Theoretical Article

Diffusion Models in Experimental Psychology

A Practical Introduction

Andreas Voss, Markus Nagler, and Veronika Lerche

Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Germany

Abstract. Stochastic diffusion models (Ratcliff, 1978) can be used to analyze response time data from binary decision tasks. They provide detailed information about cognitive processes underlying the performance in such tasks. Most importantly, different parameters are estimated from the response time distributions of correct responses and errors that map (1) the speed of information uptake, (2) the amount of information used to make a decision, (3) possible decision biases, and (4) the duration of nondecisional processes. Although this kind of model can be applied to many experimental paradigms and provides much more insight than the analysis of mean response times can, it is still rarely used in cognitive psychology. In the present paper, we provide comprehensive information on the theory of the diffusion model, as well as on practical issues that have to be considered for implementing the model.

Keywords: diffusion model, mathematical model, response times, fast-dm, EZ diffusion, DMAT

Experimental research in cognitive psychology is often based on speeded response time tasks. Typically, participants have to classify stimuli according to *category membership*, like valence (positive vs. negative), lexical status (word vs. nonword), or familiarity ("old" vs. "new" words in a memory experiment), according to *superficial stimulus properties* (e.g., color or location), or according to *stimulus identity* (e.g., in the Eriksen flanker task, Eriksen & Eriksen, 1974). Performance in such tasks is then compared between conditions (e.g., primed vs. nonprimed words), stimulus types (e.g., high frequency vs. low-frequency words), or between groups of participants (e.g., younger vs. older adults).

Depending on research tradition, either mean response time (RT) or accuracy of responses is used as measure of performance. This traditional approach to data analysis has two major drawbacks: firstly, there is the problem of a missing common metric for performance (Spaniol, Madden, & Voss, 2006; Wagenmakers, 2009) and, secondly, the degree of information usage is poor. We will discuss both problems below before introducing the diffusion model approach (Ratcliff, 1978) and its special advantages.

The problem of a (missing) common metric refers to the fact that the performance in response time tasks can be measured in terms of response times or in terms of accuracy (or using both measures). As mentioned above, research traditions differ on whether mean latencies or accuracy is considered the most important dependent variable. For example, sequential priming effects are more often analyzed in terms of response times, whereas in memory research typically the percentage of correct responses is used. The availability of two measures of performance poses several problems. Firstly, there is the risk of the accumulation of Type I error probability: It might be tempting to interpret and report a significant effect on one metric (e.g., mean latencies), and ignore nonsignificant results on the alternative metric (e.g., accuracy), without making a priori predictions (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Secondly, statistical power might be reduced whenever differences in performance spread over the two metrics, possibly resulting in nonsignificant effects for both mean RTs and accuracy. Researchers tried to respond to this latter problem by introducing response window techniques (e.g., Greenwald, Draine, & Abrams, 1996) that force the complete effect of an experimental manipulation onto the accuracy dimension. However, the extreme time pressure introduced by a response window might change the ongoing cognitive strategies and processes, thus impairing comparability of tasks with and without response windows and endangering external validity of experiments. Finally, the most important problem of the two metrics is based on the possibility of speed-accuracy trade-offs. Whenever directions of effects on RTs and accuracy diverge (i.e., responses in one condition are faster but less accurate or vice versa), it is no longer possible to interpret results in terms of overall

DOI: 10.1027/1618-3169/a000218

performance. In this case, the manipulation (or group membership, etc.) influences decisional style rather than performance (see Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010, for a neural explanation of the speed-accuracy trade-off).

The second weakness of using only mean response time or only accuracy as dependent measure for performance is the relatively poor degree of utilization of the available information. The performance of a participant working through several hundred trials of a response time task is described poorly by one single number (i.e., the mean RT). The available data comprise two RT *distributions* for the two alternative responses that are characterized by their position (e.g., mean), by their specific forms (e.g., standard deviations, skewness), and by their relative sizes representing the percentage of each response. If we use all these information we might get a better understanding of the cognitive processes that are active when the participants perform an experimental task. The information from RT distributions can help to disentangle not only whether performance differs between conditions, but also the way in which it differs and how this difference in performance can be explained in cognitive terms.

The Diffusion Model as a Theory for Binary Choice Decisions

To utilize the full information provided by position, shapes, and sizes of empirical response time distributions of a speeded response time task, it is essential to draw upon a theory explaining the composition of response time distributions from such tasks. Such a theory is provided by the diffusion model (e.g., Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999). The basic assumptions of the diffusion model approach are that during a binary decision information accumulates continuously and that this accumulation of information can be described by a Wiener diffusion process. This Wiener diffusion process is characterized by a constant systematic component, the so-called *drift*, and by normally distributed random noise. The drift rate determines the average slope of the information accumulation process, that is, the speed and direction of information accumulation. The assumption of random noise explains that repeated processing of the same stimulus – or the same type of stimulus – results in different response times, and sometimes in different (i.e., erroneous) responses. Most importantly, the diffusion model can explain the skew typically found in empirical response time distributions (Ratcliff, 2002).

Advantages of the Diffusion Model Approach

The full diffusion model is characterized by several parameters that are discussed below. In a diffusion model analysis, values for these parameters are estimated from empirical response time distributions. Although there are different ready-to-use software solutions for diffusion model analysis (Vandekerckhove & Tuerlinckx, 2007b; Voss & Voss, 2007), the analysis is more complex than simply entering mean response times into an ANOVA. The question therefore arises as to the benefits of performing a diffusion model analysis.

The major advantage of the diffusion model approach is that different cognitive processes are mapped to different psychologically meaningful parameters. Therefore, the diffusion model provides a solution to the aforementioned problem of the missing common metric of traditional analyses of response time tasks: effects do not longer spread over different measures. For example, the influence of differences in performance is disentangled from the influence of decisional styles, because these processes are mapped to separate parameters.

The estimation of different process-pure measures for different cognitive processes makes it possible to test specific theories. We get information not only *whether* participants are slower (or less accurate) in an experimental condition, but also *why* this is so. Imagine – for example – that there is a significant difference in mean RTs between two experimental conditions. This deceleration of responses can be explained (1) by slowdown of information uptake or processing, (2) by a more conservative response criterion, or (3) by a delayed (motoric) response execution. With the diffusion model, it is possible to distinguish empirically between these alternative explanations. In a recent study from our own laboratory (Voss, Rothermund, Gast, & Wentura, 2013) we found priming effects for associative and affective priming tasks that were nearly identical in terms of response times (about 10 ms). However, diffusion model analyses revealed that these effects were based on completely different mechanisms: While semantically associated primes caused a faster identification of the target word, affectively matching primes increased the speed of response execution.

Besides the fact that the diffusion model provides more specific measures, it will in many cases also provide more valid measures for specific research questions. It can also be argued that parameter estimates as process-pure measures are less noisy compared to response time means, which can improve reliability and effect sizes. For example, both cognitive speed and general speed-accuracy settings will influence latencies in a response time task, which will increase both inter- and intra-individual variance of mean response times. If a measure of cognitive speed can be extracted that is independent of other processes, this measure will have less (error) variance in it and reliability might be improved. For example, White, Ratcliff, Vasey, and McKoon (2010) show in a simulation study that the drift rate, that is, the measure for the speed of information uptake, can be more sensitive than mean RT or error rates in detecting group differences between high and low anxious participants in the processing of threatening stimuli (cf. also White, Ratcliff, Vasey, & McKoon, 2009).

The Prevalence of Diffusion Model Analyses in Psychological Research

The diffusion model approach was introduced as a tool for analyzing data from speeded response time tasks three and a half decades ago by Ratcliff (1978). Nonetheless, the usage of this kind of modeling was restricted for quite a long time to a small number of researchers who invested a lot of effort in programming their own software solutions. Recently, however, different tools were simultaneously published that allow the application of the diffusion model without extensive programming skills. These tools comprise the *EZ*-diffusion model (Grasman, Wagenmakers, & van der Maas, 2009; Wagenmakers, van der Maas, Dolan, & Grasman, 2008; Wagenmakers, van der Maas, & Grasman, 2007), *Diffusion Model Analysis Toolbox* (*DMAT*; Vandekerckhove & Tuerlinckx, 2007a, 2008), and *fast-dm* (Voss & Voss, 2007, 2008). We will discuss these programs below, in the section on parameter estimation procedures. The availability of programs for diffusion model analyses has resulted in a strong increase in interest for diffusion model analyses in different fields of psychology. This increase of interest is reflected by an exponential increase in the number of citations of the original publication introducing the diffusion model to psychology (Figure 1). We hasten to add that obviously not all articles citing Ratcliff (1978) are concerned with the diffusion model; however, clearly the vast majority of them will be.

Although the interest in diffusion modeling has grown considerably, this method is far from being a standard method in cognitive psychology. The aim of the present article is to introduce the possibilities and limitations of this form of analysis to a broader audience of researchers from different fields of psychology.

