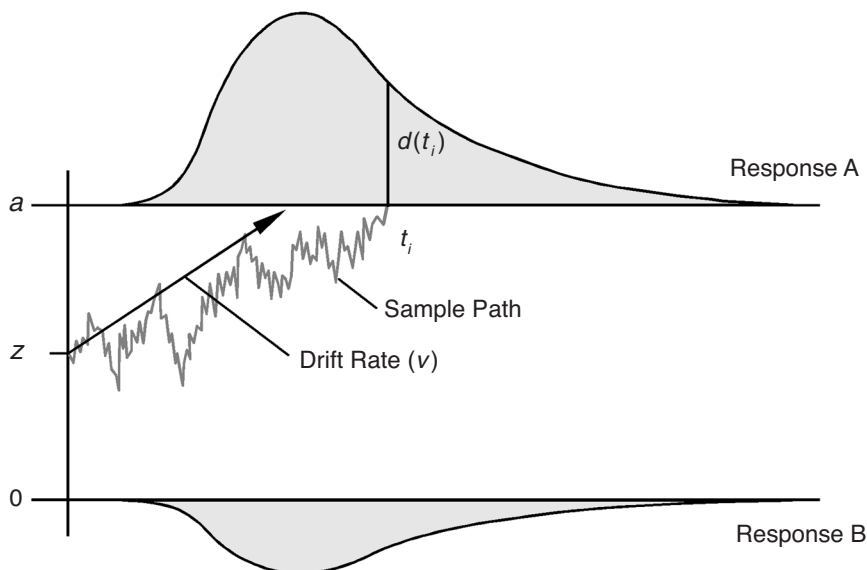
Figure 1. Schematic illustration of the diffusion model. The process starts at point *z* and moves over time until it reaches the lower threshold at 0 or the upper threshold at *a*. Depending on which threshold is reached, different responses are initiated. The erratic line gives an example of one process that reaches the upper threshold at time t_i. The mean deviance of single paths from the mean path is defined by an additional (scaling) parameter of the model (i.e., the diffusion constant *s*). Outside the thresholds, the distributions of process durations for a constant set of parameters are sketched. Note that the lower limit can be reached with random influences, although the distributions shown are based on a positive drift rate.

onto others) is a strong empirical test of the model's validity. Some of these manipulations have been used in a similar manner by others (e.g., Ratcliff, 2002; Ratcliff & Rouder, 1998; Ratcliff et al., 2001, 2003). In the present article, however, a somewhat different focus is adopted by addressing two fundamental points: First, all the important parameters of the diffusion model are validated within the same paradigm, and second, the question of validation is separated from other research questions (e.g., age-related changes in information processing); this separation allows for clear interpretation of the experimental results.

A second objective of this article—next to the validation of the model parameters—is to introduce a new method of estimating the parameters of the diffusion model based on the Kolmogorov–Smirnov (KS) test (Kolmogorov, 1941). This approach is an alternative to the more established methods (i.e., maximum likelihood, chi-square, and weighted least square; see Ratcliff & Tuerlinckx, 2002, for a detailed discussion of the different algorithms). The KS statistic also provides a test of the model fit.

In the following, we first will give a brief overview of the rationale of the diffusion model. In a second step, we will describe our estimating algorithm and the new model test. Finally, we will report two experiments with a simple color classification task in which the effects of different experimental manipulations on the parameter set of the diffusion model were investigated. Mathematical details of the diffusion model are given in the Appendix.

THE RATIONALE OF THE DIFFUSION MODEL

The diffusion model assumes that binary decision processes are driven by systematic and random influences. The systematic influences are called the *drift* rate. They "drive" the process continuously in one direction, while the random influences add an erratic fluctuation to this constant path. The process is terminated immediately when one of two thresholds is reached. In a psychological application, this means that the decision process is finished and the response system is being activated, which will initiate the corresponding response (e.g., a keypress). With a given set of parameters, the model predicts distributions of process durations (i.e., RTs) for the two possible outcomes of the process. Figure 1 provides a schematic illustration of the model: The process begins at the starting point *z* and moves over time until it reaches the lower threshold at zero or the upper threshold at *a*. Then the corresponding response is executed immediately. In Figure 1, a sample path of the process is presented. As can be seen, this path is not straight but oscillates between the two boundaries, due to random influences (i.e., the intratrial variability of the path) on the process.

Applied to a perceptual situation in which participants have to categorize ambiguous stimuli, the process describes the ratio of information gathered that fosters each of the two possible stimulus interpretations. Information is collected over time until evidence points with sufficient clearness to one interpretation; then a response is

initiated. As was mentioned above, there are random influences that cause variance in the RT distributions and also generate erroneous responses in some trials. For example, the different sizes of the shaded areas that are constituted by the time distributions in Figure 1 denote that in most cases, the process reaches the upper threshold, indicating a positive drift rate. Nonetheless, in some trials, the random influences outweigh the drift, and the process terminates at the lower threshold.

Parameters of the Diffusion Model

The model is defined by a set of parameters (see Table 1), which will be described in the following section: The *upper threshold* (a), which is also the distance between both limits, has already been introduced. This parameter is interpreted as a measure of *conservatism*: A large value of a indicates that the process takes more time to reach one limit and that it will reach the limit opposite to the drift less frequently. The second parameter is the *starting point* (z), or *bias*. This parameter reflects differences in the amount of information that is required before the alternative responses are initiated. If z exceeds $a/2$, a smaller (relative) amount of information supporting Interpretation A is needed to give Response A, as compared with a larger (relative) amount of information supporting Interpretation B that is required before Response B is executed. In other words, participants are more liberal in giving the answer connected to the upper threshold (Response A) and more conservative in giving the opposite response (Response B). If z is smaller than $a/2$, the contrary is true. If z diverges from $a/2$, the average decision times will differ for the different responses: The smaller the distance between the starting point and the limit, the shorter the process durations will be for the corresponding responses.

The third parameter of the diffusion model is the *drift rate* (v), which stands for the mean rate of approach to the upper threshold (negative values indicate an approach to the lower threshold). The drift rate indicates the relative amount of information per time unit that is absorbed. Therefore, the drift can be interpreted as a measure of perceptual sensitivity (in a between-person comparison) or as a measure of task difficulty (in a between-condition comparison). The path of the process is also influenced by (normally distributed) random influences (i.e., the *noise* in the model). The amount of the random influences is given by the diffusion constant (s), which is a

scaling parameter. This means that s is not a free parameter to be estimated but has to be fixed to any (positive) constant value. In the Appendix, the mathematical definition of the diffusion model is presented.

If the diffusion model is used to analyze RT data, it is important to note that the decision process constitutes only one component of the RTs measured. The other component consists of motor processes of the response system and of encoding processes preceding the decisional phase. These processes are not mapped by the diffusion process. Nonetheless, the model allows for estimating their total duration: The parameter t_0 indicates the duration of all extradecisional processes.

Finally, it might be necessary to allow for *inter*trial variability of the different parameters in the model. Ratcliff (e.g., Ratcliff & Rouder, 1998) uses a model with intertrial variability of the drift rate (η), the starting point (s_z), and the time constant of the motor response (s_t). The intertrial variability parameters of drift and starting point are important, because they allow for modeling of RT distributions in which error responses are slower or faster than correct responses (see Ratcliff & Rouder, 1998). Therefore, we included variability of drift and starting point. For the sake of simplicity, however, we did not allow for intertrial variance of the RT constant. Following Ratcliff, we assumed that the actual drift in one trial is normally distributed around the average drift (v), with the standard deviation η; the actual starting point is supposed to be equally distributed within the interval from $z - s_z$ to $z + s_z$.

