EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2009, 21 (5), 641–671

Psychology Press
Taylor & Francis Group



# Invited Address at the Occasion of the Bertelson Award 2007

# Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy

Eric-Jan Wagenmakers
*University of Amsterdam, Amsterdam, The Netherlands*

The Ratcliff diffusion model for simple two-choice decisions (e.g., Ratcliff, 1978; Ratcliff & McKoon, 2008) has two outstanding advantages. First, the model generally provides an excellent fit to the observed data (i.e., response accuracy and the shape of RT distributions, both for correct and error responses). Second, the parameters of the model can be mapped on to latent psychological processes such as the speed of information accumulation, response caution, and a priori bias. In recent years, the advantages of the Ratcliff diffusion model have become increasingly clear. Current advances in methodology allow all researchers to fit the diffusion model to data easily. Recent applications to ageing, lexical decision, IQ, practice, the implicit association test, and the accessory stimulus effect serve to highlight the added value of a diffusion model perspective on simple decision making.

***Keywords:*** Mathematical modelling; RT distributions; Signal-detection theory; Random walk.

For over a century, experimental psychologists have studied the complex structure of human cognition using relatively simple tasks. In many paradigms, the participant is confronted with a forced choice between two response alternatives. For instance, in a recognition memory experiment (Strong, 1912) the participant responds *old* or *new* to test stimuli; in a lexical decision experiment (Rubenstein, Garfield, & Millikan, 1970) the participant classifies letter strings as English *words* (e.g., TANGO) or *nonwords* (e.g.,

Correspondence should be addressed to Eric-Jan Wagenmakers, University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. E-mail: EJ.Wagenmakers@gmail.com

This paper is based on the Bertelson keynote lecture that was delivered at the 2007 ESCoP conference in Marseille, France.


http://www.psypress.com/ecp DOI: 10.1080/09541440802205067

DRAPA); and in an Eriksen flanker task (Eriksen & Eriksen, 1974) the participant indicates, say, whether a central target arrow among a set of distractor arrows (e.g., >>><>>>) points to the *left* or to the *right*.

In these and other two-alternative forced choice (2-AFC) tasks, participants are typically instructed to respond "as quickly and accurately as possible". Under these instructions, the observed variables are response time (RT) and response accuracy (i.e., correct/incorrect) for each trial. How should these data be analysed? Traditionally, inference is based on the mean response time for correct responses (MRT), and the proportion of correct responses (i.e., $P_c$). An experimental manipulation that increases MRT or decreases $P_c$ is thought to lower the efficiency of stimulus processing.

Despite its generality and its simplicity, the standard method of analysing data from 2-AFC tasks has several important limitations. First, the standard method ignores the shape of the RT distribution for the correct responses, focusing only on the mean, and ignores RTs for error responses altogether. Second, the standard method does not acknowledge the strong inverse relationship between response speed and response accuracy (i.e., the speed–accuracy tradeoff; Pachella, 1974; Schouten & Bekker, 1967; Wickelgren, 1977). Therefore, the standard methodology is unable to combine speed and accuracy into a single index for task difficulty or subject ability. Third, the standard method is not motivated by any substantive theory. This means that the results do not speak directly to the details of the underlying psychological processes. For instance, the fact that older adults are slower than younger adults could reflect a slowdown in decision making, a slowdown in the motor processes that are involved in pressing a response button, or both—the standard method cannot distinguish between these fundamentally different accounts.

This paper focuses exclusively on a cognitive process model that addresses all of these limitations, which plague the standard method for analysing data from 2-AFC tasks. The model was first studied in physics (e.g., Einstein, 1905) and was later implemented as a model for decision making in simple 2-AFC tasks (Ratcliff, 1978). The vanilla version of the model is widely known as the Wiener diffusion process, but here I will use the term "Ratcliff diffusion process" in acknowledgement of Roger Ratcliff's numerous attempts to modify the Wiener diffusion process and apply it to an impressive amount of phenomena.[1]

The Ratcliff diffusion model provides a detailed account of people's performance. Here, performance does not just refer to, say, mean RT for correct responses, but instead refers to proportion correct, RT distributions for correct responses, and RT distributions for error responses. These

[1] For a complete overview of Roger Ratcliff's work on the diffusion model see http://star.psy.ohio-state.edu/coglab/

measures of performance are used simultaneously to determine underlying psychological processes that are represented by model parameters. The four most important of these are the *speed of information accumulation*, *response caution*, *a priori bias*, and *nondecision time*. A statistical analysis in terms of these unobserved variables is immune to the speed–accuracy tradeoff and affords an unambiguous quantification of performance differences.

Recent years have witnessed a surge of research activity related to the Ratcliff diffusion model. This research activity has resulted in new and easy-to-use methods to fit the model to data, and has produced new insights from practical applications. The aim of this paper is to provide a selective overview of the recent practical developments in RT analysis using the Ratcliff diffusion model (for a recent, largely complementary review see Ratcliff & McKoon, 2008; for authoritative reviews on earlier developments see e.g., Luce, 1986; Townsend & Ashby, 1983).

The outline of this paper is as follows. The first section briefly outlines the Ratcliff diffusion model, and the second section summarises the pros and cons of a diffusion model analysis. The third and fourth sections summarise the current research efforts in diffusion model methodology and applications, respectively, and the fifth section concludes.



## THE RATCLIFF DIFFUSION MODEL

The diffusion model has been successfully applied to many two-choice RT paradigms, including lexical decision, short-term and long-term recognition memory tasks, same/different letter-string matching, numerosity judgements, visual-scanning tasks, brightness discrimination, and letter discrimination (e.g., Ratcliff, 1978, 1981, 2002; Ratcliff, Gomez, & McKoon, 2004a; Ratcliff & Rouder, 1998, 2000; Ratcliff, van Zandt, & McKoon, 1999; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). In all these applications, the diffusion model provided a close fit to response accuracy and the observed response time distributions for both correct and error responses. Early developments of the diffusion model are described in Edwards (1965), Laming (1968), Link (1992), Link and Heath (1975), Ratcliff (1978), and Stone (1960); a thorough account of the diffusion model is given by Luce (1986), Ratcliff (2002), Ratcliff and Smith (2004), and Townsend and Ashby (1983); details of the mathematics can be found in Gardiner (2004), Honerkamp (1994), and Smith (2000).

For concreteness, the Ratcliff diffusion model is introduced as it applies to a lexical decision task (e.g., Ratcliff et al., 2004a; Wagenmakers et al., 2008). When people have to decide quickly whether a letter string is a word (e.g., TANGO) or a nonword (e.g., DRAPA), the diffusion model assumes that people engage in a process of noisy information accumulation. That is, participants supposedly sample information sequentially, determine the

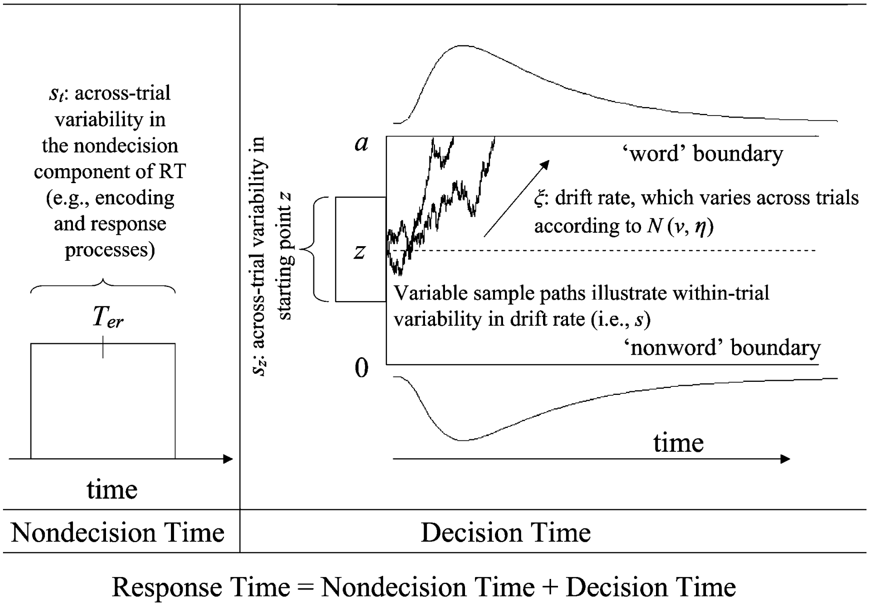


**Figure 1.** Diffusion model account of evidence accumulation in the lexical decision task (cf. Ratcliff, Gomez, & McKoon, 2004a).

extent to which each new sample of information supports the "word" versus "nonword" decision, and then add the new evidence to the old evidence. The accumulation process stops when the total evidence reaches some predetermined threshold. The process is illustrated in Figure 1. Note that, in contrast to signal-detection theory (Macmillan & Creelman, 2005), the participant is assumed to base a decision not on a single noisy sample, but on an entire sequence of noisy samples. From this perspective, the diffusion model may be thought of as a signal-detection theory for response times.

