Journal of Mathematical Psychology 53 (2009) 463–473

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

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.

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).

0022-2496/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2009.09.004

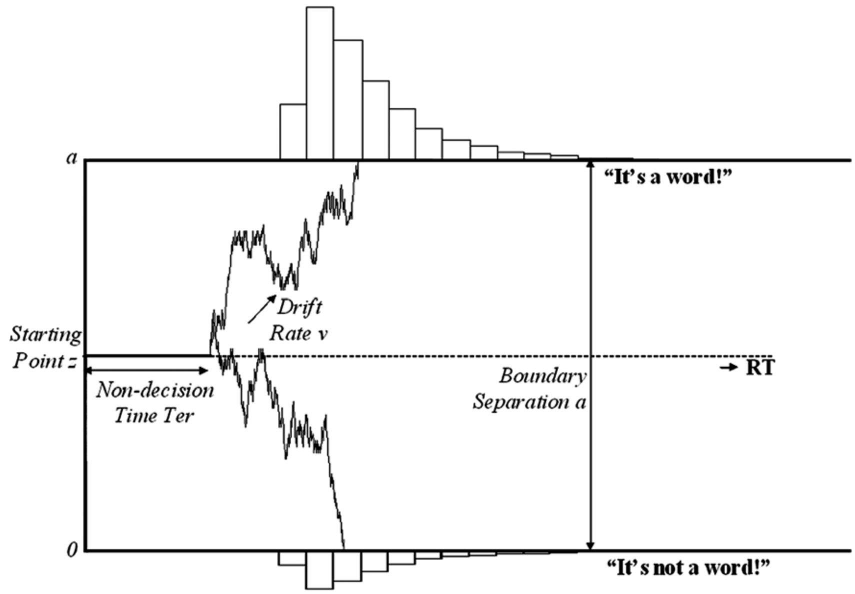

Fig. 1. Illustration of the diffusion model and parameters. The panel shows the drift rate (v), boundary separation (a), non-decision time (T_{er}), starting point (z) and illustrates how accuracy and response times for correct and error responses change as a function of the noise component within each 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_t). 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_t). 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 toolbox[4] (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/Members/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

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)

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 *mvrnorm* (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 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

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.

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
v	Sp C	.85	.70	.75
	Sp I	.93	.70	.83
	Acc C	.96	.87	.88
	Acc I	.97	.95	.91
a	Sp C	.89	.85	.95
	Sp I	.92	.84	.97
	Acc C	.92	.90	.45
	Acc I	.94	.86	.62
T_{er}	Sp C	.97	.92	.95
	Sp I	.97	.90	.95
	Acc C	.98	.95	.83
	Acc I	.98	.95	.95
η	–	–	.15	.13
s_z	–	–	−.08	.19
s_t	–	–	.86	.48

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
v	Sp C	.77	.49
	Sp I	.86	.59
	Acc C	.85	.62
	Acc I	.91	.78
a	Sp C	.83	.75
	Sp I	.85	.75
	Acc C	.73	.64
	Acc I	.77	.66
T_{er}	Sp C	.94	.87
	Sp I	.94	.85
	Acc C	.88	.83
	Acc I	.86	.88
η	–	–	.04
s_z	–	–	−.01
s_t	–	–	.71

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 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:

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

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
v	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)
a	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)
T_{er}	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)
η	–	–	–	0.41 (0.26)	−37.3 (7.2)	0.57 (0.45)	−91 (12.3)
s_z	–	–	–	0.30 (0.06)	−200 (6.4)	0.16 (0.16)	−63 (12.7)
s_t	–	–	–	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