Estimating the Parameters

The aim of the estimation of the parameters is to optimize the fit of the empirical and the predicted RT distributions for both types of responses simultaneously. For the optimization process, a criterion is needed that specifies the degree of correspondence between the theoretical and the empirical distributions (*goodness of fit*). Ratcliff and Tuerlinckx (2002) compared three different parameter estimation procedures (i.e., maximum likelihood, chi-square, and weighted least square). From our point of view, all these methods have specific problems. The maximum likelihood method seeks to maximize the predicted probability for each empirical RT. This algorithm might be especially vulnerable to the *contamination* of data (Ratcliff & Tuerlinckx, 2002). Especially, *outlier times* may strongly bias the parameter estimation; however, Ratcliff and Tuerlinckx are able to reduce this vulnerability by introducing a new parameter that controls for contamination. The chi-square and the weighted least-square methods are not based on single RTs, but on the number of times found in certain ranges on the time axis. These methods are quite coarse (because they are not based on individual responses) and have the disadvantage that the specification of the time segments is always somewhat arbitrary.[2]

An alternative to these three procedures for parameter estimation is presented in the following section. The opti-

Table 1
Parameters of the Diffusion Model

Parameter	Description
Upper threshold (a)	Response criterion (conservative vs. liberal)
Starting point (z)	Asymmetries in the response criterion
Drift (v)	Information uptake per time
Response time constant (t_0)	Nondecisional portion of the response time

Note—See text for further explanations.

mization criterion for the algorithm is provided by the KS test (Kolmogorov, 1941; see also Conover, 1999). This test is used to check whether an empirical distribution of values matches a given theoretical distribution. The test statistic (T) is the maximum vertical distance between the theoretical and the empirical cumulative distributions (see Figure 2). For a given value of T and a given number of instances (e.g., response times), the test provides the exact probability p that the empirical times fit to the theoretical distribution (Conover, 1999; Miller, 1956). A significant result (e.g., one below $\alpha = .05$) indicates that correspondence between both distributions is unlikely.

The statistic of the KS test (T) can be used as an optimization criterion for the parameter estimation procedure: Parameters of the model are varied in such a way that T (i.e., the maximum vertical distance between the empirical RT distribution and the predicted time distribution) becomes minimal. Because T is directly—and monotonically—related to the probability of match between the empirical and the predicted distributions, it allows for an identification of the optimal set of parameter values. It might also evade the disadvantages of the other methods mentioned above. First, in contrast to the maximum likelihood method, times with a predicted probability near—or even equal to—zero have no inflated influence. Second, the procedure is fine graded, because it is based on single RTs.

The application of the KS test for evaluating the model fit requires the comparison of *two* predicted distributions with *two* empirical distributions (i.e., the distributions at the upper and the lower thresholds), because the model is based on separate distributions for the RTs for the two alternative responses. To test the fit for both distributions simultaneously, we have to *associate* the two predicted distributions and the two empirical distributions, respectively. This can be done by *mirroring* the distribution of RTs for one type of response (e.g., the distribution at the lower threshold) at the zero point of the time dimension, so that all the times in this distribution receive a negative sign. Please note that, in most cases, one threshold is linked to correct responses and the other to erroneous responses. In this case, all error responses have to be mirrored (or vice versa, all correct responses). Composing a single RT distribution from the two RT distributions for the different response alternatives—with short times lying at the inner edges (near zero) and the *tails* with the longer times at the outer edges—allows for a single KS test.[3] Figure 3 illustrates this procedure. If there are different types of stimuli, different drift rates can be expected. It might be promising not to estimate completely independent models for each type of stimulus, but to determine that all parameters except for the drift take the same value in both models. In this case, the product of the probabilities of the different KS statistics—that is, the compound probability that the different models fit the data—are used as the criterion in the optimization procedure. The T statistic of the KS test is to be minimized in an iterative search process. To search for the minimum in the parameter space, the simplex method (Nelder & Mead, 1965) was used. In the Appendix, the mathematical definition of the diffusion model, as well as the formula for the KS test, is provided.

Testing the Model

The rationale for the appropriate test of model fit is based on the optimization criterion that has been described in the preceding section: Again, the KS statistic is used to evaluate the degree of match between the empirically observed RT distributions and the cumulative density function that is predicted on the basis of the parameter values of the diffusion model. However, in a typical experiment, there are different participants and/or experimental conditions for which different models are estimated. This implies that a KS test has to be conducted for each model (i.e., for participants and conditions, different KS tests have to be conducted). To get an
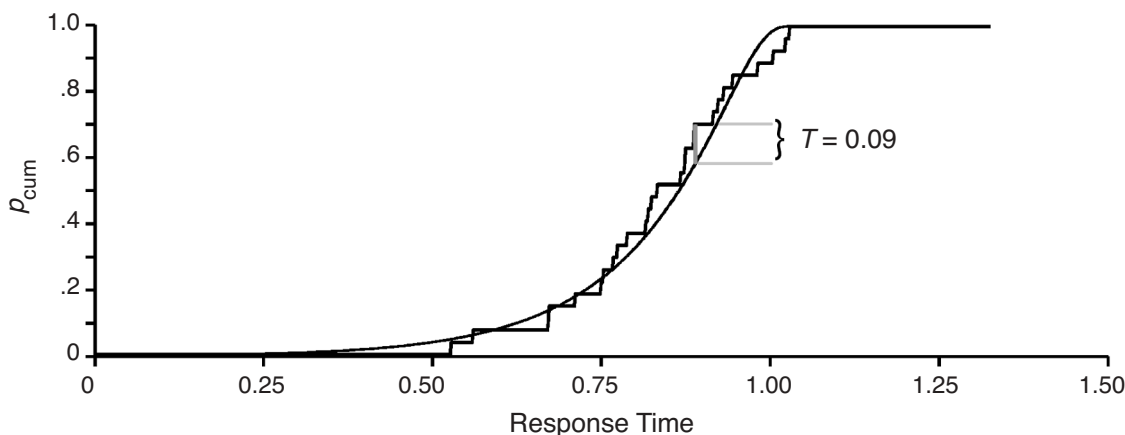

