

Speed-accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates?

Veronika Lerche¹ · Andreas Voss¹

Published online: 14 March 2018 © Psychonomic Society, Inc. 2018

Abstract

The diffusion model (Ratcliff, 1978) is a mathematical model theorized to untangle different cognitive processes involved in binary decision tasks. To test the validity of the diffusion model parameters, several experimental validation studies have been conducted. In these studies, the validity of the threshold separation parameter was tested with speed–accuracy manipulations. Typically, this manipulation not only results in the expected effect on the threshold separation parameter but it also impacts nondecision time: Nondecision time is longer in the accuracy than in the speed condition. There are two possible interpretations of the finding: On the one hand, it could indicate that speed versus accuracy instructions really have an impact on the duration of extradecisional processes. On the other hand, the effect on the measure for nondecision time could be spurious—that is, based on a problem in the parameter estimation procedures. In simulation studies—with the parameter sets based on typical values from experimental validation studies—we checked for possible biases in the parameter estimation. Our analyses strongly suggest that the observed pattern (i.e., slower nondecision processes under accuracy instructions) is attributable to a lack of discriminant validity of the manipulation rather than to trade-offs in the parameter estimations.

Keywords Diffusion model \cdot Mathematical models \cdot Reaction time methods \cdot Fast-dm

When performing a speeded decision task, individuals can prioritize either accurate or fast responses. Emphasis in one domain will result in worse performance in the other domain, producing the ubiquitous speed-accuracy trade-off (SAT; Heitz, 2014). The SAT can complicate data interpretations. For example, older adults tend to produce longer response times (RTs; e.g., Ratcliff, Thapar, & McKoon, 2001; Spaniol, Voss, & Grady, 2008; Thapar, Ratcliff, & McKoon, 2003). One explanation for this finding is that older adults are slower at accumulating the information that they need to make a decision. On the other hand, it is possible that these differences in RTs are attributable to differences in speed-accuracy settings, with older adults favoring accuracy over speed. Simply analyzing mean RTs or accuracy rates alone can lead to misinterpretations, because it is uncertain whether the results reflect genuine task difficulty or ability to perform the task, or some tendency toward favoring speed or accuracy. Several integrated measures of RT and accuracy rate have been designed to provide solutions to this problem (e.g., Bruyer & Brysbaert, 2011; Hughes, Linck, Bowles, Koeth, & Bunting, 2014; see also Vandierendonck, 2017). A different approach to disentangling speed–accuracy settings from the speed of information accumulation, and also from other processes (e.g., motor responses and decision biases), is the diffusion model.

The diffusion model (Ratcliff, 1978) is a mathematical model that allows the disentangling of different processes involved in binary decision tasks. Studies employing the diffusion model have revealed differences in speed–accuracy settings between younger and older adults (e.g., Ratcliff et al., 2001; Spaniol et al., 2008; Thapar et al., 2003). An analysis of the behavioral data alone could lead to the assumption that older adults are inferior in their information-processing capability, due to their longer RTs. However, for many tasks no differences in speed of information accumulation are reported between the different age groups (Mulder et al., 2010; Ratcliff et al., 2001; Ratcliff, Thapar, & McKoon, 2003; but see Thapar et al., 2003). Instead, older adults set more conservative decision thresholds, favoring accuracy over speed, and

Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Hauptstrasse 47–51, D-69117 Heidelberg, Germany



 [∨] Veronika Lerche veronika.lerche@psychologie.uni—heidelberg.de

also display longer durations of nondecision processes such as motor response execution (e.g., Ratcliff et al., 2001; Spaniol et al., 2008; Thapar et al., 2003).

