



How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT

Don van Ravenzwaaij^{a,*}, Klaus Oberauer^b

^a Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

^b University of Bristol, UK

ARTICLE INFO

Article history:

Received 21 January 2009

Received in revised form

13 August 2009

Available online 14 October 2009

Keywords:

Reaction times

Diffusion model

Parameter estimation

ABSTRACT

Parameter recovery of three different implementations of the Ratcliff diffusion model was investigated: the EZ model (Wagenmakers, van der Maas, & Grasman, 2007), fast-dm (Voss & Voss, 2007), and DMAT (Vandekerckhove & Tuerlinckx, 2007). Their capacity to recover both the mean structure and individual differences in parameter values was explored. The three methods were applied to simulated data generated by the diffusion model, by the leaky, competing accumulator (LCA) model (Usher & McClelland, 2001) and by the linear ballistic accumulator (LBA) model (Brown & Heathcote, 2008). Results show that EZ and DMAT are better capable than fast-dm in recovering experimental effects on parameters. EZ was best in recovering individual differences in parameter values. When data were generated by the LCA model, the diffusion model estimates obtained with all three methods correlated well with corresponding LCA model parameters. No such one-on-one correspondence could be established between parameters of the LBA model and the diffusion model.

© 2009 Elsevier Inc. All rights reserved.

Response times (RTs) are one of the prime dependent variables in experimental cognitive psychology. Despite their appeal as apparently straightforward measures of the duration of cognitive processes, several decades of research have revealed that even the RTs of relatively simple perceptual choice tasks reflect the interaction of a number of internal variables and processes (see e.g. Luce, 1986; Ratcliff, Van Zandt, & McKoon, 1999). This insight implies that the interpretation of RTs requires a measurement model that makes explicit how the latent variable of interest – e.g., the duration of a cognitive process – is translated into the observed variable, RT. Several models of RTs have been proposed (for reviews see Luce, 1986; Ratcliff & Smith, 2004), but their application has been hampered by the fact that the models were not easily applicable to the data emerging from a typical experiment, for two reasons. First, fitting the models to data is technically demanding, and second, they require a large number of data points in each experimental condition to provide a precise reflection of the underlying RT distribution (see e.g. Ratcliff & Tuerlinckx, 2002). For this reason, most experimental psychologists continue to use the mean or median of RT distributions as a direct reflection of the duration of a cognitive process of interest, thus ignoring a wealth of available information (i.e., the shape of the RT distribution, the accuracy, and the RTs of errors).

RTs are also increasingly used in psychometric research to measure individual differences in general processing speed, or of speed in specific cognitive processes (Danthiir, Roberts, Schulze, & Wilhelm, 2005; Fry & Hale, 1996; Larson & Alderton, 1990; Salthouse, 1998; Wilhelm & Oberauer, 2006). In this field, the need for an adequate and practical measurement model is equally pressing. It becomes most obvious in the shape of the speed–accuracy trade-off problem: When individuals differ in their inclination to trade accuracy for speed, individual mean RTs cannot be interpreted as reflections of a person's information processing speed without looking at their accuracy at the same time. This problem has been discussed for some time, but so far no satisfactory solution has been found for integrating individual measures of RTs and accuracies (Dennis & Evans, 1996). Therefore, individual-differences research would also benefit substantially from a measurement model that is adequate and easy to apply to RT data from individuals without making unrealistic demands on the number of data points per person.

The most thoroughly investigated model of RTs so far is Ratcliff's (1978) diffusion model for two-alternative forced-choice (2-AFC) tasks. This model has received substantial empirical support and arguably is superior to many other models (Ratcliff & Smith, 2004; Ratcliff et al., 1999; for a more recent competitor that seems to be equally successful see Brown & Heathcote, 2008). The diffusion model has been successfully applied to understand and explain the processes underlying research on lexical decision making (Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008), memory (Ratcliff, 1978, 1988), simple

* Corresponding author.

E-mail address: d.vanravenzwaaij@uva.nl (D. van Ravenzwaaij).

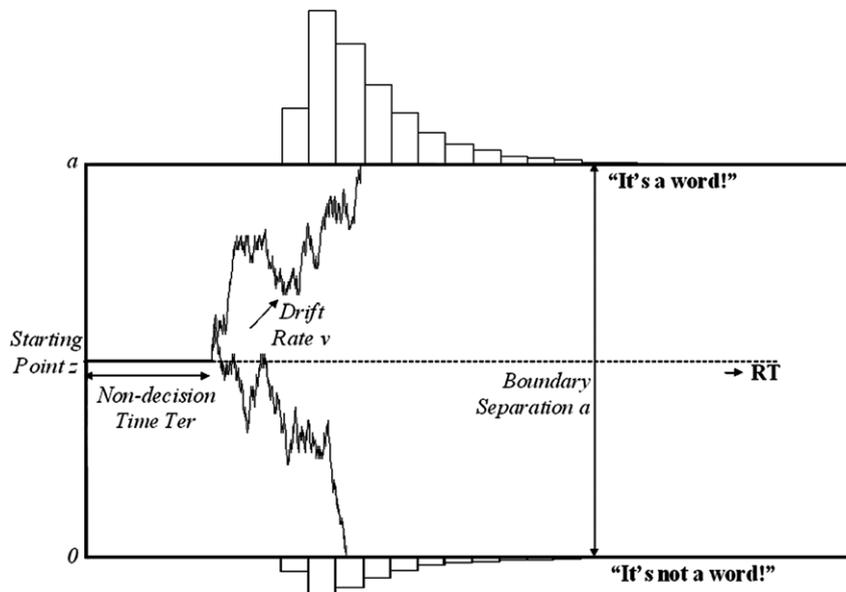


Fig. 1. Illustration of the diffusion model and parameters. The panel shows the drift rate (v), boundary separation (a), non-decision time (T_{er}), starting point (z) and illustrates how accuracy and response times for correct and error responses change as a function of the noise component within each trial (s).

reaction times (Smith, 1995), familiarity effects (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007; van Ravenzwaaij, van der Mass & Wagenmakers, submitted for publication) and perceptual judgments (Ratcliff, 2002; Ratcliff & McKoon, 2008). The diffusion model, therefore, is a promising candidate for an adequate measurement model for RT data. One of its strengths is that it integrates information from RTs and accuracies, thus solving the speed–accuracy trade-off problem. This makes the model particularly attractive for investigating individual differences, and some research has already begun to use the diffusion model to measure individual differences in the speed of cognitive processes (Oberauer, 2005; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007).

Recent years have witnessed a major advance in development of techniques for applying the diffusion model to data. Three such methods are now available – the EZ diffusion model (Wagenmakers et al., 2007), fast-dm (Voss & Voss, 2007), and DMAT (Vandekerckhove & Tuerlinckx, 2007). The purpose of the present paper is to evaluate these three methods by applying them to simulated data that were generated by the diffusion model. We ask how well each method recovers the true parameters from the data. Different research traditions are interested in different aspects of parameter recovery accuracy: for experimental research, accurate reflection of differences between experimental conditions is of primary importance, whereas psychometric research is mostly interested in accurate measurement of differences between individuals. Our work investigates these two aspects by simulating both experimental manipulations that affect individual parameters and individual differences in all model parameters. We ask how well the parameters recovered by the competing measurement methods for each individual and each condition reflect the experimental effects, and how well they correlate with the true parameter values.

In the next section, we will outline the diffusion model. Then, we will discuss the three methods for estimating parameters of the diffusion model from data, the EZ model, fast-dm, and DMAT. Using simulated data, we will investigate how they measure up to one another in terms of their capacity to recover experimental effects as well as individual differences in parameters, in particular under realistic conditions of empirical research, that is, with relatively small numbers of data points per person and condition. In our conclusion, we will argue that the method to use depends on the specific interests of the researcher.