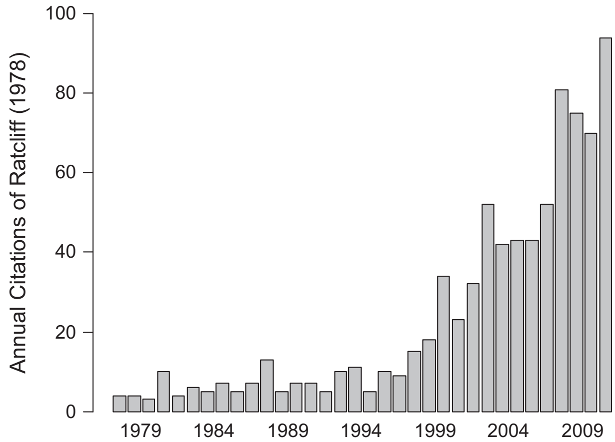

Figure 1. Annual citations of the publication introducing the diffusion model account in psychology (Ratcliff, 1978) from 1979 to 2012 (from the PsycInfo database). The slope reflects a rapid increase of interest in diffusion modeling starting about two decades after the original publication.

The Rationale of the Diffusion Model

In this section, we start with a description of a simplified four-parameter model before introducing several model extensions, followed by a short discussion on how to model performance in different experimental conditions simultaneously.

The Basic Diffusion Model

The diffusion model approach assumes that information accumulates continuously while performing a binary choice task. The accumulated information is represented by an internal counter which is driven in opposite directions by information supporting the different decisional outcomes. For example, in an evaluation task the counter might be increased by positive information and decreased by negative information. The change of the counter over time is modeled as a diffusion process that runs in a corridor between two thresholds. As soon as the upper or lower threshold is hit, the decision is reached and response A or B, respectively, is initiated.

Originally, the diffusion model was introduced as a four-parameter model (Ratcliff, 1978). In this model, performance is described by the average slope of the diffusion process (drift rate: v), threshold separation (a), starting-point (z), and duration of nondecisional processes (t_0). The basic model is depicted in Figure 2. The figure shows three sample paths for the diffusion process. The course of these paths varies from trial to trial – even if identical information is available – because of random noise.[1] This variability of process paths leads to different process durations and different process outcomes. Thus, it is possible to predict decision time distributions for both possible responses from the model parameters (Figure 2).

The most important part in a diffusion model analysis is determining which psychological processes are mapped by the parameters. There are straightforward interpretations for all parameters of the diffusion model: The drift rate (v) maps the speed of information uptake and thus provides a measure of performance. In the comparison of conditions the drift reflects task-difficulty with more difficult tasks represented by smaller drift rates. In the comparison of participants, drift is a measure for individual cognitive or

[1] The amount of noise is determined by the so-called diffusion constant (s) which is a scaling parameter that cannot be estimated but has to be chosen. *Fast-dm* (Voss & Voss, 2007) uses a diffusion constant of $s = 1$ while Roger Ratcliff usually uses a diffusion constant of $s = 0.1$. Estimates for a, z, and v depend on the chosen diffusion constant. These parameters can be transformed to the case of $s = 1$ by dividing the estimated values by the diffusion constant used for the estimation procedure.

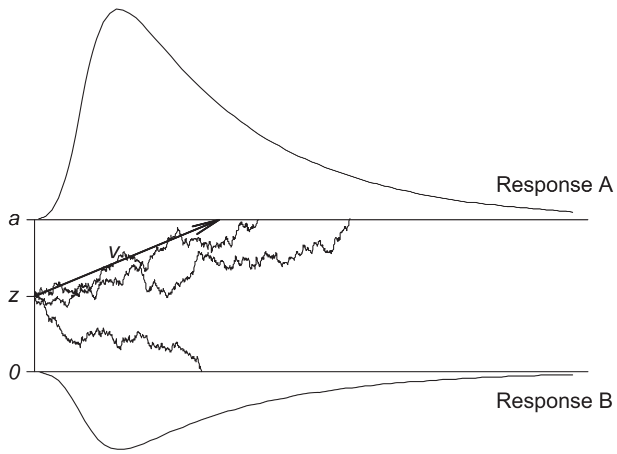

Figure 2. Simplified version of the diffusion model: An information accumulation process starts at starting point z and runs over time with the mean slope v until it hits an upper (a) or lower (0) threshold. Because of random noise, the process durations and outcomes vary from trial to trial. Outside the thresholds decision-time distributions are shown.

perceptual speed of information processing (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007).

Threshold separation (a) is defined by the amount of information that is considered for a decision. A conservative decisional style that is characterized by slow but accurate responding leads to large estimates for a, while liberal responding implies small threshold separations. Different studies have shown that the parameter a is sensitive to speed versus accuracy instructions (e.g., Voss, Rothermund, & Voss, 2004). Additionally, there is a large body of research showing that age-related slowing in response time tasks can be partially explained by more conservative styles of responding (e.g., Ratcliff, Spieler, & McKoon, 2000; Ratcliff, Thapar, & McKoon, 2006, 2010, 2011).

The third parameter of the basic diffusion model, the starting point (z), can map a priori biases in the decision thresholds. Since z can only be interpreted in its relation to a, we prefer reporting the relative starting point $z_r = z/a$. If z differs from $a/2$ (i.e., $z_r \neq 0.5$), different amounts of information are required before deciding on option A or B. Such differences might reflect situations with asymmetric pay-off matrices: For example, Voss et al. (2004) showed that the starting point is moved toward a response threshold when the corresponding response leads to greater rewards. Similarly, in the domain of motivated perception, it has been found that the starting point is closer to the "positive" threshold than to the "negative threshold" in an evaluation task, even when expectancy values for both responses were symmetric (Voss, Rothermund, & Brandtstädter, 2008).

Finally, diffusion model analysis takes into account the duration of nondecisional processes (t_0 or t_{err}). Such processes may comprise basic encoding processes, the configuration of working memory for a task, and processes of response execution (i.e., motor activity). The estimated duration of these processes is added to the decision times predicted by the diffusion process, resulting in a shift of the predicted RT distributions. A common finding is that extra-decisional processes are slowed in elder participants (e.g., Ratcliff et al., 2000). Recently, Schmitz and Voss (2012) showed that task switching costs are partially mapped onto t_0, at least when task switches cannot be predicted. In this case, obviously, the working memory has to be configured for the actual task, before the decision process can start.

Intertrial Variability

To accurately accommodate different shapes of RT distributions for correct responses and errors, Ratcliff suggested extending the model to allow for so-called intertrial variability in performance (e.g., Ratcliff & Rouder, 1998). This extension permits variability of the parameters of the basic model across trials of an experiment. For example, the drift might not be exactly the same for each trial of one condition of an experiment, either because of fluctuations of the participant's attention, or because of differences in stimuli.

Specifically, it has been proposed to model intertrial variability of drift, starting point, and of the nondecisional parameter. The drift is assumed to be normally distributed with mean v and standard deviation s_v (or η). For the sake of simplicity, uniform distributions around z and t_0 with the width s_z and s_{t0} are adopted for the starting point and the nondecisional component, respectively. Recently Ratcliff (2013) showed that these distributional assumptions still lead to valid results if the true distributions differ.

For most applications of the diffusion model, intertrial variability will be of minor interest. However, adopting these parameters sometimes notably improves model fit. For example, Ratcliff and Rouder (1998) showed that large variability of drift can explain slow errors and large variability of starting point can explain fast errors. Nonetheless, the influence of s_v and s_z on predicted response time distributions is rather limited and thus can only be estimated with any reliability from huge data sets. This is different for s_{t0}, which shows a greater effect on the shape of response time distributions (i.e., reducing the skewness).

Differences in Speed of Response Execution

Another suggestion to extend the diffusion model relates to the nondecisional component (Voss, Voss, & Klauer, 2010). Typically, it is assumed that t_0 is equal for both responses. However, this assumption might be wrong whenever motor-response programs differ in level of preactivation. For example, when one response occurs more frequently, is more urgent, or is more likely in a given situation, it is highly plausible that this response will be executed with more vigor, resulting in a faster motor response. Voss et al. (2013) showed that categorical priming (e.g., affective priming) can be explained by a faster execution of the primed response. According to this argument, the prime stimulus (pre-) activates a specific response program, which leads to a faster execution of this response, when it is finally triggered by the target stimulus.[2]

Complex Models: Mapping Different Conditions

Experimental RT paradigms typically comprise different stimulus types or experimental conditions. When modeling such data with the diffusion model, the researcher has to decide whether completely independent models should be estimated for each condition, or whether certain parameters are restricted to be equal across conditions. Especially when data sets are small to medium size (i.e., trial numbers below 200), models will be more stable when all data is fitted simultaneously.

Imagine, as an example, the case of a lexical decision task. In this case, you have at least two types of stimuli (i.e., words and nonwords) that require opposite responses. For this task the upper and lower thresholds of the model represent the responses ''word'' and ''nonword,'' respectively. Obviously two different drift rates are necessary, because for word stimuli the diffusion process will mostly rise to the upper threshold (positive drift), while nonwords will have a negative drift. Also, intertrial variability of drift may vary when stimuli from one class are more similar to those from the other class. However, it is unlikely that the remaining parameters of the model differ between stimulus types, because participants have no information on the next stimulus before it is presented, and consequently cannot adopt starting point or threshold separation to the stimulus of the next trial.

If, however, a task is considered in which participants do have information about the upcoming trial it is possible that parameters other than the drift also have to be estimated for each condition separately. For example, in a task switching paradigm, participants may be aware that switching trials are more difficult to process. Therefore, the threshold separation might be increased, if task switches can be predicted (Schmitz & Voss, 2012).