Figure 2. The statistic of the Kolmogorov–Smirnov test (T) is the maximum vertical distance between two cumulative distribution functions. Each stage in the stepped line represents one response at the time represented on the abscissa.
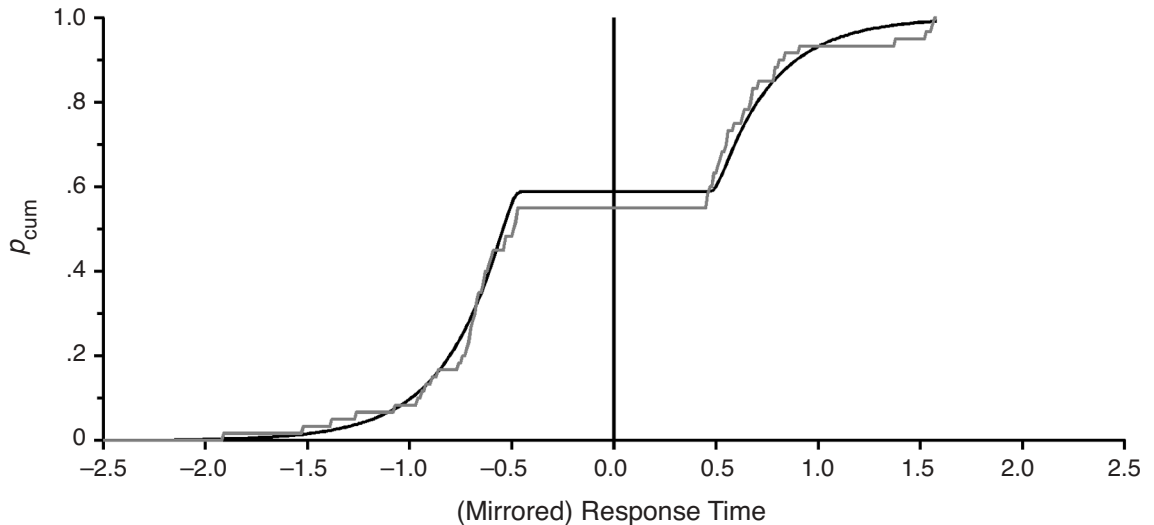
Figure 3. Comparison of empirical and predicted response time distributions. Times for one of the two responses (e.g., all left keypresses in a dual-response task) have been *mirrored* on the zero point of the time axis. The mirrored times are represented by the left part of the graph. For example, a response time of 0.850 is recoded as −0.850. The steep rise of the distribution functions at about −0.6 indicates that many responses occurred with a latency of 0.6 sec. The level of the function in the middle of the graph shows that about 60% of all the reactions have been of the mirrored type. The ascending black line is the accumulated probability function computed according to the diffusion model. The gray line shows the cumulative probability of empirical response times. Note that both cumulative functions must converge to 1.

estimate of the *overall* fit of the models with the data, the number of significant KS tests ($\alpha < .05$) can itself be subjected to a binomial test with parameters of n (number of tests) and p (α). A significant result of this overall binominal test indicates that the diffusion model in the tested form does not fit the data from the total sample. This method provides an exact test applicable to a whole sample of distributions, which we consider to be a valuable complement to graphical demonstrations of model fits (i.e., quantile probability functions; Ratcliff, 2001; Ratcliff & Tuerlinckx, 2002).

EXPERIMENT 1

In a first experiment, we manipulated the relative speed of information uptake, conservatism, and the duration of the motor response in a color discrimination task. These manipulations should specifically affect the drift rate (v), the distance between thresholds (a), and the motor component (t_0), respectively, in a diffusion model data analysis. If the predicted pattern emerges, this is strong support for the validity of the model.

Altogether, four different experimental conditions are reported. First, a standard condition, which provided baseline values for all the parameters, is presented. In a second, more difficult condition, the stimuli were harder to discriminate. Therefore, the information uptake should be reduced, which should result in a lower drift rate (v). In a third block, an accuracy motivation was induced. This manipulation should lead to more conservative response criteria and should increase the distance between thresholds (a). Finally, a more difficult form of response

had to be executed. In this condition, an increase in the value of the motor component (t_0) was predicted.

Method

Participants. Thirty-six students (27 females and 9 males) at the University of Trier participated in partial fulfillment of course requirements.

Design. Essentially, the design consisted of the within-subjects factor condition (standard, difficult, accuracy, and response handicap). The four blocks were presented in random order.

Apparatus and Stimuli. The experiment was implemented on an IBM-compatible Pentium computer connected to a 17-in. color monitor using a Turbo Pascal 7.0 program in graphics mode. As stimuli, we used colored squares (10×10 cm) consisting of 250×200 screen pixels. Each pixel was either orange (RGB color values: 63, 25, 0) or light blue (0, 25, 63). The proportion of colors within each stimulus was 53:47. In the *difficult* condition, the ratio was 51.5:48.5. Pixels were randomly intermixed. For half of the stimuli, orange was the dominant color; for the other half of the stimuli, blue was dominant. The stimuli were presented on a dark background.

Procedure. The participants started with a training block of 42 trials. Thereafter, the four experimental blocks followed in random order. Each block consisted of 2 warm-up trials and 40 experimental trials. Each trial had the following sequence. First, a gray square of the same size as the stimuli was presented. This square was replaced by the colored stimulus 750 msec later. The stimulus remained on the screen until a response was registered or the time limit of 1,500 msec (3,000 msec in the accuracy condition) was reached, whichever happened first. The remaining time was visually represented by a vertical time bar that decreased in length below the stimulus. The participant's task was to decide which of the two colors was dominant—that is, which color composed the major part of the stimulus. Two marked keys on the computer keyboard ("C" and "M") were used for the responses. The assignment of keys to colors was counterbalanced across participants. Directly after the response, the word "correct" or "wrong" was presented below the color field,

depending on whether or not the correct answer had been given. Simultaneously, the stimulus was replaced by a monochromatic field showing the formerly dominating color. Feedback was removed from the screen 1,000 msec later, and a new trial started.

Experimental manipulations. As was outlined above, the experiment consisted of a standard condition and three variations. In the *difficult* condition, the dominant color occupied only 51.5% (instead of 53% in the standard condition) of the pixels of a stimulus. Second, there was an accuracy condition, in which the RT limit was doubled (3 sec instead of 1.5 sec). In addition, the participants were instructed to work especially carefully and to avoid mistakes in this block. The third variation refers to the form of response: In the response handicap condition, the participants were told to use the same finger for both responses. To ensure that the finger was positioned between both response keys before the stimulus was shown, the participants had to press a key that lay between the two response keys (i.e., the "B" key) to start each trial.

Results

Parameter estimation. For each participant, four independent models were calculated, one for each experimental condition. Each of these models allowed for two different drift rates (for the two stimulus types) and comprised two different diffusion models (see the Estimating the Parameters section).

Therefore, there were $36 \times 4 = 144$ independent models, based on 40 trials each. For the interpretation of parameters, average values were computed across participants. All responses below 300 msec were discarded prior to parameter estimation (1% of the trials). This is important because low outliers may have a strong influence on the complete model (see Ratcliff & Tuerlinckx, 2002, for a discussion of the influence of outliers).

For all blocks of the experiment, a model was chosen in which the upper threshold was associated with the *orange* response (i.e., the press of the *orange* key), and the lower threshold was associated with the *blue* response. For the two types of stimuli (orange dominant vs. blue dominant), two different drift rates were estimated. Strictly speaking, our model is composed of two diffusion models, with the constraint that all the parameters except drift rate (a, z, t_0) are identical for both models. Please note that these two diffusion models were always estimated simultaneously in a single estimation process. Together, five parameters were estimated: a, z, v_1, v_2, and t_0. Table 2 shows the mean parameter values across participants for all blocks in the experiment. To facilitate the interpretation, z/a is presented as a measure of bias, instead of z. Values greater than 0.5 on z/a indicate that the starting point lies closer to the upper threshold (a), whereas lower values indicate a starting point closer to the lower threshold.

Within-block analyses. Table 2 shows that the processes always started roughly at the midpoint between both thresholds: One-group t tests revealed that only in the response handicap condition did z diverge significantly from $a/2$ [$t(35) = -2.39, p < .05$].