The diffusion model assumes that the signal-to-noise ratio of the information accumulation process is higher for stimuli that are easy to classify than it is for stimuli that are difficult to classify. In the diffusion model, "ease of processing" is quantified by a parameter called *drift rate*. When the absolute value of drift rate is high, decisions are fast and accurate; when the absolute value of drift rate is low, however, processing is driven to a large extent by noisy fluctuations, and as a result decisions are slow and inaccurate. In the lexical decision task, for example, classification performance for high-frequency words such as CHAIR is better than for low-frequency words such as FUME. The diffusion model accommodates this result through a change in drift rate: High-frequency words have a higher drift rate than low-frequency words. Drift rate reflects an inherent property of stimuli or participants, and, just like discriminability in signal-detection theory, it is generally not supposed to be under subjective control.[2]

More specifically, the information accumulation process is described by the following stochastic differential equation (e.g., Gardiner, 2004):

$$dX(t) = vdt + sdW(t), \tag{1}$$

where $dX(t)$ is the change in the accumulated evidence $X$ for a small time interval $dt$, $v$ is drift rate (i.e., the deterministic component of the noisy process), and $sdW(t)$ are zero-mean random increments with infinitesimal variance $s^2dt$. The factor $W(t)$ represents the Wiener noise process (i.e., idealised Brownian motion). Thus, the amplitude of the noise in the information accumulation process is governed by parameter $s$. This parameter is a scaling parameter, which means that if $s$ doubles, other parameters in the model can be doubled to obtain exactly the same result. Therefore, the choice of a specific value for $s > 0$ is arbitrary; for historical reasons, $s$ is usually fixed at 0.1.

In the diffusion model, the parameters that are under subjective control are "boundary separation" and "starting point". Both parameters are assumed to be determined by the participant before the start of each trial. Boundary separation quantifies response caution and modulates the speed–accuracy tradeoff: When the participant is careful not to make a mistake, the boundaries are set wide apart—as a result, the noisy fluctuations inherent in the accumulation of evidence are less likely to result in an incorrect response. The price that has to be paid for this decrease in error rate is an increase in response time. Thus, when boundary separation is large, decisions are slow and accurate; and when boundary separation is small, decisions are fast and inaccurate.

The other parameter that is under subjective control is starting point. Starting point reflects the a priori bias of a participant for one or the other response. This parameter is usually manipulated via payoff or proportion manipulations (Edwards, 1965; but see Diederich & Busemeyer, 2006). For instance, Wagenmakers et al. (2008, Exp. 2) used a proportion manipulation in which particular blocks of trials featured thrice as many words as nonwords. In the diffusion model, this manipulation causes the starting point to shift towards the "word" boundary. Such a shift would lead to relatively fast and accurate responding for word stimuli, but relatively slow and inaccurate responding for nonword stimuli.

The fourth key parameter of the diffusion model, $T_{er}$, quantifies the nondecision component of response time. The common interpretation of $T_{er}$

[2] To the best of my knowledge, the assertion that motivation or effort can increase drift rate has not been verified empirically.

is in terms of encoding and response processes, but this is too restrictive. For instance, suppose people have to judge whether a word represents an object that is bigger or smaller than a television. Also, suppose that each decision requires that people first construct a mental image of a television, then construct the mental image associated with the presented word, and finally engage in some sort of comparison process. The time that is taken up by the process of constructing the mental image of a television does not depend on the nature of the imperative stimulus, and is therefore clearly part of $T_{er}$. As is illustrated in Figure 1, the diffusion model assumes that the observed RT is the sum of the nondecision component and the decision component of processing (Luce, 1986):

$$RT = DT + T_{er}, \tag{2}$$

where $DT$ denotes decision time. Therefore, the nondecision time $T_{er}$ does not affect response choice and acts solely to shift the entire RT distribution by a constant amount.

Thus, the four key parameters of the diffusion model are drift rate $v$ (i.e., speed of information accumulation), boundary separation $a$ (i.e., response caution), starting point $z$ (i.e., a priori bias), and nondecision time $T_{er}$. Unfortunately, these parameters alone do not allow the model to capture all of the robust empirical phenomena that have been discovered in the RT literature, and additional parameters were introduced to address the deficiencies. In particular, uniformly distributed trial-to-trial variability in nondecision time, called $s_t$, was introduced to account for the relatively gradual rise in the leading edge of the RT distribution (Ratcliff & Tuerlinckx, 2002); uniformly distributed trial-to-trial variability in starting point, called $s_z$, was introduced to account for error responses that are systematically faster than correct responses (Laming, 1968); and normally distributed trial-to-trial variability in drift rate, called $\eta$, was introduced to account for error responses that are systematically slower than correct responses (e.g., Ratcliff, 1978; see also Ratcliff & Rouder, 1998).[3]

In sum, the parameters of the Ratcliff diffusion model are the following:

1. Mean drift rate ($v$).
2. Across-trial variability in drift rate ($\eta$).
3. Boundary separation ($a$).
4. Mean starting point ($z$).
5. Across-trial range in starting point ($s_z$).
6. Mean of the nondecision component of processing ($T_{er}$).

[3] The combination of $s_z$ and $\eta$ does not allow the model to fit slow and fast errors at will, at least not when an experiment features multiple conditions across which and $S_z$ $\eta$ are fixed. See Wagenmakers et al. (2008) for an illustration.

7. Across-trial range in the nondecision component of processing ($s_t$).

Ratcliff and Tuerlinckx (2002) proposed to estimate a mixture model, where one mixture component is associated with the diffusion model, and the other mixture component (i.e., a uniform distribution that spans the RT range) is associated with "response contaminants". This proposal adds a mixture parameter $p$ to the previous list; however, this parameter is usually not of immediate interest and it is as yet unclear whether or not it can be systematically manipulated in an experiment.

## PROS AND CONS OF THE RATCLIFF DIFFUSION MODEL

The Ratcliff diffusion model offers several advantages over a traditional analysis of RT and accuracy. First, the model provides a principled account of how RT relates to accuracy, and is able to account for changes in both dependent measures simultaneously. Second, the model requires a consideration of not only mean RT for correct responses, but of the entire RT distribution, both for correct and incorrect responses. Third, the model allows researchers to estimate unobserved psychological processes that together determine observed performance. Fourth, the model can be used to generate new insights for established problems. Specifically, a diffusion model analysis often demonstrates that a certain general phenomenon (i.e., linear Brinley plots in ageing, the worst performance rule in intelligence research) can be produced by the diffusion model through a simple change in one of the key parameters such as drift rate. Fifth, the diffusion model can act as a theoretical tool to discover new empirical regularities. For instance, Wagenmakers, Grasman, and Molenaar (2005) showed that the diffusion model predicts that an increase in task difficulty should increase RT mean and RT standard deviation at the same rate; a prediction for which Wagenmakers and Brown (2007) offer empirical support. Later sections of this paper will exemplify these claims in more concrete terms.