Finally, there are cases where – even in the presence of different stimulus types – simple models can be adopted that do not differ between different stimuli. To make this possible, data has to be recoded in terms of accuracy; the upper threshold reflects correct responses, and the lower threshold corresponds to error responses. Because there cannot be an a priori bias for or against the correct response, the starting point should be fixed to $a/2$ in this case. Recoding your data in terms of accuracy allows for a more parsimonious model (with only six parameters) at the price of the implicit assumptions that (1) drift rates for different stimulus types are identical in absolute magnitude and (2) that there is no bias in starting point. This assumption can be considered to be met when both types of stimuli lead to the same performance, that is, accuracy, position, and shape of RT distributions are similar.

Theoretical Assumptions and Prerequisites of Diffusion Model Analyses

To decide whether a task is suited for diffusion model analyses, it is important to explicitly review the theoretical assumptions and task prerequisites that have to be met. We will address all important prerequisites in the following section.

Binary Decisions

Firstly, the applicability of the diffusion model is limited by the fact that it is a model of *binary* decision making. Therefore, the diffusion model as presented here cannot account for performance of multiple responses (for a similar multiple response approach see Brown & Heathcote, 2008; Donkin, Brown, Heathcote, & Wagenmakers, 2011). However, even in the case of a multiple choice task, the diffusion model might be applied when it is reasonable to assume that the same processes underlie the different responses, and responses can be recoded as correct (upper threshold) versus error (lower threshold). Consider, for example, the Stroop task (Stroop, 1935). Although there are multiple responses (one for each color), one might try to model accuracy data with the diffusion model, allowing for different drift rates for congruent and incongruent trials. This approach averages performance over different color responses. This procedure is obviously only valid if performance (i.e., response times and accuracy) is similar across trials demanding different responses for both congruent and incongruent trials (see above).

Continuous Sampling

A second basic assumption is that decisions are based on a continuous sampling process. This assumption is obviously plausible for ambivalent visual stimuli that contain information supporting both possible responses, like fields of pixels with different colors that have been used in brightness or color-discrimination tasks (Ratcliff, 2002; Ratcliff, Thapar, & McKoon, 2003; Spaniol, Voss, Bowen, & Grady, 2011; Voss et al., 2004, 2008). However, the successful fitting of data from lexical decision tasks (Ratcliff, Gomez, & McKoon, 2004; Ratcliff, Thapar, Gomez, & McKoon, 2004) demonstrates that even the identification of words is no sudden insight but is based on a continuous – albeit rapid – increase in familiarity. In this case the information entering the decision process comes from long-term memory rather than from an external visual stimulus. The same argument applies to diffusion model accounts of recognition memory (Ratcliff, 1978; Ratcliff, Thapar, & McKoon, 2004; Spaniol, Voss, & Grady, 2008; Spaniol

[2] The possibility to map differences in t_0 between responses will be included in the forthcoming version of *fast-dm* (Voss & Voss, 2007), which will be published soon.

et al., 2006). Here, after encoding the stimuli, memory is the source of information driving the decision process. To sum up, we are confident that in many simple decision tasks information sampling can be conceived as a continuous process.

Single-Stage Decisions

More problematic might be the implicit assumption that decisions are based on a single-stage processing. Whenever participants adopt more complex strategies (e.g., double-check their solutions with an alternative strategy after an initial threshold is reached), the decision process might be divided in multiple steps with cognitive processes varying between steps. Such multiple-stage decisions cannot be mapped with a simple diffusion model as presented in this paper, although more complex models including diffusion processes for separate stages of information processing have been developed, for example, for the visual search paradigm (Guided Search 4.0, Wolfe, 2007) or for the flanker task (Hübner, Steinhauser, & Lehle, 2010; White, Ratcliff, & Starns, 2011).

Constancy of Parameter Values Over Time

Another crucial prerequisite that is related to the single-stage assumption is the assumption of constant parameter values across time. Imagine, for example, a very difficult task where a stimulus contains little or no useful information. In this case participants might be tempted to reduce threshold separation when the information accumulation did not reach the a priori set thresholds after a couple of seconds. Also, in some tasks drift might not be constant over time, for example, when the decision process starts prior to fully encoding a stimulus. In this case, the process might start with a small drift rate that increases in later stages of information processing. Therefore, paradigms with stimuli that are easy to encode are optimal for a diffusion model analysis.

Decision Times

Diffusion model analysis is often only applied to tasks with latencies below 1 s because of the assumptions described above (e.g., Ratcliff & Rouder, 1998). If decisions take notably longer, information processing might comprise qualitatively different stages or parameter values might differ over time. However, this arbitrary limit artificially restricts the area of application for diffusion models. There is no empirical evidence stating that diffusion models cannot be successfully applied for longer decisions (e.g., RT $\approx$ 10 s). For such applications, tests for model fit and parameter validity are obviously of crucial importance.

Numbers of Trials and Percentage of Errors

As well as the theoretical assumptions discussed above that might restrict applicability, there is one practical limitation of the diffusion model analysis: to get reliable estimates for seven – or more, in case of multiple conditions – diffusion model parameters a high number of decisions per participant is essential. For example, in a recent experiment by Leite and Ratcliff (2011, Exp. 2), participants had to complete 5 sessions of 64 blocks with 36 trials each (i.e., 11.520 trials per participant). However, diffusion models have been applied successfully to experiments with notably smaller number of trials: Klauer, Voss, Schmitz, and Teige-Mocigemba (2007, Exp. 2) report diffusion model results that are based on 72 trials per participant. White and colleagues show that the accuracy of parameter estimates can be increased by inserting filler trials in an experiment (White et al., 2009, 2010). For this approach some parameters (e.g., a, z, t_0, s_z, s_{t0}) are considered to be equal across experimental and filler trials which increase the accuracy of the parameters that are estimated separately from a low number of experimental trials (e.g., drift v).

As will be discussed below, the required trial number depends strongly on the estimation procedure (i.e., the adopted optimization criterion), the percentage of errors, and the complexity of the model. Rough estimates of the minimum number of trials that we consider essential for a sound analysis are presented in Table 1.

Parameter Estimation

In a diffusion model analysis, data from each participant are typically modeled separately, resulting in separate sets of

Table 1. Comparison of optimization criteria

	Optimization criterion		
	Maximum likelihood	Chi-square	Kolmogorov-Smirnov
Efficiency	High	Low	High
Robustness	Low	High	High
Computational speed	Low	High	Low
Required number of trials	Small (N > 40)	Large (N > 500)	Medium (N > 100)

Notes. The values for the required numbers of trials are just rough estimates for the lower bound of acceptable trial numbers. See text for further explanations.

estimates for all parameters, which can subsequently be used in inferential statistical analyses. However, sometimes the number of trials for each subject will be too small to estimate parameters accurately. In this case, it might be helpful to collapse data from all participants or from groups of participants with similar performance (so-called "super subjects"; e.g., Ratcliff, Perea, Colangelo, & Buchanan, 2004) for the analysis to increase the database for parameter estimation. A more sophisticated approach to deal with this problem has recently been proposed by Vandekerckhove, Tuerlinckx, and Lee (2011). They present a hierarchical diffusion model approach where participants' individual diffusion model parameters are considered to be random effects. We will address this model in more detail below.

The estimation procedure is based on a multidimensional search for the parameter estimates that leads to an optimal fit of predicted and empirical response time distributions. This search can be computationally costly because of the high number of parameters and because the calculation of predicted distributions takes some time, even for modern high-speed computers.

Figure 3 shows the predicted response time distributions from eight different parameter sets. Model A (with $a = 1$, $z_r = 0.5$, $v = 2$, $t_0 = 0.5$, $s_z = 0$, $s_v = 0$, $s_{t0} = 0$) serves as a comparison standard, and the following seven panels show how the distributions change when the value of a single parameter is modified (Panel B: increased threshold separation; C: increased starting point; D: increased drift; E: increased nondecisional parameter; F: increased variability of starting point; G: increased variability of drift; H: increased variability of the nondecisional parameter). To facilitate comparison, predicted distributions from the comparison model (A) are presented in each panel as hatched areas. A closer look on Figure 3 might help to understand the problems of parameter estimation. Firstly, there is the problem of model mimicry: if you compare, for example, the models with increased starting point (panel C) and increased drift rate (panel D), it is evident that predictions are fairly similar. Differences between both models lie in the prediction of faster error responses in the high-drift model, and in the leading edges of correct responses. Nonetheless, previous studies show that models can be separated successfully when trial numbers are sufficiently high (Ratcliff & McKoon, 2008; Ratcliff & Tuerlinckx, 2002).

A second problem is that some parameters have only minor influence on the predictions. This is especially true for intertrial variabilities of starting point and drift rate: although quite extreme values were chosen for the figure (the starting point follows a uniform distribution from 0.2 to 0.8 in panel F and drift follows a normal distribution with mean 2 and standard deviation 2 in panel G), the influence on the RT distributions is limited. Obviously, one needs large empirical distributions (large trial numbers) to estimate these parameters with any accuracy. In the case of small to medium trial numbers one might decide to fix these parameters to zero to make the model more parsimonious and to enhance stability of the estimation procedure.

Comparison of Optimization Criteria

For the multidimensional search for the optimal vector of parameter values, an optimization criterion has to be defined that quantifies the fit between predicted and empirical RT distributions. Because the choice of these criteria has influence on the speed, precision, and robustness of the estimation procedure (Ratcliff & Tuerlinckx, 2002), we will briefly discuss three different approaches in the following sections.