1. Ratcliff's diffusion model

The diffusion model was originally applied to psychology by Ratcliff (1978) and is useful for analyzing data from 2-AFC response tasks, such as the lexical decision task. In Fig. 1, the diffusion model is graphically displayed. When performing a 2-AFC task, participants are accumulating evidence in favor of either of the two response alternatives. As soon as the collected evidence reaches a certain threshold, a response is given. This 'evidence threshold' varies between people, signifying a difference in response conservativeness. From the starting point of the decision, information is accumulated in a noisy fashion toward either the upper decision boundary (corresponding to the word response) or the lower decision boundary (corresponding to the non-word response) with a certain rate.

The mean of the rate of information accumulation is the drift rate of the process, denoted by v . Furthermore, drift rate has within-trial variability, denoted by s^2 , which causes the accumulation of information to occur in a noisy fashion and leads to variation of the response time (RT) over trials.¹

The lower decision boundary is always fixed at 0, so that the upper boundary, or a , is identical to the boundary separation (see Fig. 1). Once a boundary is reached, a response is given. Occasionally, the wrong boundary is reached, resulting in an incorrect response. The model thus predicts the probability of the occurrence of an error response and its relation with RT: the larger the boundary separation, the smaller the chance of making an error, but the higher the RT. Thus, the boundary separation is a measure of response conservativeness; it reflects the individual's speed–accuracy trade-off setting.

At stimulus onset, the subject is uncertain with respect to the identity of the stimulus. This is signified by a starting point of the decision process, denoted by z , that lies somewhere between the decision boundaries (see Fig. 1). Often, the starting point is assumed to be exactly in the middle of the two boundaries, but this need not be the case, as subjects may be biased towards either of the two response alternatives.

¹ This parameter is always fixed, the magnitude of all other parameters is linearly related to this one. We opted for the value $s^2 = 1$ for this paper.

The drift rate, boundary separation and starting point together determine the decision time (DT). Other stages of information processing between stimulus onset and motor response, such as stimulus encoding, memory access, retrieval cue assembly etc. are combined in the non-decision time, or T_{er} . For simplicity, the model assumes that all these processes are totally independent from the actual decision processes and are therefore additive to DT.

In the full version of the diffusion model, there are three other parameters, corresponding to measures of variability across trials for drift rate (η), for starting point (s_z) and for the non-decision component (s_r). They are not displayed here for the sake of simplicity, but are elaborately described in Ratcliff and Tuerlinckx (2002).

2. Methods for measuring parameters of the diffusion model

We next discuss the three methods for measuring the parameters of the diffusion model that we will compare. These three methods have in common that they are available as easy-to-use program packages or codes, and make lean demands on computation time. The first method is the EZ diffusion model (Wagenmakers et al., 2007). This method is the simplest method, because there is no parameter estimation involved. Instead, the EZ model uses the mean and variance of RT and the mean accuracy and computes from them a value for drift rate, boundary separation, and non-decision time. The other parameters in the full diffusion model are not given by EZ. Code is provided in the Appendix of the paper by Wagenmakers et al. (2007), which runs on the open-source statistical package R (R Development Core Team, 2004).²

The second method is the fast-dm software package,³ which is based on a Kolmogorov–Smirnov fitting routine (Voss & Voss, 2008, 2007). Fast-dm allows for estimation of the full range of parameters, including the mean drift rate (v), boundary separation (a), mean non-decision time (T_{er}), mean starting point (z), standard deviation of the drift rate (η), the range of the starting point (s_z), and the range of the non-decision time (s_r). Also, fast-dm allows for inclusion of experimental conditions, so that particular parameters can be manipulated and others can be fixed. For instance, users can assume that an experimental manipulation affects only the mean drift rate, and then configure fast-dm such that only v is free to vary between conditions.

The last method is the DMAT toolbox⁴ (Vandekerckhove & Tuerlinckx, 2007, 2008), which runs on Matlab (Mathworks, 1994). This method is based on minimizing a negative multinomial log-likelihood function, which is conceptually similar to maximum likelihood estimation. Like fast-dm, DMAT allows for estimation of the full range of parameters and it allows for parameter restrictions.

We present two simulation studies. Simulation 1 asks how well the three methods for obtaining diffusion-model parameter estimates recover the true parameters from a data set that has been generated by the diffusion model. This simulation represents the optimistic scenario in which we assume that the diffusion model is an essentially correct model for two-choice RT data. Simulation 2 represents the more pessimistic scenario in which the diffusion model is not correct, and RT data are generated by a different process. Here we ask whether the parameter estimates obtained by the three methods nevertheless reflect parameters of the true process that generated the data in a systematic and meaningful way. To that end, we simulated data from two competing models for RTs:

the leaky, competing accumulator (LCA) model of Usher and McClelland (2001), and the linear ballistic accumulator (LBA) model of Brown and Heathcote (2008). These models have parameters that roughly correspond to the core parameters of the diffusion model, drift rate, boundary separation, and non-decision time, and we therefore investigate whether the estimated diffusion-model parameters capture individual differences in the corresponding parameters of the model that generated the data. If the answer is positive, we can use the methods for estimating diffusion-model parameters with much more confidence, because interpretation of the parameters does not depend on the unrealistic assumption that the data were generated by a diffusion process exactly as specified in Ratcliff's diffusion model.

To summarize, both simulations address the validity of the diffusion model as a measurement model: To what degree do the parameter estimates obtained from applying the model reflect the variables we intend to measure? Simulation 1 assumes that the diffusion model is essentially correct, and asks which method best recovers the true parameters of the diffusion process that generated the data. Simulation 2 assumes that the diffusion model is not correct and asks whether the parameter estimates can still be regarded as valid measurements of the variables of interest.

3. Simulation 1: Fitting data generated by the diffusion model

3.1. Method

The simulation and parameter recovery together consisted of four steps. First, we generated a set of 'true' parameters, based on an existing dataset and on a variance–covariance matrix that determined how the parameters would correlate with one another. Second, we simulated data with these parameters. Third, we applied the three diffusion measurement models to the data. Fourth and last, we compared the parameter estimates to the true parameters for all methods, evaluating their capacity to capture experimental manipulations and individual differences in the dataset, their robustness when applied to sparse data, and their bias in recovering true parameter values.

To compare performance of the three diffusion model implementations, we simulated individual differences data, based on unpublished data by Wilhelm, Keye, and Oberauer. In that study, a sample of 148 participants was tested on three two-choice RT tasks. The tasks required rapid classification of stimuli by pressing one of two buttons. One task used arrows as stimuli, one used words, and one used shapes. For each task two experimental manipulations were realized, one (stimulus–response compatibility) assumed to affect primarily drift rate, and the other (speed–accuracy instruction) assumed to affect only the boundary separation. A diffusion model analysis on this dataset using the procedure of Voss, Rothermund, and Voss (2004), which is a predecessor of fast-dm, yielded parameter estimates for each condition from three different tasks; we used these to inform the means and SDs of parameters in our simulation.

For the first step towards the simulated dataset, we calculated two mean drift rates, one for the compatible stimulus–response mapping (v_c) and one for the incompatible mapping (v_i). This was done by taking the grand mean of the drift rate estimates obtained from fitting the Voss et al. (2004) model, averaging across all participants, the three tasks, and the speed–accuracy manipulation for each mapping condition. In the same way, we computed a grand mean for boundary separation in the speed-instruction condition (a_{sp}) and one for the accuracy-instruction condition (a_{acc}). For the remaining parameters except z , we computed the grand mean across all conditions, as no meaningful variation over conditions should be expected. Parameter z was set to $a/2$ for each simulated subject, reflecting an unbiased mean starting point, because the

² R can be freely downloaded at <http://cran.r-project.org/>.

³ Fast-dm can be freely downloaded at <http://www.psychologie.uni-freiburg.de/Management/voss/fast-dm>.

⁴ The DMAT toolbox can be freely downloaded at http://ppw.kuleuven.be/okp/people/Joachim_Vandekerckhove/.