To test whether the diffusion model is able to disentangle the specific processes (such as speed of information accumulation and speed-accuracy settings), several experimental validation studies have been conducted (e.g., Arnold, Bröder, & Bayen, 2015; Voss, Rothermund, & Voss, 2004). For example, difficulty manipulations have been employed to analyze the convergent validity of the model parameter assumed to measure the speed of information accumulation. This manipulation revealed that, as expected, the parameter differed between easier and more difficult trials, providing strong support that the parameter measures what it was designed to measure. Less clear are the findings regarding the parameter that measures speed-accuracy settings. This parameter has been investigated by encouraging participants to prioritize speed or accuracy, respectively, in two blocks of the same task. Providing evidence for convergent validity, speed-accuracy instructions influenced the correct model parameter in the expected direction (e.g., Ratcliff & Rouder, 1998; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). However, in several studies, if other diffusion model parameters were also allowed to vary between the two conditions, the manipulation also unexpectedly affected the estimate for nondecision time (e.g., Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, 2004; Voss et al., 2004). Nondecision time comprises the time needed for encoding information and for executing the response. In these studies, the nondecision time was higher in the accuracy than in the speed condition.

Two different explanations have been proposed for this finding: (1) The parameter estimation procedures might have found it difficult to disentangle decision-based processes from nondecision processes, or (2) the speed–accuracy manipulations might lack discriminant validity, such that instructions to emphasize speed or accuracy might influence not only the speed–accuracy settings but also the speed of the motoric response or the speed to encode information. Even though the possibility of these two different accounts has been recognized (e.g., Dutilh et al., 2018), so far, to the best of our knowledge, no study has tried to uncover which account is more likely. To address this question, this article reports two simulation studies differing in their parameter values and data generation processes.

Note that even though the studies in this article are based on the diffusion model (Ratcliff, 1978), the findings are similarly relevant for the assessment of experimental studies using other sequential-sampling models, such as the popular linear ballistic accumulator model (LBA; Brown & Heathcote, 2008). The LBA model, amongst others, also includes a nondecision time estimate and a parameter measuring decision settings, which are both

conceptually similar to aspects of the diffusion model context (Donkin, Brown, Heathcote, & Wagenmakers, 2011).

In the following sections, we first give a short introduction to the diffusion model, followed by a summary of diffusion model studies that have analyzed the data of speed–accuracy manipulations. Finally, we present the method and results of our two simulation studies.

Introduction to diffusion modeling

The diffusion model (Ratcliff, 1978) is a mathematical model applicable to RT data from binary decision tasks. An example of such a binary task is a color discrimination task with the two stimuli "orange" and "blue" (Voss et al., 2004). In each trial, a square composed of pixels of the two colors is presented, and participants have to assess whether orange or blue dominates. The diffusion model is illustrated in Fig. 1. In this plot, a corridor is shown that is limited by two thresholds. The thresholds are associated with two colors (in the example, "orange" and "blue," respectively). Underlying the diffusion model is the assumption that in tasks like this simple color discrimination task, participants accumulate information continuously until one of the two thresholds is reached. In the example trial illustrated in the plot, the process ends at the upper threshold associated with "orange." Accordingly, after the end of the decisional process, the participant will execute the motoric response associated with the answer "orange" (e.g., left key press).

The *basic diffusion model* is made up of four main parameters. The *threshold separation* (a) defines the amount of

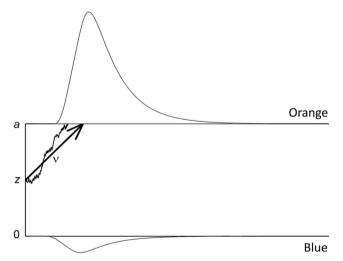


Fig. 1 Illustration of the decision process of the diffusion model, based on a color discrimination task. The process starts at the starting point z. In this example there is no decision bias, as indicated by the centered starting point. The process moves with drift rate ν until one of the two thresholds associated with the response options "orange" and "blue" has been met. Not depicted in the plot are nondecisional processes (t_0) that add to the decision process, as well as the intertrial variabilities $(s_{\gamma}, s_{z\gamma}, s_{t0})$



information that is required for the participant to make a decision. If participants adopt a conservative response strategy, the threshold separation is increased. This results in a longer time course to reach one of the two thresholds, but will produce more accurate responses. Threshold separation has been found to be higher under accuracy than under speed instructions (e.g., Ratcliff & Rouder, 1998; Voss et al., 2004) and for older than for younger individuals (e.g., Ratcliff et al., 2001; Spaniol et al., 2008; Thapar et al., 2003).