Maximum Likelihood

Maximum likelihood (ML) approaches are mathematically most efficient (cf. Klauer et al., 2007, for an example adopting this method). For this approach, the logarithmic density of predicted RT distributions is summed over all responses, and this sum is maximized in the search. The drawback of this method is that results can be strongly biased by single outliers. For example, when using a ML approach, a single fast guessing response (that does not result from a diffusion process) might force t_0 to be very small, because otherwise this fast response would lead to a density of zero, rendering the total likelihood to be zero as well (log-likelihood is no longer defined in this case). Another disadvantage is that calculation can be very slow with large trial numbers. Consequently, we recommend using a ML-based search only if data sets are so small that other optimization criteria fail and if a careful outlier analysis was conducted.

Chi Square

Most frequently used are searching algorithms based on the χ^2 statistic (e.g., Ratcliff & McKoon, 2008; Ratcliff & Tuerlinckx, 2002). This procedure uses quantiles from the empirical RT distributions to define bins on the RT axis. Ratcliff suggests using six bins for each threshold that are defined by the .1, .3, .5, .7, and .9 quantiles of the empirical RT distributions. The outer (open) bins contain 10% of data, while all inner bins comprise 20% of trials. From the predicted cumulative distribution functions the number of responses expected for each bin is calculated by multiplying the portion of the predicted distribution for each bin by the total number of trials. Then, χ^2 is calculated from the numbers of observed and predicted responses from the 12 bins (six for the upper and six for the lower threshold):

$$\chi^2 = \sum \frac{(n_{\text{observed}} - n_{\text{predicted}})^2}{n_{\text{predicted}}}$$

Advantages of the χ^2 approach are the fast calculation (independent of trial numbers), and the robustness against outliers. Since the first and last bins are open bins, only the numbers of responses within these bins are important, not the actual latencies of each response. Therefore, even an extreme outlier (e.g., RT = 0 ms) would not distort results dramatically. However, these advantages come at

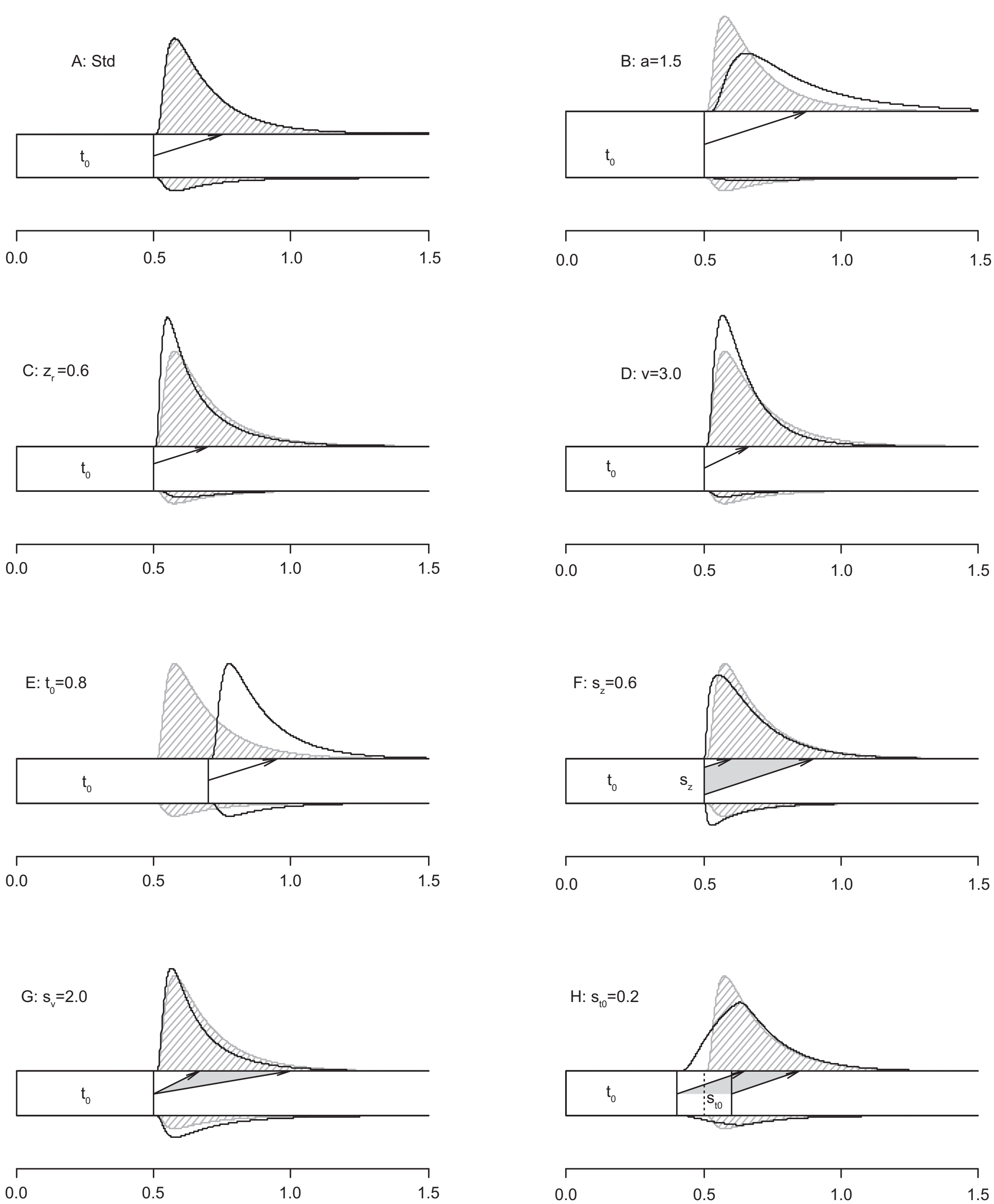

Figure 3. Predicted RT distributions from different parameter sets. Panel A shows predictions from a comparison model with $a = 1$, $z_r = 0.5$, $v = 2$, $t_0 = 0.5$, $s_z = 0$, $s_v = 0$, $s_{t0} = 0$. In the following panels, always one parameter is increased. To facilitate comparison, the distributions from Panel A are presented as hatched shapes in each display.

a cost as well: Due to the binning, information is lost and in case of small trial numbers the identification of empirical quantiles will be inaccurate. This is especially problematic in experiments with high accuracy and hence few error responses. Therefore, we recommend using a χ^2 approach only for studies with large trial numbers (i.e., at least 500 trials), and enough error trials; the smaller distribution should have at least 20 responses for each participant, so that the first and last bins each comprise two or more responses. If there are fewer errors, results of the estimation procedure tend to depend strongly on the handling of these (e.g., ignoring error information, collapsing all errors in one bin, or nonetheless using six error bins). For example, in a simulation study using a χ^2 approach White et al. (2010)

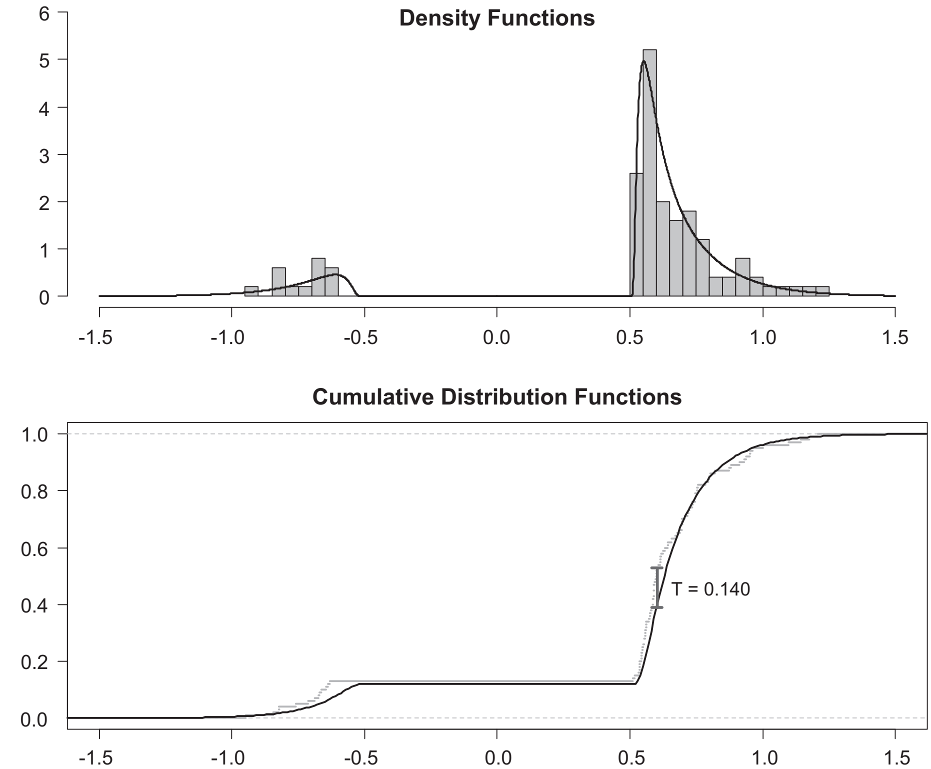

Figure 4. Illustration of the Kolmogorov-Smirnov approach. Distributions from both thresholds are combined in one distribution function by multiplying all times from the lower threshold by −1. The upper panel shows the comparison of empirical (histogram) and predicted (lines) density functions. The lower panel shows the cumulative distribution functions (CDF; data: gray line; predictions: black line). In the parameter search, the maximum vertical distance (T) between both CDFs is minimized.

found systematically biased parameter estimates from samples with 120 trials split into four experimental conditions, whereas results were accurate using 860 trials. The authors conclude that the problems with small sample sizes arise from low error numbers.