Table 1

The mean and SD of the diffusion model parameters upon which the simulation dataset is based. v_c = compatible drift rate, v_i = incompatible drift, a_{sp} = speed boundary separation, a_{acc} = accuracy boundary separation.

| | v_c | v_i | a_{sp} | a_{acc} | T_{er} | η | s_z | s_t |
|------|-------|-------|----------|-----------|----------|--------|-------|-------|
| Mean | 4.00 | 3.00 | .50 | .85 | .25 | .30 | .10 | .08 |
| SDs | .70 | .70 | .10 | .10 | .03 | .10 | .05 | .04 |

choice RT tasks in the unpublished data set that informed the simulation provided no grounds for any systematic bias in favor of one or the other response (i.e., both responses were objectively equally likely at the start of each trial), as is commonly the case in choice RT experiments. A further reason for this decision was that the EZ diffusion model is based on the assumption that $z = a/2$, and thus could not be applied if that assumption was seriously violated.⁵

These parameter values were then adjusted by hand to obtain values that generated mean RTs and accuracies, and their standard deviations, that were close to the data.⁶ The parameter values and their standard deviations that we used to create the simulated data are presented in Table 1.

The next step was to create individual differences in the dataset. This requires setting the correlations between parameters to plausible values. Based on the observation that RTs in different conditions of a within-subjects experiment are typically highly correlated (see e.g. Wagenmakers & Brown, 2007), we set the correlation between v_c and v_i to $r = .8$, and likewise, we set the correlation between a_c and a_i to $r = .8$. Based on the pervasive observation that means and standard deviations of RTs are highly correlated, we decided to assume a correlation of $r = .8$ between each mean parameter and its corresponding variability parameter. In particular, we had both v_c and v_i correlate .8 with η , and we had T_{er} correlate .8 with s_t . Also, we had both a_c and a_i correlate .7 with s_z . All other correlations were set to 0 for simplicity. Different from the parameter means, their correlations were not informed directly by the data, but rather more indirectly by common observations in RT experiments. The reason why we did not use the observed correlations between parameter estimates obtained with the Voss et al. (2004) procedure is that parameter correlations are potentially seriously distorted by parameter trade-offs during fitting. This problem has been addressed empirically and through simulation by Schmiedek et al. (2007), who developed a method for separating genuine correlations from correlation artifacts caused by parameter trade-offs. Schmiedek et al., however, used the EZ diffusion model, which does not include the variability parameters. Therefore, no reliable information exists on the true correlation between all parameters of the diffusion model. As a result, the correlation matrix underlying our simulations is to some degree an informed guess; other correlation values are conceivable, a point to which we return in the Discussion.

From the variances of the parameters and their assumed correlations we computed their variance–covariance matrix. We used *mvnrm* (available in the MASS R package) to simulate values from the multivariate normal distribution, based on the mean parameter estimates and the variance–covariance matrix. Because for all parameters except drift rate, only positive values are meaningful, we

⁵ See Grasman, Wagenmakers, and van der Maas (2009) for an extension of the EZ diffusion model that can incorporate bias.

⁶ Adjustment by hand was necessary because in the grand means of estimated parameters, the experimental manipulations of stimulus–response compatibility and speed–accuracy instruction had effects on all parameters rather than just the parameter they were intended to affect. Setting all but one parameter value equal across conditions required adjustments to parameter values because otherwise the simulated data deviated from the real data with regard to mean RT and accuracy.

Table 2

Mean RT in ms and mean accuracy in percentage (between participant SDs added in parentheses). Sp C = Speed Compatible, Sp I = Speed Incompatible, Acc C = Accuracy Compatible, Acc I = Accuracy Incompatible.

| | Sp C | Sp I | Acc C | Acc I |
|--------------|-------------|-------------|-------------|-------------|
| RT (ms) | 299 (47) | 304 (52) | 352 (79) | 374 (100) |
| Accuracy (%) | 86.5 (33.7) | 80.4 (39.2) | 95.8 (19.9) | 90.7 (28.5) |

truncated all parameter values except drift rate by setting negative values to zero (this affected less than 1% of all parameter values). In this way we generated values for 148 simulated participants for the eight parameters mentioned in Table 1. Lastly, we divided both boundary separation values (for the speed and the accuracy instruction condition) by two to get two corresponding values for z .

The final step was to use all generated diffusion model parameters to simulate 800 trials per condition for each of the 148 participants. We generated data using the procedure suggested by Ratcliff and Tuerlinckx (2002, pp. 4–5).

The resulting dataset, of which means and SDs of RTs and accuracies can be found in Table 2, was analyzed with EZ, fast-dm and DMAT. We used version 29 of fast-dm (January 13, 2008), and version 0.4 of DMAT (April 17, 2007). Since EZ is an algorithm, there is no specific version number. The EZ model was applied separately to each condition, thus yielding different parameter estimates of v , a , and T_{er} for each of the four conditions. For fast-dm and DMAT, we left the three main parameters, v , a , and T_{er} , free to vary across the four experimental conditions. The variability parameters, η , s_z , and s_t were constrained to be equal across conditions. We believe that this is a reasonable fitting strategy for most experiments, in which researchers typically are interested in which of the three main parameters is affected by an experimental manipulation, but are less interested in the variability parameters, which ought to be constrained to minimize parameter trade-offs.

3.2. Results

To see how well individual differences are captured by the parameter measurement routines, we calculated correlations between the true parameters (upon which the generated dataset was based) and the parameters estimated or computed from the data by the EZ diffusion model, fast-dm, and DMAT.⁷ The results are displayed in Table 3. These correlations can be interpreted as estimated validity coefficients for the parameters when using the diffusion model as a measurement model, because they reflect how well the measurement reflects the true variance of the variable it intends to measure (Borsboom, Mellenbergh, & van Heerden, 2004).

As can be seen from the table, the estimated parameters covary very strongly with the true parameters. Both EZ and fast-dm appear to be well capable of capturing individual differences in v , a and T_{er} . DMAT did worse on boundary separation in the accuracy conditions. For η and s_z , both fast-dm and DMAT did poorly, with correlations close to zero; s_t was recovered well by fast-dm but not by DMAT.

To see how robust the estimation routines are in the face of sparser numbers of trials per condition, we ran a bootstrap analysis for the EZ method, in which we randomly selected 80 trials per participant from the full data set 2000 times, calculated diffusion parameters based on each of these samples, correlated each of these parameter sets with the true parameters and calculated

⁷ The CPU time required by the three different methods varied strongly, with EZ taking less than a minute to calculate its parameters, fast-dm requiring a little under 50 minutes for parameter estimation and DMAT requiring about two hours.

Table 3

Correlations between the true parameters and the parameter estimates for each condition, based on the full dataset of 800 trials per condition. Sp C = Speed Compatible, Sp I = Speed Incompatible, Acc C = Accuracy Compatible, Acc I = Accuracy Incompatible.

| Parameters | Condition | EZ | Fast-dm | DMAT |
|-----------------------|----------------------|-----|---------|------|
| <i>v</i> | Sp C | .85 | .70 | .75 |
| | Sp I | .93 | .70 | .83 |
| | Acc C | .96 | .87 | .88 |
| <i>a</i> | Sp C | .89 | .85 | .95 |
| | Sp I | .92 | .84 | .97 |
| | Acc C | .92 | .90 | .45 |
| <i>T_{er}</i> | Sp C | .97 | .92 | .95 |
| | Sp I | .97 | .90 | .95 |
| | Acc C | .98 | .95 | .83 |
| <i>η</i> | – | – | .15 | .13 |
| | <i>s_z</i> | – | –.08 | .19 |
| | <i>s_t</i> | – | .86 | .48 |

the average correlation over bootstrap samples. For fast-dm, this method would have been too time-consuming. Instead, we split the set of 800 simulated trials into 10 random subsets of 80 trials, and estimated parameters for each of these subsets. DMAT was incapable of estimating parameters reliably for 80 trials per condition, as it requires at least 11 errors per RT quantile (divided in .1, .3, .5, .7, .9 and 1 quantiles), so we report results here for EZ and fast-dm only. Results for *v*, *a* and *T_{er}* can be found in Table 4.

When comparing Tables 3 and 4, it becomes apparent that the correlations between true and recovered parameters are reduced when only 80 instead of 800 trials per condition are used. In particular, the drift-rate estimates of fast-dm suffered considerably from the reduction in the data basis. Overall, EZ seems to be more robust to a smaller number of trials than fast-dm, providing estimates that correlate consistently higher with the true parameters than fast-dm.