In Fig. 1, the accumulation process is directed to the upper threshold, as illustrated by the arrow symbolizing the drift rate (ν). Drift rate measures the direction and speed of information accumulation. If the drift is stronger (i.e., the arrow is steeper), the process is more likely to reach this threshold in a shorter time. Higher drift values, thus, indicate that the process is faster and ends more frequently at the correct threshold. It has been shown that easier trials feature higher drift rates (e.g., Ratcliff, 2014; Voss et al., 2004), as do more intelligent individuals (Ratcliff, Thapar, & McKoon, 2010; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015; Schulz-Zhecheva, Voelkle, Beauducel, Biscaldi, & Klein, 2016). The decision process is influenced not only by the drift rate but also by additive Gaussian noise (the standard deviation of this noise is termed the diffusion constant and is often described as a scaling parameter of the diffusion model). Due to this noise, process durations will vary from trial to trial and processes will not all necessarily end at the threshold that the drift points toward. The figure also depicts a density distribution at each threshold indicating how often, and after how much time, a process will end at one of the two thresholds, given that the correct response is orange.

The accumulation process begins at the *starting point z* (or, the relative starting point $z_r = z/a$). The decision process is unbiased if the starting point is centered between the two thresholds, as in Fig. 1. However, if the starting point is closer to, for example, the upper threshold, it is more likely that the process will end at this threshold. In this case, the mean RTs at the threshold closer to the starting point will be shorter than mean RTs at the opposite threshold. The starting point has been the focus of a number of diffusion model studies as a measure of motivational biases in perception (e.g., Germar, Schlemmer, Krug, Voss, & Mojzisch, 2014; Voss, Rothermund, & Brandtstädter, 2008).

Finally, the duration of *nondecisional processes* (t_0), such as the encoding of information and the motoric response execution, is added to the duration of the decision process. In addition to the four main parameters of the simple diffusion model, the *full diffusion model* includes three more parameters: the intertrial variabilities of drift rate (s_v), starting point (s_{zr}), and nondecision time (s_{t0} ; e.g., Ratcliff & Rouder, 1998; Ratcliff & Tuerlinckx, 2002).



Speed–accuracy manipulations have been employed frequently in diffusion model studies (e.g., Ratcliff & Rouder, 1998) and in studies based on other sequential sampling models (e.g., Forstmann et al., 2011). In the speed conditions, participants are told to respond as fast as possible (even if this leads to an increased number of errors). Often, for slow responses (e.g., RT > 550 ms; Ratcliff & Rouder, 1998), participants get the feedback that their response was not sufficiently fast (e.g., feedback message "too slow"). In the accuracy condition, on the other hand, participants are instructed to be as accurate as possible (even if this leads to slower responses). In this condition, error feedback is typically provided.

In many diffusion model studies that used speed–accuracy manipulations, only the threshold separation (sometimes also the starting point) was allowed to vary between the conditions (Ratcliff & Rouder, 1998; Ratcliff et al., 2001, 2003; Ratcliff, Thapar, & McKoon, 2004; Thapar et al., 2003; Wagenmakers et al., 2008). These studies have consistently observed the expected effect of the speed–accuracy manipulation on the threshold separation parameter: The threshold separation was higher in the accuracy than in the speed condition.

Critically, constraints in the parameter estimation approach can distort results. Any existing differences between conditions regarding a specific parameter (e.g., nondecision time) cannot be detected if this parameter is not allowed to vary between conditions. In this case, the effects of condition will be spuriously picked up by other parameters that are estimated separately for each condition. Thus, of particular interest are studies in which not only the threshold separation but also other model parameters are allowed to vary between the speed and the accuracy conditions. In these studies, the speed-accuracy manipulations have usually influenced not only the threshold separation, but also the nondecision time estimate (e.g., Arnold et al., 2015; Dutilh et al., 2018; Voss et al., 2004). For example, Voss et al. (2004) conducted an experimental validation study using a color discrimination task in a within-subjects design. In the accuracy condition, participants were instructed to work particularly carefully trying to avoid errors. The RT limit was set at 3 s, in comparison to 1.5 s in the baseline condition. As expected, participants exhibited a higher threshold separation in the accuracy than in the baseline condition. However, additionally, nondecision time was increased by 50 ms.