Kolmogorov-Smirnov

A third possibility for the optimization criterion is based on *the Kolmogorov-Smirnov (KS) statistic* (Voss & Voss, 2007; Voss et al., 2004). This statistic is the maximum vertical distance between predicted and empiric cumulative response time distributions. The application of the KS method for diffusion model analysis is complicated by the fact that there are always two empirical and two theoretical distributions (for both thresholds) that have to be compared simultaneously. This problem has been solved by Voss et al. (2004) by multiplying all RTs from the lower distribution by −1. Thus, both distributions can be combined on a single dimension without overlapping each other (Figure 4). The KS approach can be considered as a compromise between the highly efficient ML method and the more robust χ^2 method. The KS approach – like the χ^2 method – provides robust estimates in the presence of outliers and simultaneously – like the ML method – considers the exact shape of the response time distribution without categorizing responses.

Table 1 sums the strengths and weaknesses of the three optimization criteria. Efficiency reflects the ability to accurately recover true parameter values from small data sets, robustness reflects the stability of estimates in the presence of outliers, and calculation speed points to the duration of the complete parameter estimation procedure.

It is very difficult to provide a recommendation for a minimum number of trials that is required for robust parameter estimations. Generally, estimations will be more precise for data sets comprising a high percentage of error responses (i.e., when there are distributions of reasonable size at both thresholds). Secondly, the estimation procedure tends to be more robust when the number of free parameters is reduced. Especially fixing the starting point to $z = a/2$ notably increases the stability of results. Most important, when there are participants that show virtually only one type of response (e.g., participants with perfect accuracy in a model based on accuracy data), fixing the starting point is indispensable, because then the distance from starting point to the threshold without responses is no longer defined. Finally, the necessary trial number depends on the number of experimental conditions, that is, models might be more stable when different conditions are modeled simultaneously, while some parameters are fixed across conditions.

Although the exact dependency of the required number of trials on these factors is still unclear, we decided to give rough recommendations for the minimum trial numbers we consider to be necessary for an acceptable diffusion model analysis, because this is one of the most frequent questions posed by cognitive researchers who think about applying the diffusion model to their data. Note however, that larger trial numbers are strongly recommended.

The Hierarchical Diffusion Model

Recently, Vandekerckhove et al. (2011) introduced a hierarchical approach to diffusion model analyses. The core idea of the hierarchical diffusion model (HDM) is to simultaneously estimate parameters for the complete sample while maintaining the possibility of describing and explaining interindividual differences. Like in other hierarchical (or mixed) models it is possible to include random effects. For example, it can be assumed that persons (or trials) are randomly drawn from a population and the distribution of parameters in this population can be estimated. A further advantage of the hierarchical framework comprises the possibility to include any theoretically founded restrictions (e.g., restraining parameters to be equal across participants or across experimental conditions).

Unless many restrictions are made, for a HDM analysis a high number of parameters need to be estimated simultaneously. Vandekerckhove et al. (2011) address this problem by basing the estimation algorithm on a Bayesian approach using Markov chain Monte Carlo (MCMC) simulations (see Kruschke, 2011, for an excellent introduction to Bayesian statistics). In this procedure, theoretically accounted prior distributions of all parameters gradually undergo coverage in an iterative procedure to the final posterior distributions. From these posterior distributions confidence intervals for true parameter values can be inferred (e.g., centering on the point of highest density). Vandekerckhove et al. (2011) provide an extension for the WinBugs software (Lunn, Thomas, Best, & Spiegelhalter, 2000) to conduct a HDM analysis.

Existing Software Solutions

In order to facilitate the application of diffusion model analyses different software solutions have been developed allowing parameter estimation also to researchers with limited programming experience. Important differences between these programs regard the degree of information used for parameter estimation and the number of model parameters that can be estimated.

For the *EZ-diffusion model* (Grasman et al., 2009; Wagenmakers et al., 2007, 2008) a JavaScript, an Excel sheet, R code, and a MATLAB implementation are available (see http://www.ejwagenmakers.com/papers.html for links to the respective implementations). The *EZ-diffusion model* makes use of a limited degree of information of the observed RT distributions. Only the mean and variance of the correct responses and the accuracy rate are used for parameter estimation. The estimation procedure is – as the name of the program implies – *easy* as parameter estimates can be immediately obtained by entering the three calculated values (mean, variance, and accuracy) into three equations. Thereby parameter estimates for the basic diffusion model can be obtained, that is, for drift rate, threshold separation and duration of nondecisional processes. The parameter calculation via these closed-form equations is very fast as no time-consuming iterative optimization process has to be applied. However, the restrained use of information (especially about the error trials – only the percentage of observed errors is considered) does not allow the estimation of intertrial variabilities. These are implicitly assumed to be equal to zero. In the standard *EZ* model the starting point cannot be calculated either. It is fixed to the midpoint of the threshold separation ($z = a/2$) while *EZ2* (Grasman et al., 2009) – a more recent, extended version of *EZ* – allows the calculation of an estimate for the starting point. Furthermore *EZ2*, in contrast to *EZ*, allows the estimation of different values for a parameter depending on diverse conditions (e.g., one drift rate for words, another for nonwords) while at the same time the other parameters are held constant over conditions. Another extension of *EZ*, so-called *robust-EZ* (Wagenmakers et al., 2008), deals with contaminant data.

In contrast to *EZ*, *DMAT* (Vandekerckhove & Tuerlinckx, 2007a, 2008) and *fast-dm* (Voss & Voss, 2007, 2008) utilize more information from the RT distributions to draw conclusions about the decision processes. *DMAT* is a MATLAB toolbox which is available from the website http://ppw.kuleuven.be/okp/software/dmat/. *Fast-dm* is a command-line program implemented in C that can be downloaded from the website http://www.psychologie.uni-heidelberg.de/ae/meth/fast-dm/. While *EZ* only requires the mean and variance of correct responses and the accuracy rate, for the use of *fast-dm* and *DMAT* all correct and error response times have to be supplied to the program. A file with at least two columns is needed: one coding the accuracy of the response (error vs. correct response), the other containing the response times. Considering accuracy rate and distribution of correct as well as error responses *DMAT* and *fast-dm* allow the estimation of all of the diffusion model parameters, that is, inclusive of starting point and intertrial variabilities. Like *EZ2* both *DMAT* and *fast-dm* comprise the option of restricting parameters across conditions while letting other parameters vary between these conditions.

Fast-dm and *DMAT* use the parameters v, a, and t_0 as estimated by *EZ* as starting points for an iterative optimization routine. The principal difference between *DMAT* and *fast-dm* lies in the optimization criterion used for parameter estimation. While *DMAT* is based on the χ^2 statistic *fast-dm* uses the Kolmogorov-Smirnov (KS) approach. Therefore, *fast-dm* is characterized by the usage of the complete distributional information, while DMAT draws upon the number of RTs in different "bins" on the RT axis. As outlined above (see section Comparison of Optimization Criteria) the higher degree of information usage of the KS method implies longer calculation times but on the same time might lead to higher efficiency of parameter estimation.

When comparing performance and results of different diffusion model programs it has to be considered that *fast-dm* uses a diffusion constant of $s = 1$ while *DMAT* and *EZ* fix s to 0.1. Therefore *DMAT* and *EZ* estimates for all parameters except for t_0 and st_0 have to be divided by 0.1 to establish comparability with *fast-dm* results (see Footnote 1). First simulation studies in order to compare the parameter recovery of *fast-dm*, *DMAT*, and *EZ* have been conducted by van Ravenzwaaij and Oberauer (2009). Regarding the correlation between the true parameter values (on which the simulated data were based) and the estimated parameter values *EZ* and *fast-dm* emerged as superior to *DMAT*. *EZ* and *DMAT* performed better than *fast-dm* in terms of the recovery of the mean true values. However, more studies systematically varying trial numbers, parameter ranges, and contamination by outliers are necessary to determine which algorithms are superior for which data and which research questions.

Typical Experimental Paradigms for Diffusion Model Analyses

In the following section, we will present short overviews of three experimental paradigms that have been frequently and successfully used for diffusion model analyses. These typical diffusion model paradigms described here comprise (1) brightness- or color-discrimination tasks, (2) recognition memory tasks, and (3) the lexical decision task. Additionally, we present an overview over the application of DM on data from prospective memory tasks. For studies employing the diffusion model approach to analyze general principles of information processing (e.g., age-related differences in information processing: Ratcliff et al., 2000, 2003; Ratcliff, Thapar, Gomez, et al., 2004; Ratcliff, Thapar, & McKoon, 2001, 2004; Spaniol et al., 2006; Thapar, Ratcliff, & McKoon, 2003) these well-tested paradigms are recommended because validity of parameters has been tested for these tasks.