To see how well each method is capable of capturing the mean structure of the data, we subtracted the parameter estimates from the true parameter values and divided the mean of this result by the mean true values. We multiplied the resulting proportional residuals by 100 to convert them to percentages. They are displayed in Table 5.

As can be seen from the table, EZ systematically underestimates *v* by about 6% to 13%, overestimates *a* by about 2% to 11% and underestimates *T_{er}* by about 3% to 4%. However, the bias of EZ does not change sign over conditions. Therefore, the estimates of *a* adequately reflect the true differences in boundary separation in the two speed–accuracy conditions, and the estimates of *v* reflect the true differences in drift rate between the compatibility conditions.

Fast-dm seems to be more biased than EZ, in particular for drift rate. Also, its bias is less consistent than EZ’s bias as evident by the larger standard errors of the residuals. Fast-dm underestimates *v* in the speed conditions, but overestimates *v* in the accuracy conditions. The reverse seems to hold for *a*, although less clearly so. In other words, fast-dm appears to shrink to the mean, thereby underestimating the true difference between conditions. The dispersion parameters *η* and *s_z* are recovered poorly, but *s_t* is recovered nicely.

The magnitude of DMAT’s bias seems to be the lowest of the three, except for the Acc-C condition. DMAT’s consistency appears to be somewhat in between that of EZ and fast-dm. Interestingly, DMAT overestimates both *v* and *a*, but underestimates *T_{er}*. As with fast-dm, *η* and *s_z* are recovered poorly, but the recovery of *s_t* is acceptable.

To see how the size of the bias is related to the magnitude of the parameter, we plotted residual graphs for *v* in the Sp-C condition for all three estimation routines (see Fig. 2). As shown before, EZ

Table 4

Average correlations between the true parameters and the parameter estimates for each condition, based on random samples of 80 trials per condition. Sp C = Speed Compatible, Sp I = Speed Incompatible, Acc C = Accuracy Compatible, Acc I = Accuracy Incompatible.

| Parameters | Condition | EZ | Fast-dm |
|-----------------------|----------------------|-----|---------|
| <i>v</i> | Sp C | .77 | .49 |
| | Sp I | .86 | .59 |
| | Acc C | .85 | .62 |
| <i>a</i> | Sp C | .91 | .78 |
| | Sp I | .83 | .75 |
| | Acc C | .85 | .75 |
| <i>T_{er}</i> | Sp C | .73 | .64 |
| | Sp I | .77 | .66 |
| | Acc C | .94 | .87 |
| <i>η</i> | – | – | .04 |
| | <i>s_z</i> | – | –.01 |
| | <i>s_t</i> | – | .71 |

systematically underestimates *v* (top left panel) and *T_{er}* (top right panel). This bias increases linearly with the size of the parameter. The positive bias in *a* (top middle panel) decreases as the true parameter value gets larger.

As evident from the middle panel of Fig. 2, fast-dm’s estimates have a larger bias and are less consistent than the EZ parameter estimates, basically mirroring the results presented in Table 5. The mean bias in *T_{er}* starts positive for small true values of *T_{er}*, but becomes negative for large true values of *T_{er}*. This again reflects the tendency of fast-dm to shrink individual differences towards the mean.

The bottom panels show residuals for DMAT. The bias of the DMAT estimates is relatively small. The bias in the estimates of *v* and *a* do not seem to be affected by the size of the true parameter. For *T_{er}* however, relatively large residuals arise when the true value is small.

4. Simulation 2: Fitting data generated by other models

We next created two simulated data sets using the LCA model by Usher and McClelland (2001), and the LBA model by Brown and Heathcote (2008). The data sets again represented the 2 × 2 design manipulating boundary separation (speed vs. accuracy conditions) and drift rate (compatible vs. incompatible mapping conditions).

4.1. Fitting data generated by the LCA model

Whereas Ratcliff’s diffusion model is applicable only to two-choice situations, the LCA model can be applied to an arbitrary number of alternatives. The model assumes that each alternative is represented by an accumulator collecting evidence for that choice, to which Gaussian noise with mean zero and standard deviation σ^2 is added. A decision is made as soon as one accumulator reaches a boundary θ . Different from the diffusion model, the accumulators are not linear. Rather, they lose a constant proportion of their current activation in each unit of time, so that their growth is negatively accelerated. The proportional leakage is a free parameter *k*. The accumulators for different alternatives inhibit each other, and the strength of inhibition is a free parameter β . To generate data we used Eq. (11) of Usher and McClelland (2001); this equation is an application of the model to two-choice experiments:

$$\begin{aligned}
 dx_1 &= [0.5(1 + v) - kx_1 - \beta x_2] \frac{dt}{\tau} + \xi_1 \sqrt{\frac{dt}{\tau}} \\
 dx_2 &= [0.5(1 - v) - kx_2 - \beta x_1] \frac{dt}{\tau} + \xi_2 \sqrt{\frac{dt}{\tau}}.
 \end{aligned}
 \tag{1}$$

Table 5
 Parameter estimates and proportional residuals for EZ, fast-dm and DMAT (with standard error of the mean in parenthesis). Residuals are calculated by subtracting the mean parameter estimates from the mean true parameter values, dividing these by the mean true parameters and multiplying the result by 100%. Thus, positive residuals indicate that the parameter estimates are too low, whereas negative residuals indicate that the parameter estimates are too high. Pars = Parameters, Con = Condition, Sp C = Speed Compatible, Sp I = Speed Incompatible, Acc C = Accuracy Compatible, Acc I = Accuracy Incompatible.

| Pars | Con | EZ | | Fast-dm | | DMAT | |
|-----------------------|-------|-------------|-------------|-------------|-------------|-------------------------|-------------|
| | | Estimates | Residuals | Estimates | Residuals | Estimates | Residuals |
| <i>v</i> | Sp C | 3.50 (0.65) | 12.5 (0.8) | 3.48 (1.08) | 13.0 (1.6) | 4.27 ^a (.89) | -6.8 (1.2) |
| | Sp I | 2.68 (0.64) | 10.5 (0.7) | 2.40 (0.90) | 20.0 (1.8) | 3.24 (0.89) | -8.1 (1.4) |
| | Acc C | 3.77 (0.62) | 5.9 (0.4) | 4.37 (0.69) | -9.3 (0.7) | 4.36 (1.10) | -9.0 (1.2) |
| | Acc I | 2.82 (0.65) | 6.1 (0.4) | 3.08 (0.73) | -2.8 (0.6) | 3.13 (0.86) | -4.4 (1.0) |
| <i>a</i> | Sp C | 0.56 (0.09) | -11.1 (0.7) | 0.57 (0.11) | -13.4 (1.0) | 0.52 (0.11) | -3.4 (0.6) |
| | Sp I | 0.55 (0.09) | -9.5 (0.7) | 0.55 (0.13) | -10.2 (1.2) | 0.51 (0.11) | -3.0 (0.4) |
| | Acc C | 0.88 (0.10) | -4.1 (0.4) | 0.89 (0.11) | -4.5 (0.5) | 0.98 (0.42) | -15.2 (0.4) |
| | Acc I | 0.87 (0.10) | -2.3 (0.4) | 0.83 (0.11) | 2.7 (0.5) | 0.88 (0.21) | -3.9 (0.2) |
| <i>T_{er}</i> | Sp C | 0.24 (0.02) | 4.4 (0.3) | 0.25 (0.03) | -2.0 (0.4) | 0.25 (0.03) | 0.5 (0.3) |
| | Sp I | 0.24 (0.02) | 4.1 (0.3) | 0.25 (0.03) | -2.0 (0.4) | 0.25 (0.03) | 0.3 (0.3) |
| | Acc C | 0.24 (0.03) | 3.4 (0.2) | 0.25 (0.03) | 1.1 (0.3) | 0.24 (0.04) | 2.3 (0.7) |
| | Acc I | 0.24 (0.03) | 2.7 (0.2) | 0.26 (0.03) | -4.0 (0.3) | 0.25 (0.03) | 0.1 (0.3) |
| <i>η</i> | - | - | - | 0.41 (0.26) | -37.3 (7.2) | 0.57 (0.45) | -91 (12.3) |
| <i>s_z</i> | - | - | - | 0.30 (0.06) | -200 (6.4) | 0.16 (0.16) | -63 (12.7) |
| <i>s_t</i> | - | - | - | 0.08 (0.03) | 0.5 (2.0) | 0.08 (0.05) | -5.2 (5.0) |

^a DMAT sets s^2 to 1, so all decision parameters were multiplied by 10 for consistency with the other methods.