Similarly, in a study based on a between-subjects design and a recognition memory paradigm, Arnold et al. (2015, Study 2) provided participants with negative feedback for errors in the accuracy group. In the speed group, on the other hand, negative feedback was given for responses slower than 1 s. Arnold et al. estimated the diffusion model parameters using several different methods: *fast-dm* (Voss & Voss, 2007, 2008), *DMAT* (Vandekerckhove & Tuerlinckx, 2008),



and EZ (Wagenmakers, van der Maas, & Grasman, 2007). All three methods found the expected difference between groups in the threshold separation parameter, but the analyses with fast-dm and DMAT revealed a significant difference between groups in nondecision time, as well. Specifically, the nondecision time estimated by fast-dm was 130 ms longer in the accuracy than in the speed condition, and DMAT resulted in a difference in nondecision time of 170 ms. With the EZ method, no significant effect on nondecision time was observed (difference in t_0 of about 25 ms).

In some studies, speed-accuracy manipulations have influenced other parameters in addition to nondecision time. Lerche and Voss (2017a, Study 1) observed effects of a speed-accuracy manipulation on threshold separation, nondecision time, and the intertrial variability of nondecision time for a slow binary decision task (with mean RTs of about 7 s). Furthermore, in a study by Rae, Heathcote, Donkin, Averell, and Brown (2014), the speed–accuracy manipulation not only affected the threshold separation and nondecision time, but also the drift rate, with a higher drift value in the accuracy than in the speed condition (see also, e.g., Starns, Ratcliff, & McKoon, 2012). The authors assumed that under accuracy instructions, participants put more effort into the task, resulting in the higher drift rate in this condition. Interestingly, the opposite effect was reported by Arnold et al. (2015). The EZ parameter estimation revealed a higher drift rate in the speed than in the accuracy condition. The same effect in drift rate was reported for the fast-dm estimation, but only for the new items of the recognition memory task.

The finding that speed-accuracy manipulations often lead to effects in several parameters, and not only—as expected in threshold separation, has also been discussed in the context of a large-scale validation project (Dutilh et al., 2018). Gilles Dutilh and Chris Donkin had a total of 17 research teams analyze the data from 14 pseudo-experiments. The teams were provided data in which zero, one, two, or three of the psychological constructs had been experimentally manipulated. These constructs comprised (1) the ease of processing of information (i.e., the drift rate in the diffusion model framework), (2) decisional caution (threshold separation), and (3) decisional bias (starting point). No manipulation was implemented to selectively affect nondecision time. The research teams were blind with regard to the manipulations and could use an information-sampling model of their own choice to find out which of the four psychological constructs (ease of processing information, decisional caution, decisional bias, or nondecision time) was manipulated in each experiment. Ten teams opted for a diffusion model analysis (using either the simple or the full diffusion model), five teams applied the

LBA model, and two teams used a model-free, heuristic approach.

In the nine studies with a speed–accuracy manipulation, on average 52.9% of the teams observed the typical effect in nondecision time (i.e., a longer nondecision time in the accuracy than in the speed condition). This percentage was even higher (72.2%) if only the ten teams are considered that based their decisions on a diffusion model analysis. Furthermore, in the three experiments with a manipulation of threshold separation and no simultaneous manipulation of drift rate, 62.7% of the teams found higher drift rates under accuracy than under speed conditions. This percentage is smaller (50.0%) if only the diffusion model teams are considered.