Brightness, Color, or Numerosity Discrimination

Diffusion model analyses have been applied in numerous studies requiring the classification of ambiguous stimuli with respect to brightness (Ratcliff, 2002; Ratcliff & Smith, 2010; Ratcliff et al., 2003, 2006) or color (e.g., Spaniol et al., 2011; Voss et al., 2004, 2008). In these experiments, participants see squares that are composed of a random pixel pattern of two different brightness or hue values (e.g., black vs. white or orange vs. blue). The task is to classify stimuli according to the dominating color, that is, to judge which kind of pixels appears in greater number. Across trials the frequency of occurrence of colors is varied (e.g., 44%, 48%, 52%, or 56% of pixels are white). Structurally very similar are numerosity discrimination tasks that require participants to judge whether a high or low number of stimuli (e.g., asterisks) is presented (e.g., Leite, 2012; Ratcliff, 2008).

Such color, brightness, or numerosity discrimination tasks are very well suited for diffusion model analyses because they meet the theoretical assumptions of the model particularly well. It is highly plausible that performance in these tasks is based on a one-stage continuous information accumulation process, where drift is determined, for instance, by the ratio of presented colors. Another advantage is that stimuli are artificial and initially meaningless, thus making a priori biases unlikely. Therefore, stimuli can be assigned with valence or meanings experimentally to study biases in decision making (Voss et al., 2008). Finally, such discrimination tasks are very easy to learn, making them a flexible tool for many research questions and even rendering possible the application in animal research (Ratcliff, Hasegawa, et al., 2011; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007).

Recognition Memory

Recognition memory is the paradigm for which the diffusion model was originally conceived (Ratcliff, 1978) and where it is still frequently applied (e.g., McKoon & Ratcliff, 2012; Ratcliff, Thapar, & McKoon, 2004; Spaniol et al., 2006; White et al., 2010). The task comprises a study-test paradigm, with stimuli – usually words – being presented once or more in an acquisition phase. In a later recognition phase participants have to judge whether presented stimuli are "old" or "new." Other memory tasks can be used as well. For example, participants can be asked to decide whether a pair of stimuli has been shown together before in the same way ("intact pair") or whether it has been rearranged. For recognition memory, like for the brightness discrimination paradigm, the theoretical assumptions of the model can be considered to be well met, in that there is a single fast binary decision. It is assumed that an isolated, recently formed trace in memory supplies the evidence that is accumulated.

In recent applications, Starns, Ratcliff, and White (2012) and Starns, White, and Ratcliff (2012) used the diffusion model to improve understanding of the strength-based mirror effect (Criss, 2009, 2010). This effect reflects the finding that strong memory traces for targets help to reject lures efficiently in a recognition memory task. Diffusion model results indicate that this effect is rather based on a setting of a so-called drift *criterion*, that is a subjective zero point of familiarity of a target stimulus, rather than by a better classification of lures as nontargets. This is particularly relevant for future applications of the diffusion model as it demonstrates that the drift rate is not only affected by the actual stimulus that has to be classified but also by the general context of the decisional situation.

Regarding the effect of age on recognition memory Ratcliff, Thapar, et al. (2011) and McKoon and Ratcliff (2012) report that *nondecision time* and *boundary separation* increase with age, whereas *drift* remains fairly constant (but see Spaniol et al., 2006). In contrast, for associative recognition – that is, the ability to tell whether a particular pair of word stimuli was presented jointly previously – *drift rate* decreased with age. Some studies also investigate the influence of intelligence on diffusion model results from recognition memory: Drift rate is generally positively associated with intelligence, but in the associative memory task this was much less the case for elder compared to younger participants (McKoon & Ratcliff, 2012).

Lexical Decision

Several studies used the diffusion model to analyze performance in lexical decision tasks (e.g., Ratcliff, Gomez, et al., 2004; Ratcliff, Thapar, Gomez, et al., 2004). This task requires participants to decide quickly whether presented letter strings are valid words. The diffusion model account of the lexical decision task is silent on the details of lexical access and offers a – poorly defined – concept of "wordiness" in its stead (Norris, 2009; Ratcliff, Thapar, Gomez, et al., 2004). Following this concept, words and nonwords differ in their degree of wordiness, that is, in their typicality for the category "word." One main finding of the diffusion model approach to the lexical decision task is that many aspects of stimuli map onto *drift rate*. For example, high frequency words show larger drift rates compared to low-frequency words, and random letter strings have stronger (negative) drift rates compared to word-like nonwords (Ratcliff, Gomez, et al., 2004).

Yap, Balota, Sibley, and Ratcliff (2012) found that participants' ability is reflected in *drift* and that those participants who do well in the lexical decision task show less pronounced effects of lexical variables such as length/structure, neighborhood, and frequency/semantics. Diffusion model analyses revealed lower drift rates, lower nondecision times, and larger boundary separations for reading impaired children (Zeguers et al., 2011).

Recently, the lexical decision task has also been applied in diffusion model studies analyzing different forms of sequential priming (Voss et al., 2013; Yap, Balota, & Tan, 2013). Voss et al. (2013) show that semantic (associative) priming increases *drift rate* (i.e., it facilitates lexical access), while categorical priming reduces the *nondecisional component* (i.e., speeds response execution). Yap et al. (2013) obtain a more complicated pattern of results for priming: priming mapped onto *nondecisional time* when stimuli were presented clearly visible and influenced *drift* and *nondecisional time* when stimuli were degraded (cf. also Gomez, Perea, & Ratcliff, 2013).

Event-Based Prospective Memory

Recently, the diffusion model was applied to analyze effects of prospective memory tasks on performance in the lexical decision task (Horn, Bayen, & Smith, 2011) and a color matching task (Boywitt & Rummel, 2012). The most interesting research question in this context is how and why performance in a cognitive task is impaired when participants have to remember a specific intention, for example, the intention to show a distinct reaction when a specific stimulus is presented. Theoretically, impairments could be based on (a) a draining of cognitive resources by the prospective memory task that leads to a slower processing of the targets in the ongoing cognitive task (reduction of v), (b) a more cautious speed-accuracy setting (increase of a), or (c) additional processes in the encoding or execution phase (increase in t_0).

Results from Horn et al. (2011) as well as from Boywitt and Rummel (2012) show consistently that increased latencies in prospective memory conditions can in large part be explained by more conservative speed-accuracy settings. Boywitt and Rummel (2012) demonstrate that this shift in response thresholds is of strategic nature: The setting of the speed-accuracy trade-off is based on expectations about the prospective memory task rather than on the real experience of this task.

Furthermore, there is evidence that prospective memory can reduce drift and slow down nondecisional processes as well. However, these effects were only found for high demanding prospective memory tasks, that is, tasks in which cues for prospective memory were highly similar to other targets (Boywitt & Rummel, 2012).

A Practical Guide to the Application of Diffusion Models

In the following section we give some important practical advice on how to conduct a state-of-the-art diffusion model study. Specifically, we will address issues of experimental design, data pre-treatment, model specifications, and tests of model-fit and parameter validity.

Experimental Design: Which Task Should Be Implemented?

In a first step, an adequate experimental response time paradigm must be chosen. It is always preferable to draw upon tasks that have already been tested for diffusion model analyses before. If a new task has to be used, the theoretical assumptions of the model have to be considered carefully, and the validity of the model should be analyzed empirically in elaborate pre-studies that should be conducted independently of an application question; we will address the issue of empirical parameter validation below (see Voss et al., 2004, for an example of a parameter validation study).

Once an apt task has been identified the researcher has to decide how many conditions or stimulus types should be used in the experiment. In our experience, the inclusion of different stimulus types that differ in task difficulty (e.g., four types: easy and difficult stimuli requiring response A or B) often increases robustness of the estimation procedure.

Finally, the number of trials has to be chosen. Generally large trial numbers (e.g., $N \geq 500$) support the accurate estimation of parameter values. However, such massive testing is not only costly in terms of time and money, but extensive practice might also change the underlying cognitive processes in the later blocks or sessions of an experiment (Dutilh, Krypotos, & Wagenmakers, 2011). Therefore a compromise between highly accurate parameter estimation and the practical possibilities has to be adopted. Generally, trial numbers need to be high when the model test is of key importance (i.e., when new paradigms are tested) or when parameters need to be estimated with high precision (e.g., for correlative research). Lower trial numbers might be sufficient when parameters are tested for group differences and when simplified versions of the model are applied.

Analyzing Your Data

In this paper, we cannot give a comprehensive tutorial on the different computer programs for diffusion model analysis (see section Existing Software Solutions). However, we will present an overview of the typical procedural steps and associated decisions that have to be made.

Data Pre-Treatment

Although the diffusion model is designed to predict the complete response-time distribution, the removal of outliers from the individual response-time distribution is highly recommended. For the analyses, fast outliers (e.g., fast guesses) are generally more problematic than slow outliers. Since χ^2 and KS-based estimation procedures are relatively robust, liberal criteria for lower and upper outliers (e.g., *fast outliers*: RT < 200 ms; *slow outliers:* RT > 5,000 ms) will often be sufficient. For the ML method, stricter criteria that are derived from individual response time distributions are preferable. For example, all responses 1.5 interquartile distances below the first quartile or above the third quartile of the individual RT distributions might be excluded (outliers sensu Tukey, 1977). Finally, Ratcliff and Tuerlinckx (2002) suggest a highly sophisticated procedure to remove fast outliers. The basic idea is that fast outliers should have an accuracy of 50%. An upper limit for fast outliers is increased continuously until performance rises above chance, that is, until more correct than erroneous responses are observed. All trials with RTs below the so found limit are discarded. However, this procedure is feasible only for easy tasks with low numbers of errors.