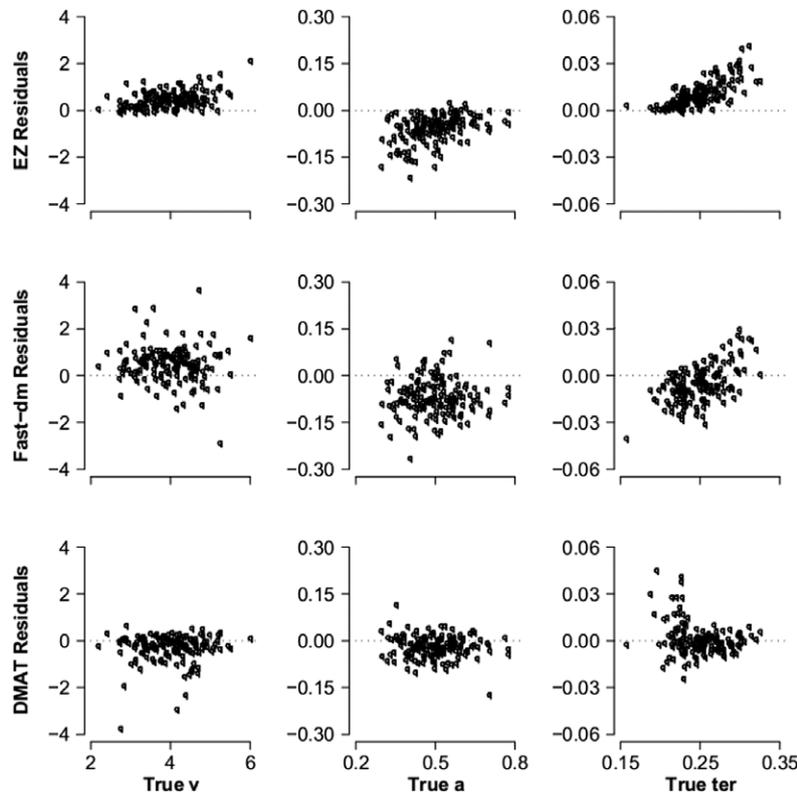


Fig. 2. Mean residuals for EZ (top row), fast-dm (middle row) and DMAT (bottom row), plotted against the magnitude of the true parameters for the Speed Compatible condition. The left column shows *v*, the middle column shows *a*, and the right column shows *T_{er}*. Residuals are calculated by subtracting the parameter estimates from the true parameter values.

Here, dx_1 and dx_2 refer to the changes per unit time in the two accumulators used in a two-choice task. The time unit dt is scaled by the time scale τ , dt/τ is fixed to 0.1. Accumulator 1 represents the correct response, its drift rate is $0.5(1 + v)$, whereas the drift rate of the competing accumulator is $0.5(1 - v)$. Thus, v represents the net drift rate, which is the difference between the drift rates of the two accumulators. Parameters k and β are the leakage and inhibition terms respectively, and ξ is the noise added at each time step. Negative values of x_i are truncated to 0.

We obtained initial parameter values for the simulation from Table 3 in Usher and McClelland (2001), which summarizes parameter estimates from an application of the model to a two-choice

task. The estimates come from five individuals and thus provide some rough indication of the standard deviation as well as the mean. We manually adjusted these values to obtain RTs and accuracies close to mean RTs and accuracies from the unpublished data by Wilhelm, Keye, and Oberauer. The parameters for the simulation are summarized in Table 6.

Individual differences were introduced by drawing 148 values for each parameter from a normal distribution with mean and standard deviation given in Table 6. Parameters for which negative values are meaningless were truncated at a low value (0.1 for θ , 0.05 for T_0 , 0.01 for σ^2 , and 0 for β and k). In practice, the truncation affected only 4 values of leakage k , and none of the

Table 6

The mean and SD of the LCA model parameters upon which the simulation dataset is based. v_c = net drift rate in compatible conditions, v_i = net drift rate in incompatible conditions, θ_{sp} = boundary in speed conditions, θ_{acc} = boundary in accuracy conditions, T_0 = time offset.

| | v_c | v_i | θ_{sp} | θ_{acc} | T_0 | β | k | σ^2 |
|------|-------|-------|---------------|----------------|-------|---------|------|------------|
| Mean | 0.8 | 0.6 | 0.6 | 1.2 | 0.25 | 0.7 | 0.1 | 0.4 |
| SD | 0.2 | 0.2 | 0.2 | 0.2 | 0.03 | 0.2 | 0.05 | 0.08 |

other parameters. Because these values were drawn independently for each parameter, the correlations between parameters were approximately zero.

We generated data by simulating, for each of the 148 subjects, 800 trials of a 2-AFC task under the four experimental conditions obtained by crossing a speed–accuracy manipulation (assumed to affect the value of the boundary θ) and a stimulus–response compatibility manipulation (assumed to affect the value of drift rate v). On each trial, the values of two accumulators were initialized at zero, and incrementally updated by Eq. (1) (Eq. (11) in Usher & McClelland, 2001) until one of them reached the boundary, at which point the number of time steps was converted into the corresponding value of seconds, and T_0 added to it to obtain RT. Accuracy was determined by taking the accumulator with the highest value as the response given. The simulation was implemented in Matlab (Mathworks, 1994).

We fitted the simulated data with the three methods for estimating diffusion-model parameters. The means and standard deviations of the parameter estimates are given in the left three columns of Table 7. The parameter estimates from all three methods were sensitive to the experimental manipulations in the expected way: The stimulus–response compatibility affected primarily drift rate, and the speed–accuracy instruction affected primarily boundary separation.

Table 7

Estimates of diffusion-model parameters from EZ, fast-dm and DMAT (with standard error of the mean in parenthesis) when applied to simulated data from the LCA and the LBA. Pars = Parameters, Con = Condition, Sp C = Speed Compatible, Sp I = Speed Incompatible, Acc C = Accuracy Compatible, Acc I = Accuracy Incompatible.

| Pars | Con | Simulated data from LCA | | | Simulated data from LBA | | |
|----------|-------|-------------------------|-------------|-------------|-------------------------|-------------|-------------|
| | | EZ | Fast-dm | DMAT | EZ | Fast-dm | DMAT |
| v | Sp C | 4.58 (1.59) | 5.09 (2.04) | 5.49 (2.59) | 3.36 (2.28) | 3.61 (3.70) | 6.26 (4.02) |
| | Sp I | 3.31 (1.33) | 3.50 (1.79) | 3.82 (1.88) | 2.73 (2.09) | 3.20 (2.75) | 5.50 (5.03) |
| | Acc C | 4.63 (1.50) | 5.00 (1.51) | 5.74 (2.33) | 2.58 (1.31) | 3.59 (1.52) | 5.38 (2.71) |
| | Acc I | 3.36 (1.30) | 3.52 (1.38) | 4.05 (2.06) | 2.22 (1.28) | 3.25 (1.46) | 4.51 (2.49) |
| a | Sp C | 0.59 (0.15) | 0.61 (0.16) | 0.65 (0.41) | 1.00 (0.32) | 0.88 (0.29) | 1.02 (0.77) |
| | Sp I | 0.57 (0.14) | 0.59 (0.16) | 0.58 (0.28) | 1.04 (0.32) | 0.94 (0.26) | 0.92 (0.63) |
| | Acc C | 1.02 (0.18) | 1.01 (0.16) | 1.33 (0.54) | 1.70 (0.24) | 1.57 (0.30) | 2.13 (0.94) |
| | Acc I | 0.98 (0.16) | 0.96 (0.15) | 1.22 (0.60) | 1.66 (0.23) | 1.64 (0.33) | 1.93 (0.77) |
| T_{er} | Sp C | 0.27 (0.04) | 0.29 (0.05) | 0.28 (0.04) | 0.15 (0.10) | 0.28 (0.09) | 0.23 (0.06) |
| | Sp I | 0.28 (0.04) | 0.29 (0.04) | 0.28 (0.04) | 0.13 (0.12) | 0.26 (0.06) | 0.23 (0.06) |
| | Acc C | 0.30 (0.04) | 0.31 (0.04) | 0.30 (0.04) | 0.25 (0.10) | 0.36 (0.07) | 0.36 (0.07) |
| | Acc I | 0.31 (0.04) | 0.32 (0.15) | 0.30 (0.04) | 0.24 (0.10) | 0.34 (0.08) | 0.37 (0.07) |
| η | – | – | 0.33 (0.15) | 0.78 (0.53) | – | 0.59 (0.57) | 2.01 (1.08) |
| s_z | – | – | 0.26 (0.07) | 0.04 (0.14) | – | 0.27 (0.16) | 0.40 (0.41) |
| s_t | – | – | 0.07 (0.03) | 0.07 (0.03) | – | 0.09 (0.04) | 0.09 (0.06) |