To account for the possibility that speed–accuracy manipulations might also influence nondecision time and drift rate, Dutilh et al. (2018) used different scoring keys. In the original scoring key, the speed–accuracy manipulation was expected to affect only the threshold separation. Thus, any other effect was coded as a false alarm. In another scoring key, an effect on nondecision time (longer nondecision times in the accuracy condition) was coded as a hit, and thus the research teams who did not report this effect had a miss. Finally, in a third scoring key, an effect on drift rate (higher drift rate in the accuracy condition) was regarded as correct, but effects on nondecision time were false alarms.

As the mere existence of these different scoring keys illustrates, there is great uncertainty regarding the true effects of speed-accuracy manipulations. Importantly, there were major differences in both the performance and the rank order of the different estimation approaches, depending on the scoring key that was used. For example, EZ2 (Grasman, Wagenmakers, & van der Maas, 2009)² performed best of all methods, with 84% correct classifications if the standard scoring key was applied. If, however, the scoring key with the assumption of an effect on nondecision time was used, EZ2 had only 68% correct classifications, which was worse than most other diffusion model methods. Accordingly, for the correct interpretation of the results obtained in this and other experimental studies based on speed-accuracy manipulations, it is of utmost importance to analyze the validity of the speed-accuracy manipulations to find out which scoring key is appropriate.

By means of analyses of model parameters gained in experimental validation studies alone, it is not possible to untangle the two possible explanations. One interesting approach to disentangling the processes involved in speed–accuracy tradeoffs has been taken by Rinkenauer et al. (2004; see also Osman et al., 2000; van der Lubbe, Jaśkowski, Wauschkuhn, & Verleger, 2001). Rinkenauer et al. analyzed an event-related potential component, the *lateralized readiness potential*, to separate motor processes from premotor

 $^{^{\}frac{1}{2}}$ EZ2 is an extended version of EZ. The probably most important extension is the incorporation of an estimate of the starting point, which is missing in EZ.



¹ Dutilh et al. (2018) use the term *pseudo-experiment* because the data sets were created on the basis of the data from one large empirical experiment. In this article, for better readability, we will use only the term *experiment*.

processes in tasks with different levels of speed stress. They found that speed–accuracy trade-offs are captured by both premotor and motor processes. Another promising approach that can be used to disentangle the different components involved in speed–accuracy trade-offs is simulation studies. This approach was taken in the present studies.

The present studies

We conducted two simulation studies based on typical parameter constellations observed in empirical, experimental validation studies that have used speed–accuracy manipulations. In the data generation process, we varied either the threshold separation or the nondecision time between two conditions. Basing the simulations on the values observed in specific experimental validation studies allows a comparison of the simulation results with the results of the empirical studies.

After the generation of data sets, we reestimated the parameters using the Kolmogorov-Smirnov optimization criterion implemented in fast-dm-30 (Voss, Voss, & Lerche, 2015).³ Next, we examined which parameters varied between the two conditions. Thus, in the simulation studies we simulated an experimental manipulation with perfect validity. Accordingly, any effect on nondecision time (or any other diffusion model parameter) observed in the condition with manipulation of the threshold separation must necessarily result from problems in the parameter estimation process. If there is no effect on nondecision time, or an effect that is clearly smaller than in the experimental validation study, it is likely that the effects reported in empirical studies are based (mainly) on a valid assessment of psychological processes (i.e., that speed-accuracy manipulations truly do influence nondecisional processes). If, on the other hand, a substantial effect on nondecision time emerged in the simulation study, we would suppose that the lack of validity was traceable to the parameter estimation procedure.

In contrast to the approach by Rinkenauer et al. (2004), this procedure allowed us to assess the size of estimation errors and made it possible to uncover whether the effects on non-decision time typically observed in empirical studies are at least partly caused by estimation problems. Furthermore, in contrast to Rinkenauer et al., we also examined varying numbers of trials, to specify how many trials are required for a clear separation of effects on threshold separation and nondecision time.

Certainly, biases in parameter estimation have been examined in several previous simulation studies (e.g., Lerche, Voss, & Nagler, 2017; Ratcliff & Tuerlinckx, 2002; van

³ Fast-dm-30 is an extended version of fast-dm, with the most important extension being that two further optimization criteria have been included: maximum likelihood and chi-square.