The Ratcliff diffusion model also comes with a few disadvantages. The most important disadvantage is that, for the uninitiated, the model used to be exceedingly difficult to apply to data; this was mainly due to the lack of publicly available, user-friendly computer programs for parameter estimation. As I will detail, recent methodological progress have now made it easy for experimental psychologists to apply the model to data without any outside help. The second disadvantage is that the model requires a fair amount of data for accurate estimation of its parameters—the reason for this is that the model requires an estimation of the RT distribution for error responses. When response accuracy is at 95%, as it typically is, then it may take as many as 200 trials to get a satisfactory estimate of the RT distribution for error responses. This disadvantage may be mitigated by including experimental manipulations that selectively affect a particular parameter. When this particular parameter is the only one that is free to vary, and when at least one of the experimental conditions features a relatively high error rate, the rule-of-thumb requirement of about 10 error responses no longer holds (Ratcliff, in press). The most promising solution to the sample size problem, however, appears to come from a Bayesian hierarchical modelling effort (Lee, Vandekerckhove, Navarro, & Tuerlinckx, 2007; Vandekerckhove, Tuerlinckx, & Lee, in press), which is discussed in more detail later.

The third disadvantage of the diffusion model is that it is only intended for two-choice response time tasks in which processing is approximately "one-shot", and RTs are mostly faster than about 1.5 s. This means that it may not be warranted to apply the diffusion model to, say, the Stroop task, in which there is a competition between prepotent processes (i.e., reading of the coloured word) and the more deliberate processes (i.e., verbalisation of the colour in which the word is printed). However, from a pragmatic standpoint the limitation to relatively fast one-shot processes is merely a guideline; the adequacy of the model can be rigorously assessed using statistical methodology, regardless of the nature of the hypothesised psychological processes.

The fourth disadvantage of the diffusion model is that it does not speak directly to the details of the neural substrate that underlies the decision-making process. In contrast, neurocomputational accounts such as those proposed by Brown, Bullock, and Grossberg (2004), Frank (2006), Lo and Wang (2006) do use results from neuroanatomy to inform their modelling. Thus, the psychological concepts from the diffusion model—drift rate, boundary separation—remain at a relatively high level of abstraction. However, the disadvantage of a high level of abstraction comes with the advantage of generalisability; the diffusion model applies to a wide range of different tasks, precisely because it does not commit to particular brain structures that are differentially involved in, say, the execution of saccades versus the execution of finger movements. Moreover, recent work has started to connect the diffusion model to a range of findings from neuroscience, mainly through the analysis of single cell recordings in monkeys, but also through mathematical modelling (e.g., Bogacz & Gurney, 2007; for reviews see Gold & Shadlen, 2007; Ratcliff & McKoon, 2008).

The fifth disadvantage of the diffusion model is that it has at least seven parameters, a number that increases when the experiment features more than one condition. This may make some people feel that the diffusion model has too many parameters; in the spirit of John von Neumann, who famously said "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk",[4] one could argue that with seven or more parameters it is possible to fit just about anything, and that consequently a good fit to the data would mean just about nothing.[5] The validity of this criticism depends to some extend on the specific data set to which the diffusion model is applied, but several arguments suggest that the problem of overparameterisation is easily overstated.

The first argument against the claim that the diffusion model is overparameterised depends on the experimental design: When an experiment features, say, three levels of task difficulty and two levels of speed stress, then the only parameter free to vary across conditions are those that are believed to be affected by the experimental manipulations; hence, drift rate is the only parameter that is free to vary across task difficulty, and boundary separation is the only parameter that is free to vary across speed stress. This means that with only two parameters free to vary, the model would account for the pronounced effects of the experimental manipulations both on response accuracy and on the shape of the RT distributions; this is quite an achievement, and one that is hard to match for competing models (e.g., Ratcliff & Smith, 2004; Wagenmakers et al., 2008). Second, the number of model parameters is not an absolute indication of model complexity. Contrary to what is suggested by von Neumann, the extent to which a model has too many parameters depends on the amount of data that the model attempt to explain. The diffusion model may have at least seven free parameters, but it also accounts for the shape of RT distributions, both for correct and for incorrect responses. Thus, the issue of deciding whether a model has too many parameters is not solved by simply counting the number of free parameters. Instead, statistical techniques (e.g., model selection; Myung, Forster, & Browne, 2000; Wagenmakers & Waldorp, 2006) can be used to determine the model that strikes the optimal balance between parsimony and goodness-of-fit. In other words, the claim that the diffusion model has too many parameters can be put to a statistical test.

The third argument against the claim that the diffusion model is overparameterised is that there are plausible patterns of results that the diffusion model fails to fit (Ratcliff, 2002). This happens because the diffusion model makes qualitative predictions that hold regardless of specific parameter values. For instance, the model predicts that RT distributions are skewed to the right, and that this skew increases with task difficulty. The

[4] http://en.wikiquote.org/wiki/John_von_Neumann

[5] An anonymous reviewer argued that a sixth disadvantage of the diffusion model is that its parameter estimates are positively correlated. Ratcliff and Tuerlinckx (2002, pp. 452–455) discuss this issue in detail. On the one hand, it is true that the parameter estimates of the diffusion model tend to be correlated, but on the other hand one should acknowledge that the same holds for many common statistical models (e.g., estimates of slope and intercept in a linear regression).

model also predicts that, when people are instructed to be less cautious, error responses speed up more than correct responses (e.g., Wagenmakers et al., 2008). Fourth, the parameters of the diffusion model respond selectively to specific manipulations. For instance, Voss, Rothermund, and Voss (2004) have shown that changes in task difficulty, speed–accuracy instructions, payoff structure, and ease of executing a motor response bring about the expected changes in drift rate, boundary separation, starting point, and nondecision time, respectively. Such a successful test of selective influence provides strong support for the hypothesised mapping between psychological processes and model parameters (for work that takes advantage of this mapping see Oberauer, 2005; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). Moreover, a model that would have too many parameters might be expected to use these excess parameters in order to provide a better fit to the data (i.e., in order to make the data "wiggle its trunk"). In the experiments by Voss et al. (2004), and many others discussed later, this was not the case. The final argument against the claim that the diffusion model is overparameterised is that the goal of using the diffusion model is often not to obtain an excellent fit to the data; instead, interest usually centres on the estimation of the latent psychological processes associated with drift rate, boundary separation, a priori bias, and nondecision time.

In sum, the diffusion model offers clear advantages for experimental psychologists who wish to analyse data from two-alternative forced choice tasks. Adoption of the diffusion model appears to come with some disadvantages, but these are to a large extent illusory.

## CURRENT METHODOLOGICAL DEVELOPMENTS

Up until a year ago, experimental psychologists who wanted to apply the diffusion model to data without expert help were confronted with a formidable challenge, namely to program a set of computer routines to carry out parameter estimation for the diffusion model. To illustrate the technical difficulties involved, consider the basic equations that need to be computed for the program to work.

Assume, first, that none of the diffusion model parameters are allowed to vary from trial to trial. In this simplified case, the probability of an error $P_e$ is given by

$$P_e = 1 - P_c = \frac{\exp\left(-\frac{2av}{s^2}\right) - \exp\left(-\frac{2zv}{s^2}\right)}{\exp\left(-\frac{2av}{s^2}\right) - 1}, \tag{3}$$

where drift rate $v$, boundary separation $a$, starting point $z$, and nondecision time $T_{er}$ are free parameters, and $s$ is the scaling parameter from Equation 1. The equation that gives the probability of an error response before time $t$ is given by (Cox & Miller, 1970)

$$G_{X,T} = P_e - \frac{\pi s^2}{a^2} \exp\left(\frac{-zv}{s^2}\right) \sum_{k=1}^{\infty} \frac{2k \sin\left(\frac{\pi k z}{a}\right) \exp\left[-\frac{1}{2}\left(\frac{v^2}{s^2} + \frac{\pi^2 k^2 s^2}{a^2}\right)(t - T_{er})\right]}{\left(\frac{v^2}{s^2} + \frac{\pi^2 k^2 s^2}{a^2}\right)}, \tag{4}$$

where $k$ indexes the infinite series. The probability of a correct response before time $t$ is obtained by replacing $z$ and $v$ by $a\text{–}z$ and $-v$, respectively.