Table 8

Correlations between true parameters of the LCA model and recovered parameters of the diffusion model.

| Diffusion & LCA parameters | Condition | EZ | Fast-dm | DMAT |
|---|-----------|-----|---------|------|
| Drift rate (v) & Noise-scaled net drift rate (v/σ^2) | Sp C | .99 | .83 | .89 |
| | Sp I | .99 | .87 | .89 |
| | Acc C | .98 | .95 | .90 |
| | Acc I | .99 | .97 | .91 |
| | | | .93 | .87 |
| Boundary separation (a) & Noise-scaled boundary (θ/σ^2) | Sp C | .93 | .87 | .69 |
| | Sp I | .92 | .83 | .80 |
| | Acc C | .65 | .61 | .54 |
| | Acc I | .67 | .65 | .51 |
| | | | .96 | .73 |
| Non-decision time (T_{er}) & Time offset (T_0) | Sp C | .96 | .73 | .87 |
| | Sp I | .94 | .91 | .87 |
| | Acc C | .92 | .88 | .83 |
| | Acc I | .89 | .85 | .78 |
| | | | .89 | .85 |

As a first step to evaluate parameter recovery, we tested to what degree the three parameters of the LCA model, drift rate, boundary and time offset, correspond to the drift rate, boundary separation, and non-decision time parameters of the diffusion model, respectively. This investigation is complicated by the fact that, in the diffusion model, noise is a constant relative to which all parameters are scaled, whereas in the LCA model noise is a parameter that varies between individuals. To make parameters comparable across models, we divided the LCA parameters net drift rate (v) and boundary (θ) by noise (σ^2), thus expressing each individual's net drift rate and boundary relative to their level of noise. Table 8 shows the correlations between the diffusion-model estimates and the corresponding noise-scaled LCA parameters.

Whereas net drift rate (v) and time offset (T_0) are recovered very well, boundary (θ) did not correlate that well with the diffusion-model estimate of boundary separation (a) especially in the accuracy condition. One source of this lack of correspondence could be the role of LCA parameters that do not have corresponding parameters in the diffusion model, that is, inhibition (β) and leakage (k). To investigate this possibility, we regressed the diffusion-model parameter estimate of a on the corresponding LCA parameter θ together with k and β as predictors. Table 9 shows the results for the accuracy-compatible condition; a similar pattern of results was obtained for the other three conditions. It is clear that the estimated boundary separation a of the diffusion model reflects a linear combination of true parameters θ and k in the LCA model; the inhibition parameter β did not contribute significantly to the regression equation. We can understand the role of k from the dynamics of the LCA model. When k is high (i.e., a large amount of leakage), the diffusion process (i.e., the growth of x_i) decelerates as it approaches the asymptote. In the Ratcliff diffusion model with its linear growth, a roughly equivalent effect can be achieved by

Table 9

Regression of the recovered boundary separation (a) of the diffusion model on true parameters boundary (θ), leakage (k), and inhibition (β) in the LCA model (condition Acc-C).

| | EZ | Fast-dm | DMAT |
|--------------------------------|------|---------|------|
| R | .72 | .68 | .57 |
| Beta (true θ/σ^2) | .62 | .62 | .53 |
| beta (true k/σ^2) | .31 | .28 | .18 |
| beta (true β/σ^2) | -.03 | -.11 | -.03 |

Table 10

The mean and SD of the LBA model parameters upon which the simulation dataset is based. v_c = compatible drift rate, v_i = incompatible drift, b_{sp} = speed boundary, b_{acc} = accuracy boundary, t_0 = non-decision component, σ = variability in drift rate.

| | v_c | v_i | b_{sp} | b_{acc} | t_0 | σ |
|------|-------|-------|----------|-----------|-------|----------|
| Mean | 1.05 | 0.95 | 0.15 | 0.40 | 0.25 | 0.35 |
| SDs | 0.2 | 0.2 | 0.1 | 0.1 | 0.03 | 0.07 |

increasing boundary separation a . Therefore, if RTs are generated by a LCA process, individual differences in leakage are captured in the boundary separation estimate of the diffusion model.

Comparing the performance of the three methods of applying the diffusion model, we found that estimates from EZ produced slightly but consistently higher correlations to corresponding LCA parameters than the other two methods, which did not differ from each other in a consistent way across parameter estimates (see Table 8). To summarize, if data are generated by a process as described by the LCA, the diffusion-model parameter estimates of drift rate, non-decision time and (with some reservations) boundary separation are still highly valid measurements of individual differences in the corresponding variables in LCA.

4.2. Fitting data generated by the LBA model

Like the LCA model, the LBA model by Brown and Heathcote (2008) assumes that evidence for each response alternative is collected in a separate accumulator, and therefore the model can be applied to choices between any number of alternatives. Like the Ratcliff diffusion model, the LBA assumes that accumulation is linear in time. Unlike Ratcliff's model, however, the accumulation process itself is ballistic, that is, it is not continuously modulated by noise. Drift rate (v) varies randomly from trial to trial according to a Gaussian distribution with mean zero and standard deviation (σ). The process stops with a decision once the first accumulator reaches a response threshold (b). Additional trial-to-trial variability arises from variation in the starting point of each accumulator, which varies between 0 and A , where A is expressed as a proportion of b .

We generated data from the LBA using the parameter values in Table 10; these values are informed by the best-fitting estimates from three experiments reported in Table 1 of Brown and Heathcote (2008), and manually adjusted to reproduce the empirical data. All parameters varied between individuals except A , which was fixed to 1 in the speed conditions and to 0.4 in the accuracy conditions. Variation between individuals was again created by drawing 148 samples from normal distributions for each parameter, using the means and standard deviations in Table 10. Values of b and t_0 were truncated at 0.05, and values of σ at 0.01. In practice, this affected only 26 values of b_{sp} , and none of the other parameters. As for the LCA, all parameter values were drawn independently and therefore correlated approximately zero.

We generated data by simulating for each of the 148 subjects 800 trials in each of the 4 experimental conditions. The simulation of each trial proceeded as follows: Starting points k_1 and k_2 for the two accumulators were drawn from a rectangular distribution

between 0 and A . The drift rates d_1 and d_2 of this particular trial were drawn from a normal distribution with mean v for the accumulator representing the correct response, and mean $1 - v$ for the accumulator representing the incorrect response (Brown & Heathcote, 2008, p. 161), both with standard deviation σ . The time to reach the threshold was computed as $(b - k_i)/d$ for each accumulator (Brown & Heathcote, 2008, p. 158). When a drift rate was negative (implying that the accumulator would never reach the threshold), we set the time to 1000 s. We determined the smaller of the two times as the decision time. In cases where both decision times exceeded a deadline of 3 s, that deadline was chosen as the decision time to avoid extreme times (this also applied to trials where both drift rates were negative). This happened on 0.5% of all trials. A trial was counted as correct if the accumulator representing the correct response had the shorter decision time.