The v_1 parameter is the drift rate for orange-dominated stimuli. In all blocks, v_1 had a positive value [all ts(35) > 3.84, $p < .001$], indicating a positive mean gradient for

Table 2
Average Parameter Values for Different Experimental Conditions of Experiment 1

Parameter	Condition			
	Standard	Difficult	Accuracy	Response Handicap
a	1.18	1.19	1.37**	1.16
z/a	0.49	0.47	0.52	0.56*
v_1	0.89	0.46*	0.91	1.02
v_2	−1.03	−0.57**	−1.18	−1.41*
t_0	0.47	0.47	0.52**	0.73**
η	0.80	0.90	0.85	0.89
s_z	0.09	0.09	0.09	0.08

Note—The parameters describe diffusion models with the *orange* response linked to the upper threshold and the *blue* response linked to the lower threshold. The parameter v_1 represents the drift for orange-dominated stimuli, and v_2 represents the drift for blue-dominated stimuli. All tests refer to a comparison with the standard condition.   *$p < .05$.   **$p < .01$.

the process. Contrarily, the drift in trials with blue-dominated stimuli (v_2) was negative in all conditions [all ts(35) < −5.35, $p < .001$]. To check for symmetry, the *absolute* values of the drift parameters were compared between orange-dominated and blue-dominated stimuli in four paired t tests. These analyses revealed a significantly larger drift rate for blue-dominated stimuli in the response handicap condition [$t(35) = -2.09, p < .05$]. No other comparison reached significance [all ts(35) < 1.32, $p > .19$].

Between-block analyses. To test the influence of the experimental manipulations, all the parameters in the three *variation blocks* were contrasted with the parameters in the standard condition. All significant effects are marked in Table 2. The absolute values of the drift parameters in the *difficult* condition were lower than those in the standard condition, both for v_1 [$t(35) = 2.44, p < .05$] and for v_2 [$t(35) = -3.19, p < .01$]. The distance between thresholds (a) was larger in the accuracy condition than in the standard condition [$t(35) = -4.63, p < .001$]. In addition, we found an increased value for t_0 in this condition [$t(35) = -3.71, p < .01$]. Finally, in the response handicap condition, the RT constant (t_0) significantly exceeded the value in the standard condition $t(35) = -8.27$, $p < .001$. In this condition, the relative starting point (z_r) also was somewhat increased [$t(35) = -2.39, p < .05$], and the drift parameter v_2 was decreased [$t(35) = -2.32$, $p < .05$]. No other contrasts reached significance.

Model fit. As was outlined above, we performed individual tests of fit separately for each diffusion model (i.e., for each participant and for each condition). There were 2 (stimulus types) × 4 (condition) models for each of the 36 participants, resulting in a total of 288 models. The KS tests revealed no significant result at the .05 level. It can be concluded that the individual models describe the individual RT distributions well.

To demonstrate model fit in a more descriptive way, empirical and predicted values of the medians of the RT distributions for correct and erroneous responses and of
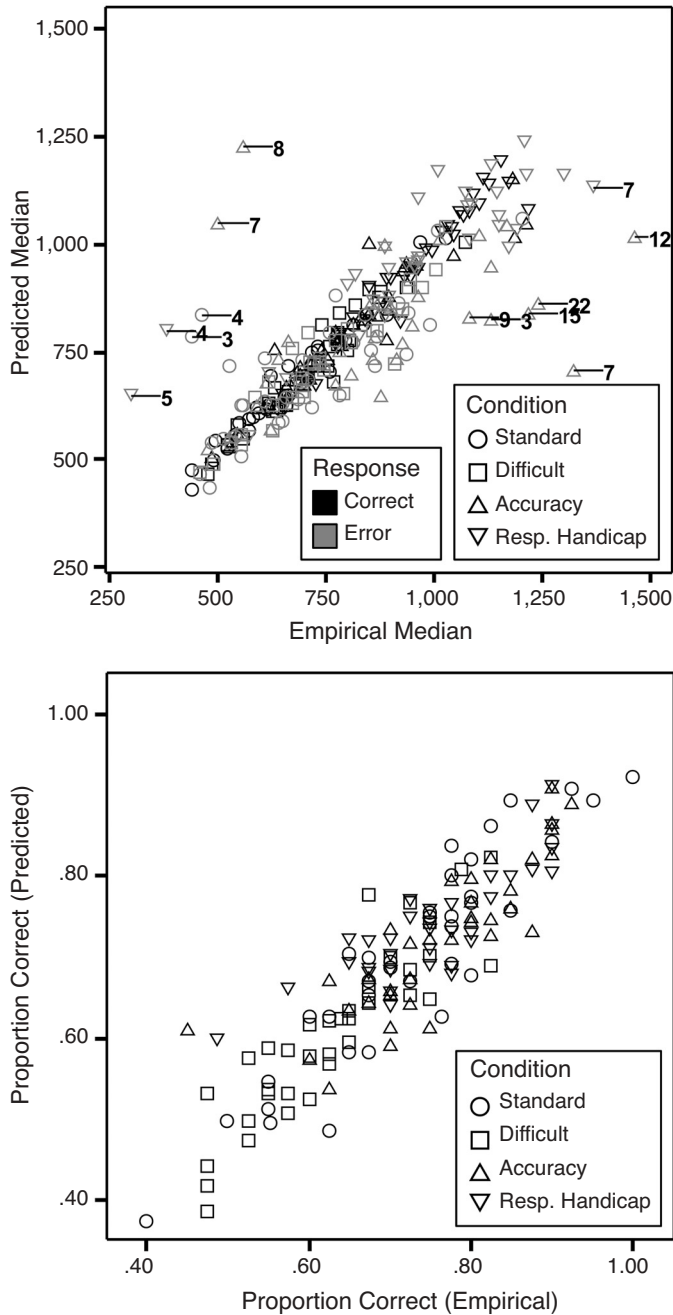
Figure 4. Individual model fit (Experiment 1). The top panel shows the concordance of the medians of the empirical and the predicted response time (RT) distributions. For each participant and experimental condition, the distributions of correct responses and errors are represented by single symbols. In case of large misfit, the number of RTs building the empirical distribution is also shown. The bottom panel shows the concordance of the proportions of correct responses.

proportions of correct responses for all the conditions in Experiment 1 are presented in Figure 4. As can be seen from the high concordance of the empirical and the predicted values, there is a good fit for most of the participants. However, there are some points with rather large

deviations of predicted and empirical medians. These large deviations are found only for distributions of error responses, and especially where very few errors were made (in most cases, fewer than 10). This indicates that reliable predictions require a somewhat higher number

of RTs within each empirical distribution. This is also illustrated by the fact that there is a significant negative correlation of the absolute deviations between the medians of the empirical and the predicted RT distributions and the number of RTs making up the empirical distribution ($r = -.42$, $p < .001$].