Grouping of Data

Typically, data is modeled for each participant separately. This will require saving data in separate data files. If the individual data sets are too small for a sound parameter estimation, one might consider collapsing data across the complete sample or across so-called super-subjects (Ratcliff, Perea, et al., 2004), that is, across participants with similar response time distributions and error percentages. Grouping has always the disadvantage that it is unclear whether the estimated parameters are valid measures, because cognitive processes might differ between participants, even if overall performance is similar. Another problem is that subsequent statistical comparisons are impossible or lack power.

Mapping Actual Responses Versus Accuracy

For the diffusion model analysis, responses have to be linked to thresholds. In a binary choice task, this can be done by maintaining the actual responses (e.g., upper threshold = "word," lower threshold = "nonword" in a lexical decision task), or by linking thresholds to accuracy (i.e., upper threshold = correct response, lower threshold = error). In the first case, one has to use separate drift rates for alternate stimulus types, and drift for the stimuli requiring the response linked to the lower threshold will be negative. To compare speed of information uptake between stimulus types, *absolute values* of the drift have to be checked against each other. In this model the starting point reflects a decisional bias, with shorter distances between starting point and threshold signaling a bias in favor of the outcome connected to this threshold.

In the second case – that is, when accuracy data is modeled – one drift rate is sufficient, and implicitly the assumption is made that performance is equal across stimulus types.[3] When thresholds are linked to accuracy the starting point should normally be fixed to $z = a/2$, because – logically – there cannot be an a priori bias toward (or against) the correct response. However, if there is a priori information available in a trialwise fashion that can be congruent or incongruent with the target stimulus (e.g., primes in a sequential priming paradigm) it is possible that this information influences the starting point. Such an effect does not reveal a general bias toward one of the response categories but rather gives information on how the prime information does influence decision making (Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Voss et al., 2013).

Varying Parameters Between Stimuli or Conditions

Often, diffusion model analyses are employed to test empirically which parameters account for the effect of an experimental manipulation. To answer this question, *different*

[3] Different drift rates could be used in the model on accuracy data as well. However, then recoding will not make the model more parsimonious and parameter estimation will be less stable because the distribution at the lower (error-) threshold will be rather small.

parameters must be allowed to vary between conditions or stimulus types. If only one parameter is allowed to vary, any present effect will be mapped on this parameter. One possibility is to split the data and estimate completely independent models for each condition.[4] Thus, all parameters could possibly account for an effect of the manipulation. Often more parsimonious models can and should be chosen. For example, it might make sense to assume that threshold separation and starting point are constant across trials. Also, for the sake of simplicity, it might often be helpful to fix variability parameters across conditions and stimulus types.

Selection of an Optimization Criterion

As discussed above (see section Comparison of Optimization Criteria) the criteria have different advantages (see Table 1 for a comparison). We recommend using the ML approach for small, the KS approach for medium, and the χ^2 approach for large trial numbers. Until now, these criteria are implemented in different software solutions. In the forthcoming version of *fast-dm* it will be possible to choose between the three optimization criteria discussed above (ML, KS, or χ^2).

Interpretation of Parameter Values

In the last step, the estimated parameter values are entered as dependent variables into statistical analyses. Thus, it is possible to check which parameters account for differences in performance between groups (e.g., younger vs. elder participants, Ratcliff et al., 2000), stimulus types (e.g., high frequency vs. low frequency words in lexical decisions, Ratcliff, Gomez, et al., 2004), experimental manipulations (e.g., speed vs. accuracy instructions, Voss et al., 2004), and so on. Another strategy is to use parameter estimates for correlational analyses (e.g., predicting intelligence scores, Schmiedek et al., 2007). However, there are two caveats that should be considered prior to the interpretation of diffusion model results. Results can only be considered valid if, firstly, the chosen model fits the data well, and, secondly, if one can be sure about the psychological meaning of the parameters. Both issues will be discussed in the following sections on Model Fit and on Empirical Validation Studies, respectively.

Model Fit

A crucial precondition for the interpretation of diffusion model results is an acceptable model fit. Only if the model can recover response time distributions and accuracy rates adequately results might reflect the ongoing cognitive processes. Different strategies have been developed to assess model fit.

Statistical Tests of Model Fit

Firstly, it is possible to assess model fit via statistical tests: The χ^2 statistic as well as the KS statistic can be directly translated into *p*-values from the corresponding statistical test for the comparison of predicted versus empirical response time distributions.[5] Small values of p (e.g., $p < .05$) indicate that the diffusion model cannot account for the data. However, the interpretation of these *p*-values has several problems: (1) Firstly, because the shapes of the predicted RT distributions are fitted to the empirical distributions the tests will tend to be too conservative, that is, models will be rejected too seldom (D'Agostino, 1986). (2) Secondly, if several conditions are fitted simultaneously (i.e., if at least one parameter is free to vary between conditions), *fast-dm* reports the product of all *p*-values from the different conditions. Therefore, the displayed *p*-values might be very small in case of multiple conditions. (3) Thirdly, results from statistical tests depend strongly on the number of trials: In case of small or medium trial numbers, on the one hand, the power of both χ^2 test and KS test is small and – consequently – misfits will often not be detected. On the other hand, an *exact* model fit cannot be expected, because – albeit being quite sophisticated – diffusion models as any theory propose a simplified model of reality. Therefore, statistically significant misfits are to be expected in case of large trial numbers and might be considered rather unproblematic (e.g., Ratcliff, Thapar, Gomez, et al., 2004; White et al., 2009).

Graphical Displays of Model Fit

Because of the problems related to the statistical tests of model fit, graphical displays of concordance of predictions with data have been proposed. A good way to do this is to present a display of fit for each person and each condition separately by plotting the predicted and empirical cumulative distribution functions (CDF) in the same graph.[6]

[4] If all parameters are allowed to vary between stimulus types it is highly recommended to use separate data files and thus separate runs of the estimation program. Theoretically *fast-dm* or *DMAT* could estimate all parameters for all conditions in one search. However, the multidimensional search procedure (the SIMPLEX algorithm, Nelder & Mead, 1965) has problems finding the optimal solution when too many parameters are optimized simultaneously.

[5] For the χ^2 statistic, the degrees of freedom are calculated as $df = (K*(N-1)) - M$, where K is the number of conditions, N the number of bins for correct responses and errors (usually 12), and M the number of free diffusion model parameters (White et al., 2010). The KS statistic can only be translated into a *p*-value for one condition at a time. Here, the degrees of freedom are given by the number of trials within the condition.

[6] Predicted CDFs can be computed, for example, by the plot CDF routine from the *fast-dm* software (Voss & Voss, 2007).

Another possibility is provided by so-called quantile probability plots that display quantiles of the RT distributions as a function of response probabilities for different conditions or stimulus types (e.g., Ratcliff & Smith, 2010). However, in many studies there are too many participants to present separate figures for each model. In this case it might be helpful to average CDFs across participants (Schmitz & Voss, 2012, Appendix A). Another possibility to display model fit simultaneously for many participants is provided by scatter plots. Here, empirical values (x-axis) of accuracy and of quartiles of the RT distributions are plotted against predicted values (y-axis; Voss et al., 2013, Appendix B). The main problem of all types of graphical display of model fit is the ambiguity of interpretation: There is no clear criterion on how much deviance of data from predictions is acceptable.

Monte-Carlo-Simulations

Monte-Carlo simulations provide a highly sophisticated possibility of overcoming the discussed biases of p-values from statistical model tests (Clauset, Shalizi, & Newman, 2009). For this purpose, many (e.g., 1,000) data sets have to be simulated from the diffusion model, matching the characteristics of the empirical data. That is, parameter values for the simulation should be based on the estimated parameter values, and the numbers of trials, conditions, etc., should be equivalent to those used in the experiment. To accomplish this, parameter values for the simulation study might be generated following the multivariate normal distribution defined by the mean values and the variance-covariance matrix from the estimated parameters (e.g., using the *mvrnorm* routine from the MASS R-package). For the simulation, the construct-samples routine from *fast-dm* (Voss & Voss, 2007) can be used.[7] These simulated data sets are then re-analyzed with the diffusion model. From the results a distribution of fit-values (e.g., p-values provided by *fast-dm*) can be obtained, and the 5% quantile of this distribution can be taken as a critical value to assess model fit of the empirical models.

Interpretation of Model Fit

If suspicious fits are found for only a low percentage of participants (e.g., less than 5%), it can be assumed that the diffusion-model generally describes data well. Obviously, a good model fit does not prove that the diffusion model is the ‘‘correct'' model. Especially for small samples, nonsignificant results have to be interpreted with great caution. In any case, it is recommended to discard data from participants with bad model fit before running further analyses.

If data from a substantially larger portion of participants cannot be fitted, the diffusion model in the applied form has to be discarded. Possibly, a stricter exclusion of outliers or a more complex model with fewer restrictions can fit the data in this case. Often, however, bad model fit indicates that the theoretical assumptions of the model are not met.

Empirical Validations

It is difficult to assess the adequacy of diffusion models with an analysis of model fit alone. Additional support should be provided by empirical validation studies. Face-valid experimental manipulations have to be used for each parameter of the diffusion model. Then, independent models are estimated for each experimental condition, where manipulations could map on all parameters. If each manipulation successfully and exclusively influences the expected parameter(s), the validation can be considered successful. An example for an elaborated validation study for the color-discrimination task is provided by Voss et al. (2004). In a series of experimental manipulations with high face validity the authors showed that it is possible to manipulate single parameters in the proposed way: For example, task-difficulty exclusively mapped on the drift parameter while speed-accuracy instructions did influence threshold separation. Another example for empirical parameter validation in a prospective memory task is provided by Boywitt and Rummel (2012).