The first complication is that Equation 4 contains an infinite sum (i.e., the $\Sigma_{k=1}^{\infty}$ part) that contains an oscillating series. Therefore, Equation 4 has to be approximated by a finite partial sum, which is truncated when the absolute values of the last two consecutive terms are smaller than some constant (Tuerlinckx, 2004, p. 703). An alternative procedure is to determine the number of terms that are needed to achieve a specific accuracy of approximation (Voss et al., 2004, p. 1218). Despite this complication, this model is relatively easy to program and estimate. However, in the full Ratcliff diffusion model, parameters $v$, $z$, and $T_{er}$ are allowed to vary from trial to trial. Consequently, Equation 4 needs to be integrated over these three sources of variability, and the equation that gives the probability of an error response before time $t$ is given by (Tuerlinckx, 2004)

$$\int_{-\infty}^{\infty} \int_{z-\frac{s_z}{2}}^{z+\frac{s_z}{2}} \int_{Ter-\frac{s_t}{2}}^{\min\left(Ter+\frac{s_t}{2},t\right)} G_{X,T} U\left(T_{er} - \frac{s_t}{2}, Ter + \frac{s_t}{2}\right) U\left(z - \frac{s_z}{2}, z + \frac{s_z}{2}\right) N(u, \eta^2) dT_{er} dz dv, \tag{5}$$

where $U$ denotes the uniform distribution, and $N$ the normal distribution. Evaluation of Equation 5 requires numerical integration techniques, and at this point most experimental psychologists would decide that the expected payoff from the model is not worth the expected effort associated with its implementation.

In the last year, however, this pain-to-gain ratio has been dramatically lowered by the appearance of several freely available computer programs for diffusion model analyses. The goal of these computer programs is to allow the nonexpert user to estimate parameters of the Ratcliff diffusion model. The next sections provide an overview of these and other recent methodological advances.

## EZ and EZ2

The EZ-diffusion model (Wagenmakers, van der Maas, & Grasman, 2007) was developed in order to make it as easy as possible for experimental psychologists to apply the diffusion model to their data. The EZ-diffusion model consists of a set of three equations that transform $P_c$ (i.e., the proportion of correct responses), MRT (i.e., the mean RT for correct responses), and VRT (i.e., the variance of RT for correct responses) into estimates for drift rate $v$, boundary separation $a$, and nondecision time $T_{er}$. In order to accomplish this transformation, the EZ-diffusion model makes the simplifying assumptions that the starting point is unbiased (i.e., $z = a/2$) and the across-trial variabilities $\eta$, $s_z$, and $s_t$ are all zero. The EZ-diffusion model is shown in Figure 2.

Under these simplifying "EZ" assumptions, the first equation gives the probability correct as

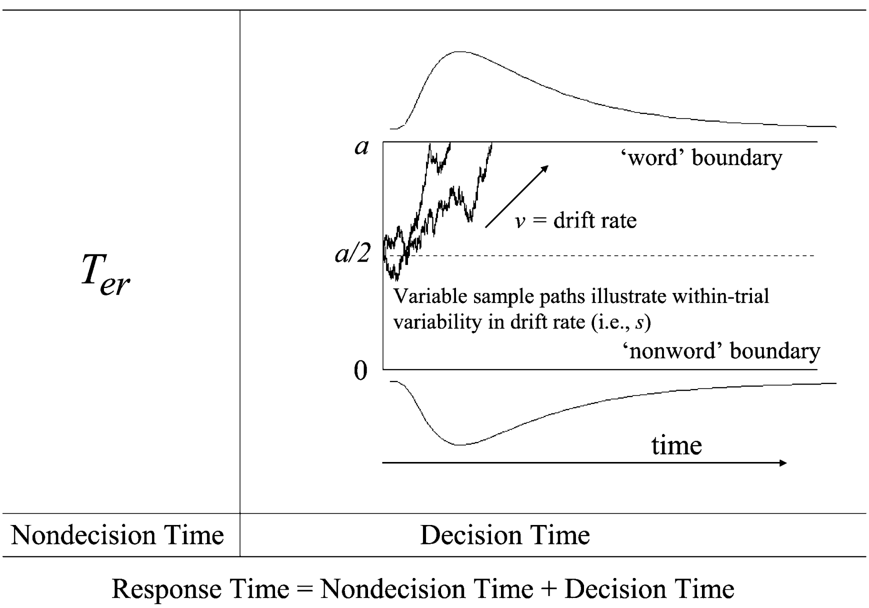


**Figure 2.** The EZ-diffusion model.

$$P_c = \frac{1}{1 + \exp(-av/s^2)}, \tag{6}$$

where $s$ is again the irrelevant scaling parameter. The second equation refers the RT variance of a diffusion process with unbiased starting point (Wagenmakers et al., 2005), and is given by

$$\mathrm{VRT} = \left[\frac{as^2}{2v^3}\right] \frac{2y\exp(y) - \exp(2y) + 1}{(\exp(y) + 1)^2}, \tag{7}$$

where $y = -va/s^2$ and $v \neq 0$. If $v = 0$, $\mathrm{VRT} = \dfrac{a^4}{24s^4}$.

From these two equations it is straightforward to extract estimates for $v$ and $a$. After $a$ and $v$ have been determined from Equations 6 and 7, $T_{er}$ can then be computed from the third equation, which incorporates the mean time until arrival at a response threshold (i.e., MDT, mean decision time):

$$\begin{aligned} MRT &= MDT + T_{er} \\ &= \left(\frac{a}{2v}\right)\frac{1 - \exp(y)}{1 + \exp(y)} + T_{er}, \end{aligned} \tag{8}$$

where, again, $y = -va/s^2$. Thus, the EZ-diffusion model has three equations that feature the observed quantities $P_c$, VRT, and MRT, and from these one can uniquely determine values for the model parameters $v$, $a$, and $T_{er}$. None of this requires any parameter fitting; all that is needed to determine the parameters is a straightforward computation. Hence, the EZ-diffusion model is straightforward to implement in JavaScript, R, or Excel.[6]

As shown by Figure 3, the EZ-diffusion model is conceptually similar to signal-detection theory. Both procedures involve a simple transformation from observed behaviour to unobserved processes, and both procedures differentiate between processes that quantify task difficulty or subject ability (i.e., discriminability for signal-detection theory, and drift rate for the EZ-diffusion model) and processes that are under control of the participant (i.e., bias for signal-detection theory, and boundary separation for the EZ-diffusion model).

Despite the obvious practical advantages of the EZ-diffusion model, some researchers may believe that the model is perhaps *too* easy. An in-depth discussion of the pros and cons of the EZ-diffusion model will take place elsewhere (e.g., Ratcliff, in press). It should be kept in mind, however, that the goal of the EZ-diffusion model is not to replace the more complete diffusion model analysis; rather, the goal of the EZ-diffusion methodology is to provide an admittedly rough first estimate of the underlying processes,

[6] Code is freely available from http://users.fmg.uva.nl/ewagenmakers/EZ.html, http://users.fmg.uva.nl/ewagenmakers/2007/EZ.R, and http://users.fmg.uva.nl/ewagenmakers/2007/EZ.xls, respectively.

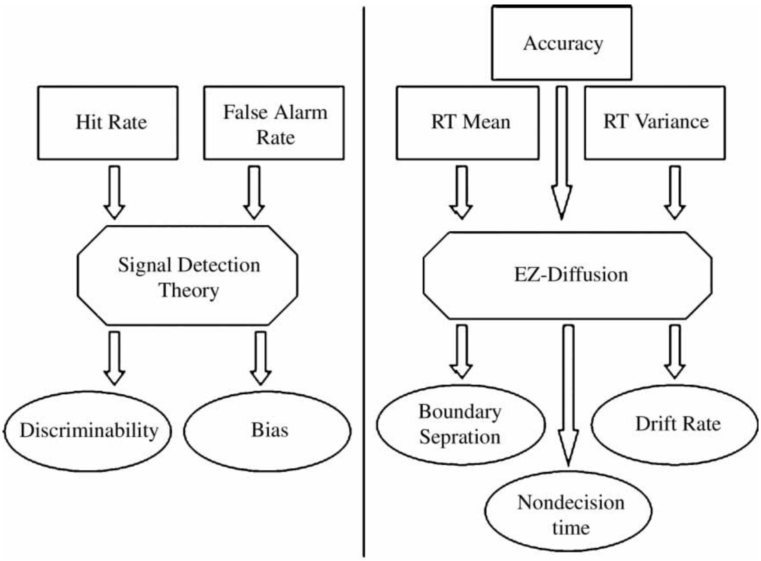


**Figure 3.** Schematic representation of the similarity between a signal-detection analysis and an EZ-diffusion model analysis. The circles at the bottom denote unobserved variables, and the squares at the top denote observed variables.

which should, if possible, be later augmented with a more complete analysis. Also, the simplifying assumptions of the EZ-diffusion model can be verified by statistical tests (Wagenmakers et al., 2007).