Discussion

The results provide evidence that the diffusion model adequately represents the different process components that are involved in speeded binary color dominance decisions. The within-block analyses showed that the decision process is roughly symmetric: The process starts at the midpoint between thresholds and is driven upward by orange-dominated stimuli or downward by blue-dominated stimuli. More interesting are the between-block comparisons. The different conditions had been implemented to investigate the validity of the interpretation of the parameters. It should be demonstrated that each parameter reflects specific components of the decision process and is not influenced by other processes. Only if the validity of the parameters is demonstrated can diffusion model results be interpreted clearly. In detail, we found decreased drift rates when we made stimuli harder to discriminate, an increased distance between thresholds when we induced an accuracy motivation, and a substantial increase in the motor component when we introduced a more time-consuming form of response. Some unexpected effects emerged as well: There was an increase of t_0 in the accuracy condition. This finding can be easily accommodated, because it is plausible that participants execute their responses more slowly if time pressure is reduced, which was the case in this condition. The changes in drift rate and starting point in the *difficult* condition, however, cannot easily be explained.

The KS tests revealed a good model fit for all the models (i.e., for all participants and all conditions). This is an additional clue that the diffusion model was appropriate for describing the data without losing much information. Descriptive data on the model fit showed that fit was worst for the error distributions of the accuracy condition. This was probably due to the low numbers of errors in this condition.

EXPERIMENT 2

In Experiment 1, the parameter z, which reflects the starting point of the diffusion process, was not manipulated. Therefore, Experiment 2 focused on the empirical validation of the interpretation of the z parameter. Again, a color discrimination task was used. This time, however, the two responses were not equivalent anymore: The participants collected points during the experiment, and an asymmetric payoff matrix promoted one response over the other. We expected that this manipulation would bias the starting point of the diffusion process toward the threshold that was linked with the *rewarded* response— that is, less information would be needed to reach the corresponding threshold, as compared with the opposite

Table 3
Payoff Matrix Used in Experiment 2

Response	Stimulus Color		Σ (Response)
	Color 1	Color 2	
Color 1	10	-5	$+5$
Color 2	-10	5	-5
Σ (stimulus)	0	0	

Note—Selection of the promoted response (Color 1) was counterbalanced across participants.

threshold. As a second variation, the stimuli were presented in two different color ratios.

Method

Participants. Twenty-four students (17 females and 7 males) at the University of Trier participated in partial fulfillment of course requirements.

Design. One response was fostered by an asymmetric payoff matrix. Which of the two responses was selected as the privileged response (*orange* vs. *blue*) was counterbalanced across participants.

Apparatus and Stimuli. The procedure was the same as that in the previous experiment, with the following exception: The stimuli were presented in two different color ratios, 52:48 and 51:49.[4]

Procedure. The experiment consisted of 12 training trials and 160 experimental trials, which were presented without intermission in one block; each of the four stimulus types (orange dominant vs. blue dominant × easy vs. difficult) was presented 40 times. Trials were presented in random order. The participants received points for correct answers and lost points for errors. These point credits were exchanged into monetary payoffs later. For each participant, one *color response* had a higher expected value because of an asymmetric payoff matrix (Table 3): If the promoted response was given correctly, 10 points were credited; if it was given mistakenly, however, only 5 points were lost. Opposite rules held for the nonpromoted response: Only 5 points were gained when the color response was given correctly, but 10 points were subtracted when an error occurred. These rules were explained carefully in the instructions. During the experiment, the current amount of points was visible at the bottom of the screen. In addition, gains of 5 points were signaled by one high-frequency beep, and gains of 10 points by two high-frequency beeps. Accordingly, losses of 5 or 10 points were accompanied by one or two low-frequency beeps, respectively. At the end of the experiment, the participants received 20 Euro Cent (about $0.18 at that time) for each 100 points they reached. The procedure resulted in a mean benefit of 1.28 Euro (about $1.15).

Results

Parameter estimation. RTs below 300 msec (0.3% of all the times) were excluded from all analyses. In Experiment 2, a model with four drift parameters, one for each stimulus type, was estimated (i.e., this model consisted of four *connected* diffusion models; see the Estimating the Parameters section for further explanations). The parameters v_1 to v_4 described the drift rate for a descending proportion of orange pixels (52%, 51%, 49%, and 48%; see the Apparatus and Stimuli section). Furthermore, there were the same parameters as above: a, z, t_0, η, and s_z, resulting in a total of nine free parameters for each model. For each participant, one model was computed on the basis of all 160 experimental trials. The average parameter values for both groups of participants are shown in Table 4. We expected that the manipulation

should bias the *relative* starting point of the process—that is, the position of z in relation to a. Therefore, the relative starting point (z/a) is presented in Table 4. A value of .50 for z/a indicates that the process starts at the midpoint between both thresholds; larger values indicate

that the starting point is closer to the *orange* threshold than to the *blue* threshold, and vice versa.

Statistical analyses. All the parameters were tested for between-group differences (see Table 4). A significant difference emerged only for the relative starting
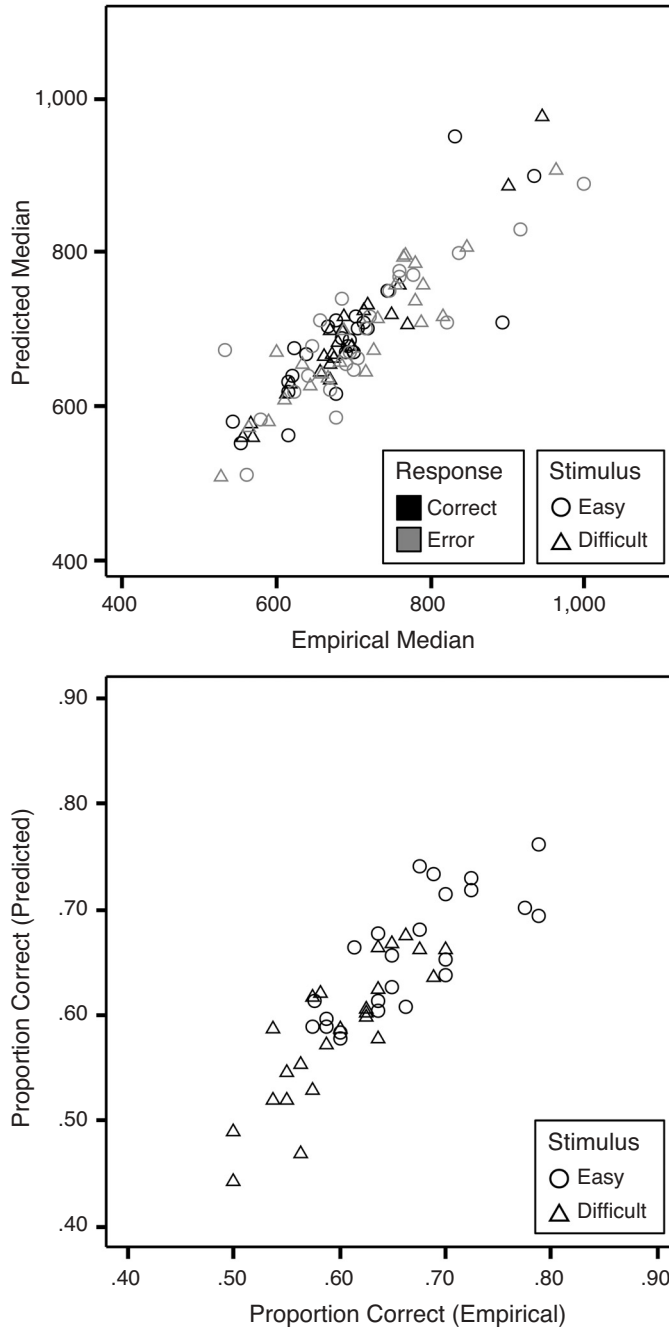

Figure 5. Individual model fit (Experiment 2). The top panel shows the concordance of the medians of the empirical and the predicted response time distributions. For each participant and each of the two difficulty levels, the distributions of correct responses and errors are represented by single symbols. The bottom panel shows the concordance of the proportions of correct responses.

point z/a [$t(22) = 4.08$, $p < .01$]. The effect reveals that the starting point was biased toward the *promoted* response—that is, the participants were more liberal in giving the promoted answer and more conservative regarding the nonpromoted answer. In the between-group comparison, no other effects reached significance.