Parameter estimates were then obtained by applying the three diffusion-model methods to the simulated data. The means and standard deviations of the parameter estimates can be found in the right three columns of Table 7. The experimental manipulations generated the expected effects on these parameter estimates only partially: The speed-accuracy manipulation affected boundary separation, but also had an effect on non-decision time estimates. The stimulus-response compatibility manipulation had only a very small effect on drift rates.

We evaluated the three methods by correlating the diffusion-model parameter estimates with corresponding true parameters of the LBA. Specifically, we correlated diffusion drift rate (v) with LBA drift rate (v), boundary separation (a) with response threshold (b), non-decision time (T_{er}) with non-decision time (t_0), and trial-to-trial variability in drift rate in the diffusion model (η) with the corresponding parameter in the LBA model (σ). Table 11 shows the results. The estimates of none of the three methods reflected the corresponding true parameters particularly well. For all three methods, estimates of diffusion drift rate reflected some variance of the LBA drift rate, and estimates of diffusion boundary separation reflected some variance of the LBA response threshold. The non-decision component was recovered only by DMAT with reasonable accuracy.

What, then, do the parameter estimates reflect? Table 12 gives a partial answer. The table summarizes the results of regression analyses, predicting the diffusion-parameter estimates for v , a , and T_{er} by a range of true parameter values of LBA. We present only the results for one of the four conditions because it was representative for the pattern of beta-weights obtained in all four conditions.

The diffusion-model drift rate (v) is a weighted function of drift rate (v) and variability in drift rate (σ) in LBA, with higher σ being reflected in lower diffusion-model v estimates. The diffusion-model boundary separation (a) depends on different combinations of true parameters according to the three methods. The EZ model computes a higher value of a when the LBA boundary (b) is high, and when variability in drift rate (σ) is high; fast-dm tends to behave in the same way. DMAT estimates a higher value of a when LBA parameters b and v are high. Finally, the non-decision parameter T_{er} as computed by EZ depends primarily on the LBA value of variability in drift rate σ . When estimated by fast-dm or DMAT, T_{er} seems to depend to some degree on all four LBA parameters.

Comparing the three methods, none of the models fared well in recovering individual LBA parameters by corresponding diffusion-model parameters. DMAT did better than the other two methods in recovering t_0 by T_{er} . EZ again outperformed the two competitors in terms of its multiple correlations with true parameters of the LBA model. Thus, the EZ parameters, although not reflecting a single LBA parameter, can at least be interpreted as good estimates of linear combinations of two LBA parameters.

To summarize, when data are generated by a process as described by LBA, diffusion-model parameter estimates of drift

Table 11

Correlations between true parameters of the LBA model and recovered parameters of the diffusion model.

| Diffusion & LBA parameters | Condition | EZ | Fast-dm | DMAT |
|---|-----------|-----|---------|------|
| Drift rate (v) & drift rate (v) | Sp C | .66 | .45 | .63 |
| | Sp I | .67 | .60 | .52 |
| | Acc C | .76 | .76 | .67 |
| | Acc I | .79 | .84 | .69 |
| Boundary separation (a) & Response threshold (b) | Sp C | .58 | .45 | .40 |
| | Sp I | .44 | .40 | .41 |
| | Acc C | .63 | .55 | .42 |
| | Acc I | .59 | .50 | .50 |
| Non-decision time (T_{er}) & Non-decision component (t_0) | Sp C | .29 | -.04 | .66 |
| | Sp I | .26 | .46 | .63 |
| | Acc C | .27 | .40 | .43 |
| | Acc I | .26 | .37 | .52 |
| Drift rate variability (η & σ) | | - | .38 | .25 |

Table 12

Regression of diffusion-model parameter estimates on true parameters of the LBA model (condition Acc C).

| Criterion | Regression estimate | EZ | Fast-dm | DMAT |
|-----------|-----------------------|------|---------|------|
| v | R | .96 | .88 | .75 |
| | beta (true v) | .79 | .79 | .69 |
| | beta (true σ) | -.55 | -.32 | -.31 |
| | beta (true b) | -.18 | -.30 | -.15 |
| a | beta (true t_0) | -.02 | .03 | 0 |
| | R | .82 | .64 | .63 |
| | beta (true b) | .62 | .55 | .38 |
| | beta (true v) | .20 | .16 | .42 |
| T_{er} | beta (true σ) | .47 | .27 | -.20 |
| | beta (true t_0) | .06 | -.02 | .09 |
| | R | .91 | .73 | .70 |
| | beta (true t_0) | .26 | .38 | .43 |
| | beta (true b) | .33 | .47 | .34 |
| | beta (true v) | .13 | -.26 | -.37 |
| | beta (true σ) | -.79 | -.29 | -.24 |

rate and boundary separation still retain some limited validity as measurements of drift rate and response threshold, respectively, though contaminated by contribution from other LBA parameters. Estimates of non-decision time have some degree of validity only when obtained by DMAT.

5. Discussion

Both experimental cognitive psychology and the psychometrics of mental speed rely heavily on measurements of RT. A decade-old conundrum is that there was no measurement model for RTs that is well supported by data and can be applied to data sets without specialized technical knowledge. The advent of a new generation of algorithms for fitting the diffusion model to data, or computing its parameters directly from the data (as in the case of the EZ model), promises to deliver such a practically applicable measurement model. Here we comparatively evaluated how well EZ, fast-dm, and DMAT could recover diffusion model parameter values from a simulated data set. We had two criteria: their ability to accurately reflect individual differences in model parameters and their ability to recover the experimental effects on parameter means. In addition, we investigated what the diffusion-model parameter estimates obtained by the three methods reflect when the data were generated by a process different from the one described by Ratcliff's diffusion model.

5.1. Summary of findings

Regarding individual differences, the results show that EZ recovers drift rate, boundary separation and non-decision time in the dataset reasonably well. EZ did consistently better than fast-dm and DMAT, both with large data sets (800 trials per condition) and

with small data sets (80 trials per condition). There was no consistent difference between fast-dm and DMAT in how well their parameter estimates correlated with true parameter values.

Both fast-dm and DMAT are incapable of recovering the individual differences in the dispersion parameters for drift rate and for starting point, η and s_z , respectively. EZ provides no estimate of the dispersion parameters. Thus, at present there is nothing to be gained from these dispersion parameters when it comes to measuring individual differences in cognitive processes. One possible reason for the poor recovery of these two dispersion parameters is that they were highly correlated with their corresponding mean parameters in Simulation 1. As a consequence, they added little unique variance to the simulated data, thereby producing only a weak signal to be picked up by the parameter computation or fitting procedures. We tested this explanation by running a further simulation identical to Simulation 1 but with all parameters uncorrelated. The results were essentially the same, including the poor recovery of η and s_z . This result rules out parameter correlations as a cause of the poor recovery of the dispersion parameters.

Regarding parameter means, EZ has a bias to underestimate both drift rate and non-decision time, whereas it overestimates boundary separation. While the bias is modest (between 2% and 13%), it is systematically present. This bias has already been documented by Wagenmakers and Brown (2007), and it is due to the fact that EZ simplifies the diffusion model, implicitly assuming that all dispersion parameters are zero. This implies that the effect of the dispersion parameters is picked up by the mean parameters, generating a systematic bias in them. Other than that, EZ captures the mean structure of the data well, showing mean parameter differences between experimental conditions in the correct directions. DMAT tends to overestimate both drift rate and boundary separation, whereas it underestimates non-decision time. Its bias is smaller than that of the other estimation methods and the size of the bias is not linearly related to the size of the true parameter. DMAT is also well capable of capturing mean parameter differences between experimental conditions. However, DMAT does have problems with low values of the non-decision time. Also, in the accuracy-compatible condition, the bias in parameter estimates of DMAT was relatively large. Fast-dm seems to do relatively poorly with respect to estimation bias. With an underestimation of up to 20% of drift rate in the speed-incompatible condition, and a tendency for parameter values in different conditions to converge towards each other, the fast-dm method yields smaller differences between conditions than are actually there in the simulated data.