In an attempt to generalise and improve on the original EZ-diffusion model, Grasman, Wagenmakers, and van der Maas (2008) recently proposed the EZ2 model. In the EZ2 model, the starting point is allowed to vary freely, and parameters can be constrained across conditions. The key equation of the EZ2 model (which is of course hidden from the user) gives the variance of the decision time distribution for diffusion process with a priori bias, conditional on the boundary that was reached first. A software routine that implements the EZ2 program is freely available online[7]; Figure 4 shows a screenshot of the EZ2 web application (courtesy of Raoul Grasman).

The EZ2 model takes advantage from the fact that in a two-alternative RT task, there are two stimulus categories that may have different drift rates but share values for decision criteria. Thus, for words and nonwords in a lexical decision task, the EZ2 model takes as input $P_c^{words}$, $P_c^{nonwords}$,

[7] http://users.fmg.uva.nl/rgrasman/jscript/EZ2_new.html; this web application works fastest under Safari or Microsoft Internet Explorer Version 7. Excel code and R code available from Raoul Grasman upon request.


**Figure 4.** Screenshot of the EZ2 web application.

$VRT^{words}$, $VRT^{nonwords}$, $MRT^{words}$, and $MRT^{nonwords}$, and returns as output estimates for $v^{words}$, $v^{nonwords}$, $a$, $z$, $T_{er}^{words}$, and $T_{er}^{nonwords}$. Note that $a$ and $z$ are response criteria that are assumed to be determined prior to stimulus processing, so that they are independent of whether the stimulus is a word or a nonword. If desired, the EZ2 program can determine a common estimate for $T_{er}$ using a least-squares fitting procedure.

## DMAT

The Diffusion Model Analysis Toolbox (DMAT) is a Matlab application developed by Joachim Vandekerckhove and Francis Tuerlinckx (Vandekerckhove & Tuerlinckx, 2007, 2008).[8] The DMAT program is able to estimate the full Ratcliff diffusion model, including the trial-to-trial variability parameters $\eta$, $s_z$, and $s_t$. By default, the parameter estimation routine first computes the EZ-diffusion model estimates to provide useful starting values, and then uses multinomial maximum likelihood to determine the best-fitting parameter values.

The DMAT program comes with many options, such as the possibility to simulate data from the model, the possibility to constrain parameters through easy-to-use design matrices, the possibility to do model selection based on indices such as BIC (e.g., Raftery, 1995; Wagenmakers, 2007), the possibility to filter out uniformly distributed contaminants using mixture modelling, and the possibility to automatically determine the optimal cutoff

[8] DMAT is freely available online at http://ppw.kuleuven.be/okp/dmatoolbox

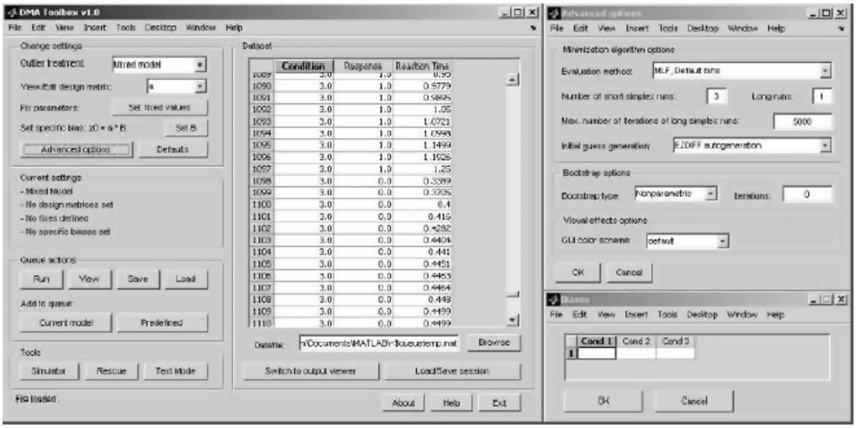

**Figure 5.** Screenshot of DMAT's graphical user interface.

point for the elimination of fast guesses. Last but not least, the DMAT program comes with a user-friendly graphical user interface (GUI), a screenshot of which is shown in Figure 5 (courtesy of Joachim Vandekerckhove). The GUI generates syntax that a more experienced user may choose to modify for batch processing (i.e., fitting data from multiple participants one after another).

The DMAT GUI, the clear documentation, and the excellent online support make the program so easy to use that—at least at the methodology unit of the University of Amsterdam—graduate and even undergraduate students can fit data without outside help.

The main disadvantage of this truly excellent toolbox is that it does require Matlab, a relatively expensive computer program. Also, a user without any programming experience may find it somewhat difficult to adjust the Matlab syntax by hand, an activity that is not strictly necessary but can often be very useful.

## Fast-dm

The fast-dm program is a platform-independent command line tool developed by Andreas and Jochen Voss (Voss & Voss, 2007, 2008).[9] Just as the DMAT program, fast-dm is able to estimate all of the Ratcliff diffusion model parameters, including $\eta$, $s_z$, and $s_t$. Fast-dm's parameter estimation routine first computes the EZ-diffusion model estimates to

[9] fast-DM is freely available online at http://www.psychologie.uni-freiburg.de/Members/voss/fast-dm

provide useful starting values, and then determines the best-fitting parameter values by minimising the Kolmogorov-Smirnov statistic, that is, the maximal vertical distance between the predicted and the empirical cumulative RT distributions (Voss et al., 2004). It should further be pointed out that the fast-dm program does not use the analytical solution of the partial differential equation (PDE) that defines the diffusion process (i.e., Equation 4 with its infinite sum); instead, fast-dm uses a numerical solution of the PDE, and Voss and Voss (2008) show that this numerical PDE method can greatly lower the computational costs of parameter optimisation.

In fast-dm, the program settings are specified in a control file. In this control file, the user can indicate the input files that contain the data, the format of the data, the required accuracy of the predicted RT distributions, and the parameters that are either fixed or free to vary across specific conditions. The output is provided on the screen and is written to text files. The fast-dm program is easy to use and comes with a good manual (i.e., Voss & Voss, 2007).

The fast-dm program currently offers fewer modelling options than DMAT, but then again fast-dm does not require that the user owns a copy of the relatively expensive Matlab program. Both DMAT and fast-dm are excellent, user-friendly programs that have made it easy for experimental psychologists to fit the full Ratcliff diffusion model to their data.

## Bayesian estimation

A age-old problem in psychology is how to deal with individual differences. In contrast to the field of psychometrics, where individual differences are the main topic of investigation, the field of experimental psychology has traditionally ignored individual differences, pretending instead that each new participant is a replicate of the previous one (Batchelder, 2007) As Bill Estes and others have shown, however, individual differences that are ignored can lead to so-called *averaging artifacts* in which the data that are averaged over participants are no longer representative for any of the participants (Estes, 1956, 2002; Heathcote, Brown, & Mewhort, 2000). One way to address this issue, popular in psychophysics, is to measure each individual participant extensively, and deal with the data on a participant-by-participant basis.

In between the two extremes of assuming that participants are completely the same and that they are completely different lies the compromise of hierarchical modelling. In hierarchical modelling, individual-level parameters are drawn from a group distribution—hence, the hierarchical model takes both differences and similarities between participants into account. In a diffusion model context, hierarchical modelling allows variability across participants and items to be accounted for by higher order distributions that inform and constrain the estimation of lower level parameters for individual participants or items.

In the field of psychology, Jeff Rouder and colleagues have repeatedly illustrated the theoretical advantages and practical relevance of a *Bayesian* hierarchical analysis of common experimental data (Rouder & Lu, 2005; Rouder, Lu, Speckman, Sun, & Jiang, 2005; Rouder et al., 2007). Although hierarchical analyses can be carried out using orthodox methodology (i.e., Hoffman & Rovine, 2007), there are strong philosophical and practical reasons to prefer the Bayesian methodology (e.g., Lindley, 2000; Gelman & Hill, 2007, respectively). In the Bayesian methodology, uncertainty is represented by probability distributions, and probability theory is used to update prior knowledge of the parameters in light of the observed data (e.g., Lee & Wagenmakers, 2005; Wagenmakers, 2007).