The drift parameters were analyzed additionally in a 2 (group) × 4 (color ratio) analysis of variance. There was a strong effect of color ratio [$F(3,20) = 89.55$, $p < .001$], indicating larger (i.e., more positive) drift rates for higher proportions of orange in the stimuli. No other effects reached significance.

Model fit. In this experiment, 4 (stimulus types) × 24 (participants) = 96 different probability distributions were predicted by the models. Six of the KS tests detected a significant ($\alpha < .05$) deviation of the empirical from the predicted distribution. The probability of finding six or more mismatches with $\alpha = .05$ in 96 tests was $p = .35$, indicating that the models describe the empirical RT distributions well. Again, model fit should also be demonstrated descriptively. The coherence model predictions and the empirical distributions can be seen on an individual level in Figure 5: Again, the medians of the empirical RT distributions are plotted against the medians of the predicted distributions, both for errors and for correct responses.

Discussion

The second experiment was carried out to validate the interpretation of the bias parameter (z/a) in the diffusion model. Specifically, we investigated whether biasing response tendencies in a binary decision task by means of an asymmetric payoff manipulation is mapped specifically onto the bias parameter in a diffusion model data analysis. The parameter z indicates the starting point of the diffusion process and, thus, defines—in combination with the total distance between thresholds—the necessary amount of information that is needed to draw a specific conclusion. In an unbiased model, z will equal $a/2$—

that is, the same amount of information is required for each of both decisions. The closer z moves to one of the boundaries, the more biased is the process. In this case, the preferred response is initiated quickly and often.

In Experiment 2, we successfully manipulated z by "rewarding" the participants for executing one of the two responses. The starting point moved toward the threshold that was connected with the rewarded answer. No other parameter was affected by this manipulation. As revealed by the KS tests, model fit was good.

GENERAL DISCUSSION

Diffusion models were introduced in psychology by Ratcliff (1978) two and a half decades ago. These models describe a theoretical process of decision making. Interestingly, recent neurobiological studies based on single-cell recording have shown activation patterns prior to action and decision making that resemble the path of the diffusion process (see Schall, 2001, for an overview).

Only the recent developments in computing power allow for efficient parameter estimation. Our procedure differs from the methods employed by Ratcliff in a couple of ways: (1) We use distributions of single RTs in the estimation process, instead of using only certain percentiles of the RT distribution. (2) We introduced a new optimization criterion to estimate the parameters that was deduced from the KS test (Kolmogorov, 1941; see also Conover, 1999). (3) The KS test is also used as a test of the model fit. This method allows a clear-cut statistical test of whether or not a model fits the data.

The estimation procedure presented in this article has been developed on the basis of theoretical considerations. Because the KS statistic provides an exact test for the match of an empirical and a predicted distribution, this measure for the *probability of match* seems to be an optimal criterion for parameter estimation. So far, no empirical comparisons with other estimation procedures have been made (e.g., concerning the influence of contaminants; see Ratcliff & Tuerlinckx, 2002), although this is clearly an interesting challenge for the procedure. Another important question for the application of the model is how reliably the different procedures (i.e., KS, maximum likelihood, and chi-square) can estimate the parameters in cases in which there are the relatively small sets of data (i.e., few RTs per participant) typically used in cognitive research. It is plausible that procedures based on single RTs (such as the KS-based procedure or the maximum likelihood procedure) may be better at this, because they use the total amount of information. To empirically answer such questions, more simulation studies have to be conducted.

The diffusion model provides plausible interpretations for the estimated parameters. The distance between thresholds (a) is interpreted as a measure of *conservatism*—that is, the larger a is, the more information is collected prior to a response being given. The starting point (z) is an estimate of *relative* conservatism, or *bias*:

Table 4
Average Values of the Diffusion Model Parameters for Both Groups of Experiment 2

| Parameter | Promoted Response | | |
	Orange	Blue	Difference
a	1.13	1.19	−0.06
z/a	0.54	0.44	0.10**
v_1	0.70	0.69	0.01
v_2	0.46	0.03	0.42
v_3	−0.47	−0.69	0.22
v_4	−0.70	−1.02	0.32
t_0	0.44	0.49	−0.05
η	1.02	1.11	−0.09
s_z	0.13	0.12	−0.01

Note—The parameters describe diffusion models with the *orange* response linked to the upper threshold and the *blue* response linked to the lower threshold. The drift parameters v_1 to v_4 are linked to the four different stimulus types with a descending proportion of orange pixels. In addition to the estimated parameter, the relative starting point (z/a) is shown (see the text for further details).   **$p < .01$.

If the process starts above or below the midpoint between the two thresholds, different amounts of information are required for both responses; that is, a more conservative decision criterion is applied for one response, and a more liberal criterion for the opposite response. The drift rate (v) indicates the (relative) amount of information gathered per time, denoting either perceptual sensitivity or task difficulty. The parameter t_0, finally, provides a measure of the duration of nondecisional components of the response process—that is, first of all, of the motor processes.

A possible criticism of the diffusion model is that it uses a fairly large number of parameters to model RT distributions. Given that it is always easy to fit any possible data set, provided that enough independent parameters are available, this leads to the question of whether unambiguous interpretations of the parameters of the model are warranted. In a recent set of simulation studies, Ratcliff (2002; Ratcliff & Tuerlinckx, 2002) demonstrated that diffusion models impose a number of important restrictions on data and that violations of these restrictions can be detected by a low fit of the model. However, this still leaves open the question of whether the parameters of the model yield unique estimates of distinct components of real decision processes. For example, the model might be insensitive in disentangling different components of the decision process and might show a strong tendency to represent different components of the process on only one parameter. Alternatively, the model might have difficulty in uniquely identifying a certain component of the decision process with only one parameter: Variations in different parameters of the model might be equifunctional in representing differences in the same component of the decision process, so that only slight variations in the data set can lead to dramatic changes of the parameter estimates. Ratcliff and Tuerlinckx have shown that parameters can be reproduced reliably from a set of simulated data, even if *contaminated* times are added after the simulation. However, this does not demonstrate whether the different parameters represent specific *cognitive* processes. An evaluation of the *empirical* validity of the parameter estimates cannot be decided in theory but requires empirical testing. The experiments reported in this article are a step in this direction. We used simple, straightforward experimental manipulations that should uniquely influence certain components of the decision process. The face validity of the manipulations is of utmost importance, because it is a prerequisite for taking the observed effects as a test of the model's validity, rather than as a test of what was being manipulated. In contrast to other studies based on similar paradigms (e.g., Ratcliff et al., 2001, 2003), in the present article systematic manipulations within one paradigm have been used to check for cognitive processes related to the different model parameters in the absence of any other research question.