When fitting data that are based on the leaky, competing accumulator model by Usher and McClelland (2001) with the diffusion model, we found that both drift rate and non-decision time are recovered very well. The EZ-model seems to fit best, with little difference between fast-dm and DMAT. The correspondence between boundary separation in the diffusion model and boundary in LCA was not as close as for the other two parameters, especially in

the accuracy condition. Regression analysis showed that boundary separation a in the diffusion model must be interpreted as a combination of boundary (θ) and leakage (k) if the data are assumed to be generated by a LCA process.

There seems to be no simple one-to-one mapping of diffusion model parameters and parameters of the linear ballistic accumulator process by Brown and Heathcote (2008). Regression analyses revealed that the diffusion drift rate (v) is primarily a linear combination of the LBA drift rate (ν) and between-trial variance in drift rate (σ). Boundary separation of the diffusion model (a) mainly reflects boundary (b) and between-trial variance in drift rate (σ) in the LBA model. Non-decision time (T_{er}) seems to be a combination of all four LBA parameters.

5.2. Conclusions

What do we learn from these results? If researchers are willing to assume that their data have actually been generated by a diffusion process similar to the one described in Ratcliff's diffusion model, they can use one of the three routines to measure the model's parameters with reasonable accuracy, provided they have sufficient data points from each individual in each condition. Whereas DMAT requires relatively large numbers of data points, EZ and fast-dm provide useful estimates even with as little as 80 trials per condition. As always, more data points lead to more accurate measurement; we have not mapped out the increase in measurement precision with increasing sample size systematically, but it is probably safe to assume that the unsystematic component of measurement error (i.e., the standard deviation of the residuals, which is responsible for the less-than-perfect correlation between true parameters and estimates) decreases with the square root of the number of trials (Ratcliff & Tuerlinckx, 2002), whereas the systematic component (i.e., the bias) remains unaffected by the number of trials. Therefore, increasing the number of trials (beyond about 100) is probably worth the effort in particular when individual differences are of interest.

One issue not addressed in the present work is the distortion of parameter estimates due to trade-offs between parameters. Such trade-offs are likely to be responsible for a substantial portion of the inaccuracy in parameter recovery. A partial solution for this problem has been developed by Schmiedek et al. (2007), who showed that structural equation models can be used to obtain latent factors of diffusion-model parameters that reflect the true covariance between parameters without distortion by parameter trade-offs.

Among the three competing routines for estimating the diffusion-model parameters, none is the best for all purposes. EZ and DMAT are better capable than fast-dm of reflecting experimental effects on parameters. When the researcher is interested in an unbiased estimate of parameter means, DMAT is the best option, though occasional biases in individual conditions may arise. When differences between experimental conditions are important, EZ or DMAT are to be preferred. When individual differences in parameter values are of primary interest, EZ is superior to both fast-dm and DMAT, in particular with smaller numbers of trials.

If researchers are not willing to assume that their data are generated by a process as specified by the diffusion model, the diffusion-parameter estimates could still provide useful information. We have shown that the three main diffusion-model parameters provide good estimates for the corresponding parameters in the LCA model (with the proviso that boundary separation a also reflects variance in leakage). However, if the data are assumed to be generated by a different model such as the LBA model, interpretation of the diffusion-model parameters is no longer straightforward. Each parameter estimate still reflects some aspect of performance, but the parameters do not decompose the variance in

the data into theoretically meaningful components. We conclude that the validity of diffusion model parameters does not depend on the correctness of all assumptions of the Ratcliff diffusion model, but on the correctness of a set of relatively general assumptions that are shared between the diffusion model and the LCA model but not the LBA model.

One feature that the Ratcliff diffusion model and LCA have in common that distinguishes them from LBA is that the first two models assume a noisy accumulation process. Variability between trials arises, to a large degree, from that noise. The contribution of inter-trial variability in parameter values (i.e., η , s_z , and s_t) is relatively small and can be ignored with little loss, as witnessed by the success of EZ. In LBA, in contrast, all variability between trials arises from variability of drift rate and the starting point; these parameters therefore have a huge impact on the shape of the RT distributions. The three methods of estimating diffusion model parameters do not capture these effects by the corresponding diffusion model parameter (e.g., mapping σ of LBA onto η of the diffusion model), but rather use a mix of parameters to model them.

Thus, although interpretation of diffusion model parameters is not bound to the correctness of Ratcliff's diffusion model, it is contingent on the validity of at least some theoretical assumptions. This is an unsurprising conclusion; any measurement model that aspires to estimate theoretically interpretable variables must fail if the underlying theory is seriously wrong. The only way to address this limitation is to search for evidence to decide between competing models.

To conclude, estimating the parameters of mathematical models has long been known to be tricky and technically demanding, in particular when the aim is to obtain precise estimates for individual subjects, as is necessary in individual-differences research. Significant progress has been made in recent years in developing methods for obtaining parameter estimates for the diffusion model which are fast and easy to use. Our analysis shows that, with some notable exceptions, these methods also recover the true parameters quite accurately, even with relatively small numbers of data points. In response to some criticism of the EZ model (Ratcliff, 2008, but see Wagenmakers, van der Maas, Dolan, & Grasman, 2008), an EZ2 model is in the making (Grasman et al., 2009). At the same time, Vandekerckhove, Tuerlinckx, and Lee (submitted for publication) are working on a hierarchical version of the diffusion model. With such new diffusion model implementations in hot pursuit, we can be optimistic that both experimental and individual-differences research will soon have a formidable toolbox of measurement models for analyzing and interpreting RT data at their disposal. The times when disciples of mental chronometry got little more out of their efforts than estimates of mean RT seem to finally be coming to an end.

References

- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071.
- Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.
- Danthiir, V., Roberts, R. D., Schulze, R., & Wilhelm, O. (2005). Mental speed: On frameworks, paradigms, and a platform for the future. In O. Wilhelm, & R. W. Engle (Eds.), *Understanding and measuring intelligence* (pp. 27–46). Thousand Oaks, CA: Sage.
- Dennis, I., & Evans, J. S. B. T. (1996). The speed–error trade-off problem in psychometric testing. *British Journal of Psychology*, *87*, 105–129.
- Fry, A., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence. *Psychological Science*, *7*, 237–241.
- Grasman, R. P. P., Wagenmakers, E.-J., & van der Maas, H. L. J. (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.
- 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*(3), 353–368.

- Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence, 14*, 309–325.
- Luce, R. D. (1986). *Response times*. New York: Oxford University Press.
- Mathworks. (1994). <http://www.mathworks.com/products/matlab/>.
- 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.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review, 95*, 238–255.
- Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9*, 278–291.
- Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review, 15*, 1069–1084.
- Ratcliff, R., Gomez, P., & McKoon, G. (2004). Diffusion model account of lexical decision. *Psychological Review, 11*, 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., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333–367.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*(3), 438–481.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.
- van Ravenzwaaij, D., van der Mass, H., & Wagenmakers, E. -J. (2009). Does the racial Implicit Association Test measure prejudice?. Manuscript (submitted for publication).
- R Development Core Team, (2004). *R: A language and environment for statistical computing*. Vienna: Author.
- Salthouse, T. A. (1998). Relation of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence, 26*, 153–166.
- 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.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review, 102*, 567–591.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550–592.
- Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & 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. (2009). Hierarchical diffusion models for two-choice response times. Manuscript (submitted for publication).
- 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. (2008). A fast numerical algorithm for the estimation of diffusion-model parameters. *Journal of Mathematical Psychology, 52*, 1–9.
- Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39*(4), 767–775.
- 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., van der Maas, H. L. J., Dolan, C., & Grasman, R. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review, 15*, 1229–1235.
- Wagenmakers, E.-J., van der Maas, H. L., & Grasman, R. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*, 3–22.
- 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.
- Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice–reaction-time tasks. *European Journal of Cognitive Psychology, 18*, 18–50.