Recently, Michael Lee and colleagues implemented two Bayesian (hierarchical) estimation routines for diffusion model analyses (Lee, Fuss, & Navarro, 2006; Lee et al., 2007). The work by Lee et al. (2006) depends on an analytic approximation to Equation 4 that does not contain the infinite sum, and the work by Lee et al. (2007) rests on the implementation of Equation 4 for use in the popular WinBUGS program (e.g., Spiegelhalter, Thomas, Best, & Lunn, 2003). The Bayesian estimation routine from Lee et al. (2007) is not yet publicly available, but this is expected to happen in the near future.

The Bayesian estimation of diffusion model parameters constitutes an exciting new development, and it offers several advantages:

1. In the Bayesian framework, it is easy to carry out hierarchical analyses. As mentioned earlier, hierarchical analyses take into account differences and similarities between participants and between items. Moreover, in the case of few data, extreme individual estimates will shrink towards the mean of the group distribution.
2. Prior distributions for the parameters can be based on estimates from previous experiments. The use of such informative priors will lead to more sensible results.
3. Uncertainty with respect to parameters is directly reflected in the posterior distribution of the parameters. The knowledge gained from the experiment is quantified by the difference between the prior distribution and the posterior distribution. Combined with the first two advantages, this means that posterior distributions can be calculated even when some participants or items do not produce many error responses. This addresses an important limitation of current non-Bayesian methods.

4. The adequacy of the model can be quantified by generating synthetic data from the model, and comparing these data to those that were actually observed (i.e., so-called posterior predictive checks, e.g., Meng, 1994). When data are generated from a Bayesian model, uncertainty with respect to all parameters in the model is automatically and properly taken into account.
5. In many situations, researchers might want to select between several competing diffusion models, for instance one in which an experimental effect is in drift rate versus another one in which the effect is in boundary separation. In the Bayesian paradigm, model selection is accomplished using Bayes factors, a number that quantifies the change from prior to posterior odds (Jeffreys, 1961; Kass & Raftery, 1995). Such a Bayesian hypothesis test has many advantages over the orthodox likelihood ratio test and their associated *p*-values (Wagenmakers, 2007).
6. Using the power of graphical modelling, the Bayesian framework can easily be adjusted to account for a host of interesting findings (e.g., Shiffrin, Lee, Wagenmakers, & Kim, in press).

In sum, the methodological developments that are presented here range from the very simple (i.e., the EZ-diffusion model) to the complex (i.e., hierarchical Bayesian estimation). Programs such as DMAT and fast-dm have made the Ratcliff diffusion model accessible to a broad audience of experimental psychologists, and with these programs comes the hope that the Ratcliff diffusion model will see many more applications in the near future.

## CURRENT EMPIRICAL DEVELOPMENTS

The goal of this section is to provide a brief overview of recent diffusion model applications. These applications span a wide range of phenomena, and they are intended to show the added value of decomposing performance into the constituent psychological processes that are hypothesised by the diffusion model. In many of the applications, a diffusion model analysis provides a serious challenge to entrenched verbal theories and explanations.

### Phenomenon 1: Ageing

One of the most popular explanations for the cognitive effects of ageing is provided by the "general slowing" hypothesis. According to this hypothesis, the main effect of ageing is to slow down all cognitive processes by the same rate (e.g., Brinley, 1965; Cerella, 1985; Salthouse, 1996). One source of support for the general slowing hypothesis comes from the analysis of Brinley plots, in which averaged mean RT of a group of young participants is plotted against the averaged mean RT of a group of older participants, across a range of different conditions. The Brinley plot is often linear with a slope greater than 1, which supposedly indicates that older people have a slower rate of processing than younger people.

In several articles, Roger Ratcliff, Dan Spieler, and Gail McKoon have demonstrated that Brinley plots are potentially misleading (Ratcliff, Spieler, & McKoon, 2000, 2004). The details of their arguments and the ensuing discussion are not important here. What is important is that the "global slowing" hypothesis states that the main age-related change in processing is in the rate of information accumulation, that is, in drift rate. This is a hypothesis that can be tested by applying the diffusion model to empirical data. Note that this test involves not just an analysis of group-averaged mean RTs, but an analysis of proportion correct, RT distributions for correct responses, and RT distributions for incorrect responses.

In recent years, the diffusion model has been applied to ageing across a wide range of tasks (e.g., Ratcliff, Thapar, Gomez, & McKoon, 2004c; Ratcliff, Thapar, & McKoon, 2001, 2003, 2004d, 2006a, 2007; Spaniol, Madden, & Voss, 2006; Thapar, Ratcliff, & McKoon, 2003). The overall results show that the effects of ageing cannot be attributed to a change in a single process. The most reliable age-related changes are an increase in nondecision time, and an increase in boundary separation (i.e., older people may take longer to execute a motor response, and they are more careful not to make errors). Depending on the task, age-related changes in drift rate may or may not be observed. These results are based on a detailed analysis of the data and appear to have convincingly falsified the global slowing hypothesis of ageing.



## Phenomenon 2: Lexical decision

In the lexical decision task, participants are required to quickly decide whether a visually presented letter string is a word (e.g., MANGO) or a nonword (e.g., DROPA). The lexical decision task is one of the most often used tasks in the field of visual word recognition, and several models have been proposed to account for its key findings (e.g., Norris, 2006; Plaut, 1997; Wagenmakers et al., 2004).

One of the most influential models for lexical decision is the deadline model (i.e., MROM, the Multiple Read-Out Model, Grainger & Jacobs, 1996; or DRC, the Dual Route Cascaded model of visual word recognition and reading aloud, Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). The characteristic feature of the deadline model is that it makes "nonword" decisions *by default*; that is, the system is assumed to accumulate evidence for the hypothesis that the presented letter string is a word. If this accumulation process has not reached threshold levels before some temporal deadline $T$, the system responds "nonword" (for details see Wagenmakers et al., 2008).

The deadline model for lexical decision is radically different from the diffusion model for lexical decision (Ratcliff et al., 2004a; Yap, Balota, Cortese, & Watson, 2006). In the diffusion model, there is no qualitative distinction between "word" and "nonword" responses; a "nonword" response is not based on a temporal deadline but on accumulated evidence. In an attempt to distinguish between the models, Wagenmakers et al. (2008) conducted two experiments and fit both the diffusion model and the MROM to the data. The results showed that the diffusion model accounted nicely for the experimental data, which included effects of word frequency, speed–accuracy stress, and word–nonword proportion. In contrast, the deadline model could not handle the experimental data. In particular, the deadline model failed to account for the experimental conditions in which "nonword" responses were systematically faster than "word" responses. The reason for the failure of the deadline model is in the deadline mechanism; when "nonword" responses can only be given after the "word" response has timed out, "nonword" response are necessarily slow.[10]

## Phenomenon 3: IQ

People with a high IQ tend to respond faster than people with a low IQ, a regularity that is reported across a wide range of cognitive tasks (Jensen, 1998, 2006). Somewhat surprisingly, the effect of IQ on response speed is observed even in the simplest tasks, such as when people have to decide which of two lines, presented side by side, is the longest. Moreover, this statistical association with IQ is more pronounced for the slowest responses than it is for the fastest responses, an effect that Larson and Alderton (1990) coined the "worst performance rule" (WPR). Figure 6 shows data from Larson and Alderton's seminal study. These and other data show that people's slowest responses are the best predictors of their intelligence.

The origin of the WPR is subject to debate (see Coyle, 2003, for a review). Although several verbal theories have been proposed to explain the WPR, almost no quantitative models have tried to account for the phenomenon. Recently, Ratcliff, Schmiedek, and McKoon (2008) showed that, in theory at

[10] Simulations with the deadline model indicated that the model is able to generate fast "nonword" responses, but that this leads to unacceptably large error rates.