Two experiments were carried out to test whether simple experimental manipulations aimed at influencing specific components of the decision process in a straightforward way map onto corresponding parameters of the model while leaving others unaffected. In Experiment 1, stimuli that were difficult to categorize produced smaller drift rates than did *easy* stimuli, an induction of accuracy motivation increased the conservatism parameter, and a response handicap did not affect the decision-related parameters but led to increments in the motor component. In Experiment 2, an asymmetric payoff matrix promoting one of the two responses moved the starting point toward the threshold that was linked to the promoted response. In sum, our findings clearly support the validity of the previous interpretations of the parameters of the diffusion model.[5]

CONCLUSION

RT data from binary decision tasks are a major source of information in experimental psychological research. The diffusion model (Ratcliff, 1978) offers a powerful approach for analyzing these data. The separation and identification of different components of the decision process allow fine-grained interpretations of the findings that also provide interesting information for theoretical debates regarding the processes that underlie or mediate experimental effects. To give just a few examples, a diffusion model data analysis might help to clarify the nature of the processes involved in cognitive aging (Ratcliff et al., 2000; Ratcliff et al., 2001, 2003), of the underlying processes in associative or affective priming (Wentura, 2000), or of what drives effects in the IAT (Brendl, Markman, & Messner, 2001; Rothermund & Wentura, 2001). Another advantage of this form of data analysis is that it is able to integrate a large amount of the available information in only one analysis (e.g., RTs for correct responses and errors and percentages of errors). These examples reveal that the diffusion model is a highly promising statistical tool with a very wide range of potential applications. In the present article, we have reported some new strategies that further improve the application of the model (new techniques of parameter estimation and model testing). The main focus of the article, however, was two validation experiments that yielded further evidence in support of the model's capacity to identify hidden cognitive processes in binary decision tasks.

REFERENCES

BRENDL, C. M., MARKMAN, A. B., & MESSNER, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality & Social Psychology*, **81**, 760-773.

CONOVER, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

FELLER, W. (1971). *An introduction to probability theory and its applications*. (Vol. 2). New York: Wiley.

KOLMOGOROV, A. (1941). Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics*, **12**, 461-463.

MILLER, L. H. (1956). Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association*, **51**, 111-121.

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

RATCLIFF, R. (2001). Diffusion and random walk processes. In *International encyclopedia of the social and behavioral sciences* (Vol. 6, pp. 3668-3673). Oxford: Elsevier.

RATCLIFF, R. (2002). 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. *Psychonomic Bulletin & Review*, **9**, 278-291.

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

RATCLIFF, R., SPIELER, D., & MCKOON, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review*, **7**, 1-25.

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., THAPAR, A., & MCKOON, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, **65**, 523-535.

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

RATCLIFF, R., VAN ZANDT, T., & MCKOON, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261-300.

ROTHERMUND, K., & WENTURA, D. (2001). Figure–ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie*, **48**, 94-106.

SCHALL, J. D. (2001). Neural basis of deciding, choosing, and acting. *Nature Reviews Neuroscience*, **2**, 33-42.

WENTURA, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: Yes versus no responses to word targets reveal evaluative judgment tendencies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 456-469.

NOTES

1. Whether the model adequately represents the underlying cognitive processes in a task remains an open question even when the model fits the data well, because there are always infinitely many theoretically distinct models that fit the data equally well.

2. The number (and the width) of the time segments may influence the result of the estimation process: (Slightly) different parameter values may, for example, emerge in estimation processes based on five or six segments. It is hard to pronounce upon what would be an optimal number of segments: If few large segments are chosen, much information is lost (concerning the exact shape of the distribution); if many small segments are chosen, the number of empirical RTs in each is largely chance dependent.

3. We thank Heinrich von Weizsäcker for this ingenious hint.

4. A third stimulus condition with color ratios of 53:47 was also included in the experiment. Trials in this condition were not entered into the analyses, because model fit was not especially good for this condition. The low model fit is probably due to the extremely low error rates in this condition (median of the number of errors = 8). Nonetheless, the results regarding the starting point of the diffusion process do not change substantially if this condition is included in the analysis.

5. With only two data points for each manipulation, it is not possible to describe the exact functions relating the parameter values to physical properties of the stimuli. However, it is clear that monotonic relationships will emerge. (Ratcliff & Rouder, 1998, point out a probability-matching explanation of the drift rate.)

APPENDIX
Mathematical Details

Mathematical Definition of the Diffusion Model

The probability that a diffusion process with constant drift v and start at $z > 0$ will hit $a > z$ before time t and before the first visit at 0 has a density, which we shall denote by $g_+(t, z, a, v)$. We have two different representations for this density (all formulas are derived from Feller, 1971):

$$g_+(t,z,a,v) = \frac{\exp\left[(a-z)v - 0.5v^2 t\right]}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} \exp\left(-\frac{\left[(1+2n)a - z\right]^2}{2t}\right) \cdot \left[(1+2n)a - z\right] \tag{A1}$$

and

$$g_+(t,z,a,v) = \frac{\pi}{a^2} \cdot \exp\left[(a-z)v\right] \sum_{n=1}^{\infty} n \cdot \sin\left[\frac{\pi(a-z)n}{a}\right] \cdot \exp\left[-0.5 \cdot \left(v^2 + \pi^2 n^2/a^2\right) \cdot t\right]. \tag{A2}$$

By symmetry, we get the corresponding densities for the probability that a diffusion process with constant drift v and start at $z > 0$ will hit 0 before time t and before the first visit at $a > z$ as $g_-(t, z, a, v) = g_+(t, a-z, a, -v)$. This gives

$$g_-(t,z,a,v) = \frac{\exp\left(-zv - 0.5v^2 t\right)}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} \exp\left[-\frac{\left(2na + z\right)^2}{2t}\right] \cdot \left(2na + z\right) \tag{A3}$$

and

$$g_-(t,z,a,v) = \frac{\pi}{a^2} \cdot \exp\left(-zv\right) \sum_{n=1}^{\infty} n \cdot \sin\left(\frac{\pi z n}{a}\right) \cdot \exp\left[-0.5 \cdot \left(v^2 + \pi^2 n^2/a^2\right) \cdot t\right]. \tag{A4}$$

The case of a diffusion constant s other than 1 can be reduced to the case above: If we scale down everything by the factor s, we get an ordinary diffusion with diffusion constant 1, with constant drift v/s, start at z/s, and the upper threshold a/s. So the probability that a diffusion process with diffusion constant s, drift v, and start at $z > 0$ will hit $a > z$ before time t and before the first visit at 0 has the density $g_+(t, z/s, a/s, v/s)$.