We recommend employing such validation studies whenever the diffusion model is applied to a new task. A successful validation might even be considered to be more important than the demonstration of an excellent model fit, because model fit will often be good in case of small trial numbers.

Conclusions and Perspectives

The diffusion model is not a new approach in psychology. However, in the first two decades after its initial proposal (Ratcliff, 1978), it was used rather rarely. This slow increase in popularity has several reasons: Firstly, the implementation of this method was difficult, before software solutions like *fast-dm* or *DMAT* were published. Secondly, access to computational power, which is necessary for this kind of analyses, was limited. And finally, there is still a general lack of knowledge about the possibilities and problems of diffusion model analyses. With the present paper we hope to address this last hurdle, and help to clear the way to diffusion model analyses for a broad community of cognitive researchers.

Diffusion modeling is still far from being a standard method in cognitive sciences. Nonetheless, in recent years a rapid increase in methodological developments as well as in applications of the diffusion model in many research domains is evident. New fields of application comprise, for example, research on effects of practice (Dutilh et al., 2011), intelligence (Ratcliff et al., 2010; Schmiedek et al., 2007), or clinical psychology (White et al., 2009, 2010).

[7] If there are multiple conditions, data sets have to be simulated separately for each condition, and afterwards combined into one file.

It is hard to predict further developments in this highly dynamic field of research. One trend seems to be an increased interest in the measurement of interpersonal differences in cognitive abilities and mental settings with the diffusion model. Before this model can be used as a diagnostic tool, much more needs to be known about stability of estimation procedures and of the psychometric quality of diffusion model parameters. Of special importance are clear recommendations for standards for experimental setups, parameter estimation, analyses of model fit, etc. With the present paper, we try to contribute to the discussion about these points and to help to spread knowledge about the diffusion model.

Acknowledgments

This research was supported by a grant from the German Research Foundation to Andreas Voss (Grant No. VO1288/2-1).

References

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences, 33*, 10–16. doi: 10.1016/j.tins.2009.09.002

Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition, 40*, 70–82. doi: 10.3758/s13421-011-0128-6

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*, 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51*, 661–703. doi: 10.1137/070710111

Criss, A. H. (2009). The distribution of subjective memory strength: List strength and response bias. *Cognitive Psychology, 59*, 297–319. doi: 10.1016/j.cogpsych.2009.07.003

Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 484–499. doi: 10.1037/a0018435

D'Agostino, R. B. (1986). Tests for the normal distribution. In R. B. D'Agostino & D. A. Stephens (Eds.), *Goodness-of-fit techniques*. New York, NY: Marcel Dekker.

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes?. *Psychonomic Bulletin & Review, 18*, 61–69. doi: 10.3758/s13423-010-0022-4

Dutilh, G., Krypotos, A.-M., & Wagenmakers, E.-J. (2011). Task-related versus stimulus-specific practice: A diffusion model account. *Experimental Psychology, 58*, 434–442. doi: 10.1027/1618-3169/a000111

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked vs. unmasked priming: Are they qualitatively diferent? *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0032333

Grasman, R. P., Wagenmakers, E.-J., & van der Maas, H. L. (2009). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology, 53*, 55–68.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science, 273*, 1699–1702. doi: 10.1126/science.273.5282.1699

Horn, S. S., Bayen, U. J., & Smith, R. E. (2011). What can the diffusion model tell us about prospective memory? *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 65*, 69–75. doi: 10.1037/a0022808

Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review, 117*, 759–784. doi: 10.1037/a0019471

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology, 93*, 353–368.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial introduction with R and BUGS*. Burlington, MA: Academic Press.

Leite, F. P. (2012). A comparison of two diffusion process models in accounting for payoff and stimulus frequency manipulations. *Attention, Perception, & Psychophysics, 74*, 1366–1382. doi: 10.3758/s13414-012-0321-0

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making, 6*, 651–687.

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS. A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325–337.

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language, 66*, 416–437. doi: 10.1016/j.jml.2011.12.001

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience, 32*, 2335–2343. doi: 10.1523/jneurosci.4156-11.2012

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

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review, 116*, 207–219. doi: 10.1037/a0014259

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

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. doi: 10.3758/bf03196283

Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging, 23*, 900–916. doi: 10.1037/a0013930

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review, 120*, 281–292. doi: 10.1037/a0030775

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model account of the Lexical Decision Task. *Psychological Review, 111*, 159–182.

Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Childers, R., Smith, P. L., & Segraves, M. A. (2011). Inhibition in superior colliculus neurons in a brightness discrimination task? *Neural Computation, 23*, 1790–1820. doi: 10.1162/NECO_a_00135

Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology, 97*, 1756–1774. doi: 10.1152/jn.00393.2006

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922.

Ratcliff, R., Perea, M., Colangelo, A., & Buchanan, L. (2004). A diffusion model account of normal and impaired readers. *Brain and Cognition, 55*, 374–382. doi: 10.1016/j.bandc.2004.02.051

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

Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General, 139*, 70–94. doi: 10.1037/a0018128

Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin and Review, 7*, 1–25.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A Diffusion Model analysis of the effects of aging in the Lexical-Decision Task. *Psychology and Aging, 19*, 278–289.

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

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics, 65*, 523–535. doi: 10.3758/bf03194580

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*, 408–424.

Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review, 13*, 626–635. doi: 10.3758/bf03193973

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology, 60*, 127–157.

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*, 464–487. doi: 10.1037/a0023810

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

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

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*, 414–429.

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 222–250. doi: 10.1037/a0026003

Spaniol, J., Madden, D. J., & Voss, A. (2006). A Diffusion Model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 101–117.

Spaniol, J., Voss, A., Bowen, H. J., & Grady, C. L. (2011). Motivational incentives modulate age differences in visual perception. *Psychology and Aging, 26*, 932–939. doi: 10.1037/a0023297

Spaniol, J., Voss, A., & Grady, C. L. (2008). Aging and emotional memory: Cognitive mechanisms underlying the positivity effect. *Psychology and Aging, 23*, 859–872.

Starns, J. J., Ratcliff, R., & White, C. N. (2012). Diffusion model drift rates can be influenced by decision processes: An analysis of the strength-based mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1137–1151. doi: 10.1037/a0028151

Starns, J. J., White, C. N., & Ratcliff, R. (2012). The strength-based mirror effect in subjective strength ratings: The evidence for differentiation can be produced without differentiation. *Memory & Cognition, 40*, 1189–1199. doi: 10.3758/s13421-012-0225-1

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

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

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Weasley.

van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: *Ez*, *fast-dm*, and *DMAT*. *Journal of Mathematical Psychology, 53*, 463–473. doi: 10.1016/j.jmp. 2009.09.004

Vandekerckhove, J., & Tuerlinckx, F. (2007a). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review, 14*, 1011–1026. doi: 10.3758/bf03193087

Vandekerckhove, J., & Tuerlinckx, F. (2007b). Fitting the Rateliff diffusion model to experimental data. *Psychonomic Bulletin & Review, 14*, 1011.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods, 40*, 61–72.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16*, 44–62. doi: 10.1037/a0021765

Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology, 44*, 1048–1056.

Voss, A., Rothermund, K., Gast, A., & Wentura, D. (2013). Cognitive processes in associative and categorical priming: A Diffusion Model analysis. *Journal of Experimental Psychology: General, 142*, 536–559. doi: 10.1037/a0029459

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*, 1206–1220.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39*, 767–775.

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology, 52*, 1–9.

Voss, A., Voss, J., & Klauer, K. C. (2010). Separating response tendency and decision biases: Arguments for an additional parameter in Ratcliff's diffusion model. *British Journal of Mathematical and Statistical Psychology, 63*, 539–555.

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*, 641–671.

Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review, 15*, 1229–1235. doi: 10.3758/PBR.15.6.1229

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*, 3–22. doi: 10.3758/BF03194023

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. M., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*, 615–631.

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology, 63*, 210–238. doi: 10.1016/j.cogpsych.2011.08.001

White, C. N., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition and Emotion, 23*, 181–205.

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*, 39–52.

Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York, NY: Oxford University Press.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 53–79. doi: 10.1037/a0024177

Yap, M. J., Balota, D. A., & Tan, S. E. (2013). Additive and interactive effects in semantic priming: Isolating lexical and decision processes in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 140–158. doi: 10.1037/a0028520

Zeguers, M. H. T., Snellings, P., Tijms, J., Weeda, W. D., Tamboer, P., Bexkens, A., & Huizenga, H. M. (2011). Specifying theories of developmental dyslexia: A diffusion model analysis of word recognition. *Developmental Science, 14*, 1340–1354. doi: 10.1111/j.1467-7687.2011.01091.x

Received December 21, 2012
Revision received April 4, 2013
Accepted April 5, 2013
Published online July 30, 2013

Andreas Voss

Psychologisches Institut
Ruprecht-Karls-Universität Heidelberg
Hauptstrasse 47-51
69117 Heidelberg
Germany
E-mail andreas.voss@psychologie.uni-heidelberg.de