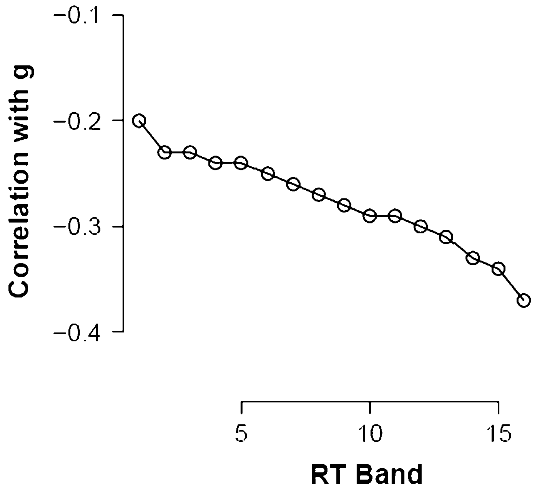


**Figure 6.** The worst performance rule: The association between response speed and intelligence is most pronounced for slow RTs. The figure is based on data reported in Larson and Alderton (1990, Table 4, p. 317).

least, the WPR is automatically generated by a diffusion model in which individual differences in IQ map on to individual differences in drift rate or boundary separation.[11] The driving force behind the diffusion model account of the WPR is the fact that RT distributions skew out when drift rate decreases or boundary separation increases—that is, the difference between two RT distributions is larger for the slow part of the distributions than it is for the fast part.

In order to test whether the diffusion model account of the WPR holds up in practice, Weeda, Wagenmakers, and Huizenga (2007) applied the model to data from a perceptual discrimination experiment with 44 high-school students. A median split on their Raven scores divided the students in a high IQ group and a low IQ group. Contrary to our expectation, the behavioural results did not indicate the presence of the WPR—performance of the two IQ groups did not differ significantly with respect to proportion correct or mean RT. A diffusion model analysis, however, showed that the high IQ participants had a higher drift rate and a lower boundary separation than did the low IQ participants, consistent with the theoretical account by Ratcliff et al. (2008). In addition, however, the high IQ participants had a

[11] In order to produce the WPR, the diffusion model also requires that the nondecision time $T_{er}$ varies randomly from participant to participant (Ratcliff et al., 2008).

longer nondecision time $T_{er}$ than the low IQ participants, an unexplained result that effectively masked the WPR in the behavioural measures.

In sum, the diffusion model provides an account of IQ-related differences in response times that is much more detailed than the one provided by standard analysis. The diffusion model explains the WPR in terms of individual differences in drift rate or boundary separation. Weeda et al. (2007) showed that the IQ-related changes in these key parameters can be obtained even when changes in other parameters mask the WPR in the behavioural measures.

## Phenomenon 4: Practice

The effect of practice is ubiquitous; in almost every cognitive task, whether it involves cigar rolling, card sorting, pencil mazes, or alphabet arithmetic, performance increases as a function of training time (e.g., Crossman, 1959; Logan, 1992). One of the unresolved issues in experimental psychology concerns the shape of the learning curve: Does performance increase as a power function (Logan, 1988; Newell & Rosenbloom, 1981) or as an exponential function (Heathcote et al., 2000)? In much of the research on practice, the dependent variable of interest is either proportion correct or the mean RT for correct responses (but see Logan, 1992). One of the advantages of the Ratcliff diffusion model is that it takes into account the full range of data: proportion correct, the RT distributions for correct responses, and the RT distributions for error responses (Ratcliff, Thapar, & McKoon, 2006b). For this reason alone, it would be useful to apply the diffusion model to data from a practice experiment.

From a diffusion model perspective, one might hypothesise that the practice effect is associated with an increase in drift rate (see Brown & Heathcote, 2005, for a detailed analysis). One may further hypothesise that such an increase in drift rate may lead participants to lower their boundary separation, a strategic adjustment that keeps error rate constant but leads to more gains in RT. In order to test these hypotheses, Dutilh, Wagenmakers, Vandekerckhove, and Tuerlinckx (2008) carried out a 5-day, 10,000-trial lexical decision experiment in which participants completed 25 blocks of 400 stimuli each. Two participants were instructed to pay attention mainly to response accuracy, and two participants were instructed to pay attention mainly to response speed.

The behavioural results showed that the accuracy-instructed participants improved mostly on response speed, whereas speed-instructed participants improved mostly on response accuracy. The diffusion model analysis showed that, as expected, drift rate increased with practice for all participants. For the accuracy-instructed participants, practice also decreased boundary separation,[12] and—to our surprise—practice caused a 100 ms decrease in nondecision time. For the speed-instructed participants, boundary separation, starting point, and nondecision time did fluctuate over practice, but not in a systematic manner (for details see Dutilh et al., 2008).

Thus, a diffusion model analysis of the practice effect showed that the effect of practice may involve more than one underlying process. This conclusion seriously calls into question the usefulness of fitting power laws or exponential laws to a sequence of mean RTs.

## Phenomenon 5: Implicit Association Test

As is evident from the silly article by Bones and Johnson (2007),[13] the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is one of the most popular tools in experimental social psychology (see Fazio & Olson, 2003, and Nosek, Greenwald, & Banaji, 2006, for reviews). The IAT consist of two tasks. For example, the first task may be to classify faces according to race (i.e., Black or White), and the second task may be to classify words according to valence (i.e., positive or negative). In the two critical phases of the IAT, these tasks are performed in alternation. In one phase, the response key for the Black faces is the same as the one for negatively valenced words, and the response key for the White faces is the same as the one for positively valenced words. In this phase, the response mapping is compatible with a prejudice that favours White people over Black people. In another phase, the response mapping is incompatible with this prejudice, such that the response key for Black faces is the same as that for positively valenced words, and the response key for White faces is the same as that for negatively valenced words. In the IAT, responses are generally faster and more accurate for the compatible response mapping than for the incompatible response mapping, and the size of this difference may be taken as a measure of the implicit attitude under consideration.

One problem with the IAT is that people may be aware of the fact that the incompatible mapping is more difficult, and, in order to prevent too many error responses, they may adopt a more conservative threshold setting in the incompatible phase than in the compatible phase (Brendl, Markman, & Messner, 2001). Thus, performance differences between compatible and incompatible response mappings are not pure measures of implicit prejudice, as IAT dogma would have it. Instead, these performance differences could be partly due to strategic considerations that reflect a participant's uncertainty.

[12] This decrease in boundary separation was clearly observed for one of the participants, and clearly absent for the other.

[13] Yes, this really is a silly article.

In order to separately estimate the different components of processing that together determine performance on the IAT, Klauer, Voss, Schmitz, and Teige-Mocigemba (2007, Exp. 1) applied the diffusion model to data from a flower–insect IAT. The results showed that, compared to the compatible response mapping, the incompatible response mapping lead to an increase in boundary separation (i.e., more response caution), an increase in nondecision time, and a decrease in drift rates. Follow-up experiments showed that only the difference in drift rate correlated significantly with attitude ratings.

In sum, the application of the diffusion model to the IAT shows that the compatibility effect can be attributed to three different psychological processes, that is, boundary separation, nondecision time, and drift rate (Klauer et al., 2007). As only drift rate reflects the attitude process of interest, the diffusion model can be used to filter out unwanted strategic variance in the IAT, thereby providing a cleaner estimate for implicit attitudes.



## Phenomenon 6: Accessory stimulus effect

The accessory stimulus effect refers to the finding that response times speed up when the imperative stimulus is accompanied by a salient but task-irrelevant stimulus that is presented in a different perceptual modality (Bernstein, Clark, & Edelstein, 1969). Usually, the imperative stimulus is presented in the visual modality, and the accessory stimulus (AS) is presented in the auditory modality. For instance, Jepma, Wagenmakers, Band, and Nieuwenhuis (in press) presented an auditory tone immediately prior to the onset of a letter string that required a lexical decision; in this case, the beneficial effect of the AS on mean lexical decision RT was on average 24 ms.

The literature shows little consensus with respect to the origin of the AS effect. At least four accounts have been proposed:

1. The *energy-integration hypothesis*, which postulates that the AS facilitates stimulus encoding (Bernstein, Rose, & Ashe, 1970).
2. The *drift rate hypothesis*, which postulates that the AS increases the rate with which information from the imperative stimulus is accumulated (Hackley & Valle-Inclán, 1999).
3. The *boundary separation hypothesis*, which postulates that the AS leads to a more risky threshold setting (Posner, 1978).
4. The *motor time hypothesis*, which postulates that the AS decreases the time needed to execute the motor response (Sanders, 1980).