We need to evaluate these series numerically. The speed of convergence in these series is determined by the exponential terms in the sum. So the series in Formula A1 will converge quickly when t is small, in comparison with $2a^2$, and the series in Formula A2 will converge quickly when t is large, in comparison with $\pi^2/2a^2$. An advantage of Formula A2 is that we can calculate the distribution function: The probability that a diffusion process with constant drift v and start at z will hit $a > z$ before time t and before the first visit at 0 is

$$\int_0^t g_+(r,z,a,v)dr$$

$$= \frac{\pi \cdot \exp\left[(a-z)v\right]}{a^2} \sum_{n=1}^{\infty} n \cdot \sin\left[\frac{\pi(a-z)n}{a}\right] \cdot \int_0^t \exp\left[-0.5 \cdot \left(v^2 + \pi^2 n^2/a^2\right) \cdot r\right]dr$$

$$= \frac{\pi \cdot \exp\left[(a-z)v\right]}{a^2} \sum_{n=1}^{\infty} n \cdot \sin\left[\frac{\pi(a-z)n}{a}\right] \cdot \frac{2\left\{1 - \exp\left[-0.5 \cdot \left(v^2 + \pi^2 n^2/a^2\right) \cdot t\right]\right\}}{v^2 + \pi^2 n^2/a^2}$$

$$= 2\pi \cdot \exp\left[(a-z)v\right] \sum_{n=1}^{\infty} n \cdot \sin\left[\frac{\pi(a-z)n}{a}\right] \cdot \frac{1 - \exp\left[-0.5 \cdot \left(v^2 + \pi^2 n^2/a^2\right) \cdot t\right]}{v^2 a^2 + \pi^2 n^2}. \tag{A5}$$

For our parameter estimation, we use Formula A5. Note that t is not the RT but the duration of the decision process. Therefore, $t = rt - t_0$ has to be substituted in the equation.

Necessary Number of Steps in the Calculation of the Infinite Series of Formula A5

Here, we derive an upper bound for the difference between the exact value of the infinite sum from Formula A5 and the finite sum, which we will calculate using our program. We can then choose the required number of steps, in order to keep the error below a given bound ε.

1. For the calculation below, we will need a bound for

$$\sum_{k=0}^{n} \sin(ck)$$

that does not depend on n. Since

$$\sum_{k=0}^{n} \sin(ck) = \text{Im}\left(\sum_{k=0}^{n} e^{ick}\right) = \text{Im}\left(\frac{1 - e^{ic(n+1)}}{1 - e^{ic}}\right),$$

we will use

$$S(c) = \sup\left\{\left|\text{Im}\frac{1 - e^{ix}}{1 - e^{ic}}\right| \,\bigg|\, x \in \mathbb{R}\right\} \tag{A6}$$

for this bound. Because the set

$$\left\{1 - e^{ix} \big| x \in \mathbb{R}\right\}$$

is a circle in the complex plane, the possible values of the fraction in Formula A6 still form a (scaled and translated) circle in the plane.

Using $a = \text{Re}(1 - e^{ic}) = 1 - \cos(c)$, $b = \text{Im}(1 - e^{ic}) = -\sin(c)$, and $r = |1 - e^{ic}| = \sqrt{a^2 = b^2} = \sqrt{2 - 2\cos(c)}$, the center of this circle has the real part a/r^2 and the imaginary part $-b/r^2$. The radius of the circle is l/r, so we conclude that

$$S(c) = \left|\frac{b}{r^2}\right| + \frac{1}{r} = \frac{|b| + r}{r^2} = \frac{|\sin(c)| + \sqrt{2 - 2\cos(c)}}{2 - 2\cos(c)}. \tag{A7}$$

2. The next step is to obtain a general result about the tails of oscillating series. Let a_n be a sequence monotonically decreasing to 0. Analogous to partial integration, we find

$$\sum_{a=N}^{M} \left(a_n - a_{n+1}\right) \sum_{k=N}^{n} \sin(ck) = \sum_{a=N}^{M} a_n \cdot \sin(cn) - a_{M+1} \sum_{k=N}^{M} \sin(ck)$$

and, thus,

$$\left|\sum_{n=N}^{M} a_n \sin(cn)\right| \leq \sum_{n=N}^{M} \left(a_n - a_{n+1}\right) \cdot \left|\sum_{k=N}^{n} \sin(ck)\right| + a_{M+1} \cdot \left|\sum_{k=N}^{M} \sin(ck)\right|$$

$$\leq \left(a_N - a_{M+1}\right)S(c) + a_{M+1}S(c).$$

Letting $M \to \infty$ proves

$$\left| \sum_{n=N}^{\infty} a_n \sin(cn) \right| \le a_N S(c).$$

This helps to estimate the tail of the series, which is omitted from the calculation.

3. In Formula A5, we have

$$c = \frac{\pi(a-z)}{a}$$

and

$$a_n = 2\pi e^{(a-z)v} n \frac{1 - \exp\left(-\dfrac{a^2 v^2 + \pi^2 n^2}{2a^2} t \right)}{a^2 v^2 + \pi^2 n^2}$$

$$= 2\pi e^{(a-z)v} \frac{1 - \exp\left(-\dfrac{a^2 v^2 + \pi^2 n^2}{2a^2} t \right)}{a^2 v^2 / n + \pi^2 n}.$$

From c, we can calculate $S(c)$ with Formula A7 and then use the following estimate:

$$\left| \int_0^t g_+(r,z,a,v)dr - \sum_{n=1}^{N} a_n \sin(cn) \right| = \left| \sum_{n=1}^{\infty} a_n \sin(cn) - \sum_{n=1}^{N} a_n \sin(cn) \right|$$

$$= \left| \sum_{n=N+1}^{\infty} a_n \sin(cn) \right|$$

$$\le a_{N+1} S(c). \qquad (A8)$$

Formula A8 gives an upper bound for the difference between the theoretical value from Formula A5 and the value that our program calculates (summing up to N). To achieve a prescribed precision ε, we simply include enough terms in the sum to make $a_{n+1}S(c)$ smaller than ε.

4. A simplification is possible, because for large n the exponential term and the term $a^2v^2/2$ in the definition of a_n make almost no difference. So we get

$$a_n \le 2\pi e^{(a-z)v} \frac{1}{\pi^2 n} = \frac{2 e^{(a-z)v}}{\pi} \frac{1}{n}.$$

If we want to estimate the integral up to accuracy ε, we can use

$$N = \left[\frac{2 e^{(a-z)v} S(c)}{\pi \varepsilon} \right] \qquad (A9)$$

to get

$$a_{N+1} S(c) \le \frac{2 e^{(a-z)v}}{\pi} \frac{1}{N+1} S(c) \le \frac{2 e^{(a-z)v}}{\pi} \frac{\pi \varepsilon}{2 e^{(a-z)v} S(c)} S(c) = \varepsilon.$$

Formula A9 gives the required number of steps to keep the error smaller than ε.

5. *Examples*. For the results from Experiment 1 (see Table 2), we get the following required numbers of steps.

a	z	v	ε	N
1.18	$0.49 \cdot a$	0.89	0.01	108
1.18	$0.49 \cdot a$	-1.03	0.01	34
1.19	$0.47 \cdot a$	0.46	0.01	85
1.19	$0.47 \cdot a$	-0.57	0.01	44
1.37	$0.52 \cdot a$	0.91	0.01	115
1.37	$0.52 \cdot a$	-1.18	0.01	29
1.16	$0.56 \cdot a$	1.02	0.01	107
1.16	$0.56 \cdot a$	-1.41	0.01	30

APPENDIX (Continued)

The Probability of the KS Statistic

A formula for the probability of a given value of the KS statistic, depending on the number of observations, is provided by Conover (1999):

$$p = 2T \sum_{j=1}^{[n(1-T)]} \binom{n}{j} \left(1 - t - \frac{j}{n}\right)^{n-j} \left(t + \frac{j}{n}\right)^{j-1},$$ (A10)

where T is the observed KS statistic, n is the number of observations, and $[n(1-T)]$ is the maximum integer less than or equal to $n(1-T)$. Note that Formula A10 is valid for the two-sided test.