It is evident that the different hypotheses make specific predictions with respect to the effect of the AS on response accuracy and the distribution of RTs. In an attempt to differentiate between the various hypotheses, Jepma et al. (in press) applied the diffusion model to the data from a lexical decision experiment with an auditory tone as AS. The data showed that the effect of the AS was to shift the entire RT distribution by the same amount, and that there was no effect on response accuracy. This pattern of results suggests that the AS effect does not affect the decision process. Indeed, the diffusion model analyses provided overwhelming support for the hypothesis that the AS effect is to decrease the nondecision time $T_{er}$. Hence, the experiment by Jepma et al. undercuts both the drift rate hypothesis and the boundary separation hypothesis. In a separate ERP experiment, Jepma et al. found that the AS modulated the P1 and N1 components, and this supports Bernstein's original energy-integration hypothesis.



## CONCLUDING REMARKS

For over a century, experimental psychologists have studied the complex structure of human cognition using relatively simple tasks, and, indeed, a simple methodology. The present paper suggests that considerable advantages can be gained by using a more sophisticated methodology. The recent developments sketched in this paper confirm that the Ratcliff diffusion model is a promising candidate to replace or supplement the simple methods of analysis that have up till now dominated experimental psychology.

Original manuscript received January 2008
Revised manuscript received April 2008
First published online September 2008

## REFERENCES

Batchelder, W. H. (2007, October). *Cognitive psychometrics: Combining two psychological traditions.* CSCA lecture, Amsterdam, The Netherlands.

Bernstein, I. H., Clark, M. H., & Edelstein, B. A. (1969). Effects of an auditory signal on visual reaction time. *Journal of Experimental Psychology*, *80*, 567–569.

Bernstein, I. H., Rose, R. G., & Ashe, V. (1970). Energy integration in intersensory facilitation. *Journal of Experimental Psychology*, *86*, 196–203.

Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, *19*, 442–477.

Bones, A. K., & Johnson, N. R. (2007). Measuring the immeasurable: Or "could Abraham Lincoln take the Implicit Association Test?". *Perspectives on Psychological Science*, *2*, 406–411.

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 760–773.

Brinley, J. F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In A. T. Welford & J. E. Birren (Eds.), *Behavior, aging and the nervous system* (pp. 114–149). Springfield, IL: Thomas.

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, *17*, 471–510.

Brown, S., & Heathcote, A. (2005). Practice increases the efficiency of evidence accumulation in perceptual choice. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 289–298.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, *98*, 67–83.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.

Cox, D. R., & Miller, H. D. (1970). *The theory of stochastic processes.* London: Methuen.

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, *31*, 567–587.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, *2*, 153–166.

Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception and Psychophysics*, *68*, 194–207.

Dutilh, G., Wagenmakers, E.-J., Vandekerckhove, J., & Tuerlinckx, F. (2008). A diffusion model account of practice. *Manuscript submitted for publication*.

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, *2*, 312–329.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen [On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat]. *Annalen der Physik*, *17*, 549–560.

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

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin and Review*, *9*, 3–25.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327.

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, *19*, 1120–1136.

Gardiner, C. W. (2004). *Handbook of stochastic methods* (3rd ed.). Berlin, Germany: Springer-Verlag.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge, UK: Cambridge University Press.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.

Grasman, R. P. P. P., Wagenmakers, E.-J., & van der Maas, H. L. J. (2008). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Manuscript submitted for publication*.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Hackley, S. A., & Valle-Inclán, F. (1999). Accessory stimulus effects on response selection: Does arousal speed decision making? *Journal of Cognitive Neuroscience*, *11*, 321–329.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review*, *7*, 185–207.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*, 101–117.

Honerkamp, J. (1994). *Stochastic dynamical systems*. New York: VCH Publishers.

Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger Publishers.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Amsterdam: Elsevier.

Jepma, M., Wagenmakers, E.-J., Band, G. P. H., & Nieuwenhuis, S. (in press). The effects of accessory stimuli on information processing: Evidence from electrophysiology and a diffusion model analysis. *Journal of Cognitive Neuroscience*.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 377–395.

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.

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

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*, 309–325.

Lee, M., Vandekerckhove, J., Navarro, D. J., & Tuerlinckx, F. (2007, July). *A hierarchical Bayesian account of human decision-making using Wiener diffusion*. Paper presented at the 40th annual meeting of the Society for Mathematical Psychology, Irvine, USA.

Lee, M. D., Fuss, I., & Navarro, D. (2006). A Bayesian approach to diffusion models of decision-making and response time. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (Vol. 19 pp. 809–815). Cambridge, MA: MIT Press.

Lee, M. D., & Wagenmakers, E.-J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review*, *112*, 662–668.

Lindley, D. V. (2000). The philosophy of statistics. *The Statistician*, *49*, 293–337.

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

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

Lo, C. C., & Wang, X. J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience*, *9*, 956–963.

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

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

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

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Meng, X. L. (1994). Posterior predictive p-values. *Annals of Statistics*, *22*, 1142–1160.

Myung, I. J., Forster, M. R., & Browne, M. W. (Eds.). (2000). Model selection [Special issue]. *Journal of Mathematical Psychology*, *44*(1–2).

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

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327–357.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2006). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). London: Psychology Press.

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, *134*, 368–387.

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

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, *12*, 765–805.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge, UK: Blackwells.

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

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

Ratcliff, R. (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 and Review*, *9*, 278–291.

Ratcliff, R. (in press). The EZ diffusion method: Too EZ? *Psychonomic Bulletin and Review*.

Ratcliff, R., Gomez, P., & McKoon, G. (2004a). Diffusion model account of lexical decision. *Psychological Review*, *111*, 159–182.

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

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

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

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence*, *36*, 10–17.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

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., Spieler, D., & McKoon, G. (2004b). Analysis of group differences in processing speed: Where are the models of processing? *Psychonomic Bulletin and Review*, *11*, 755–769.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004c). 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 and Psychophysics*, *65*, 523–535.

Ratcliff, R., Thapar, A., & McKoon, G. (2004d). 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. (2006a). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin and Review*, *13*, 626–635.

Ratcliff, R., Thapar, A., & McKoon, G. (2006b). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging*, *21*, 353–371.

Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging*, *22*, 56–66.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times 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*, *102*, 261–300.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin and Review*, *12*, 573–604.

Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin and Review*, *12*, 195–223.

Rouder, J. N., Lu, J., Sun, D., Speckman, P., Morey, R., & Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika*, *72*, 621–642.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*, 487–494.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Sanders, A. F. (1980). Stage analysis of reaction process. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 331–354). Amsterdam: North-Holland.

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.

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143–153.

Shiffrin, R. M., Lee, M. D., Wagenmakers, E.-J., & Kim, W. J. (in press). Model evaluation and selection: Established methods and recent developments. *Cognitive Science*.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.

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.

Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS version 1.4 user manual*. Cambridge, UK: Medical Research Council Biostatistics Unit.

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

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review*, *19*, 447–462.

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.

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

Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behavior Research Methods, Instruments, and Computers*, *36*, 702–716.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin and Review*, *14*, 1011–1026.

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. (in press). A Bayesian approach to diffusion process models of decision-making. *Proceedings of the 30th annual conference of the Cognitive Science Society.*

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory and 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.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review*, *14*, 779–804.

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*, 830–841.

Wagenmakers, E.-J., Grasman, R. P. P. P., & Molenaar, P. C. M. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology*, *49*, 195–204.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*, 140–159.

Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology*, *48*, 332–367.

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

Wagenmakers, E.-J., & Waldorp, L. (Eds.). (2006). Model selection: Theoretical developments and applications [Special issue]. *Journal of Mathematical Psychology*, *50*(2).

Weeda, W. D., Wagenmakers, E.-J., & Huizenga, H. M. (2007). Empirical support for a diffusion model account of the worst performance rule. *Manuscript submitted for publication*.

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

Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (2006). Single versus dual process models of lexical decision performance: Insights from RT distributional analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1324–1344.