Ravenzwaaii & Oberauer, 2009: Vandekerckhove & Tuerlinckx, 2007). However, few studies have generated differences in threshold separation between conditions, and none, as far as we know, have generated a difference in nondecision time. Moreover, these previous studies were usually not based on the specific parameter sets observed in empirical studies with speed-accuracy manipulations, or they relied on parameter estimates obtained in a fitting process with only one single nondecision time. Furthermore, importantly, these previous studies were usually based on very high trial numbers. For example, Ratcliff and Tuerlinckx (2002) used a minimum of 1,000 trials per data set. In the experimental validation studies, however, trial numbers were much smaller. For example, Voss et al. (2004) used only 40 trials, which made difficulties with disentangling different parameters more likely. We were interested in whether, even for such small trial numbers, effects on threshold separation can be separated clearly from effects on nondecision time. We therefore conducted simulation studies that were explicitly based on the parameter sets observed in experimental validation studies in order to compare the effects observed in the simulation studies with the effects in empirical studies.

In the following sections, we first present a simulation study based on parameter estimates of the experimental validation study by Voss et al. (2004, Study 1). In this study, we used several different parameter sets and generated a large number of data sets for each parameter set. In Study 2, the data were generated on the basis of parameter estimates from Experiment 3 by Dutilh et al. (2018). In contrast to Study 1, we created parameter sets based on a multivariate normal distribution. Thus, in the two studies we employed different strategies of data simulation. The main aim was the same: an examination of the causes underlying the effects of speed–accuracy manipulations on nondecision time.

Study 1: Simulation study based on Voss et al. (2004)

Simulation Study 1 was based on parameter estimates reported by Voss et al. (2004, Study 1). If the diffusion model is capable of disentangling threshold parameter settings from nondecision time effects, an effect should appear only for the specific parameter that was varied in the generation of the data sets. Thus, if two different threshold parameters were used for the data generation, differences in the reestimated parameters should be observed only in this parameter. Likewise, if the two conditions differed only in nondecision time, the effect should solely manifest in the nondecision time estimate and in no other parameter.

Method

Data generation The data generation was based on the parameter values reported in the study by Voss et al. (2004, Study 1). The authors estimated parameters separately for the "standard" (in the following termed "speed") and "accuracy" conditions. For our simulation, we assumed a within-subjects design with two conditions. Either the threshold separation or the nondecision time was varied between the two conditions. For all parameters that were not varied, we used the mean over the two conditions.

For the manipulation of the threshold separation, one parameter set included the threshold separation estimates by Voss et al. (2004) for the two conditions (i.e., $a_{spd} = 1.18$, a_{acc} = 1.37). We further generated two more parameter sets: one parameter set with a smaller difference in the threshold separation between the two conditions, namely half of the original difference; and one parameter set with a larger difference, namely a doubled difference4 (see Table 1 for the exact parameter values). For the manipulation of nondecision time, one parameter set included the nondecision time estimates by Voss et al. (2004) for the two conditions (i.e., $t_{0.\text{spd}} = 0.47$, $t_{0,\rm acc} = 0.52$). In addition—as for the manipulation of the threshold separation—we used one parameter set with a smaller and one with a larger difference (see note 4). In sum, Study 1 was based on a 2 (manipulated parameter: a or t_0) × 3 (effect size: small, medium, or large) design.

The thresholds were associated with the two response alternatives from the color discrimination task. Specifically, the drift rate ν_0 refers to the blue stimulus, associated with the lower threshold, and the drift rate ν_1 to the orange stimulus, associated with the upper threshold. We set the intertrial variability of nondecision time (not estimated in the study by Voss et al., 2004) to 0. One thousand data sets were generated for each parameter set and each of six different numbers of trials (40, 100, 200, 400, 1,000, and 5,000). In half of the simulated trials, the stimulus associated with the upper versus the lower threshold was the correct response.

Parameter estimation For the parameter estimation, we used the same procedure as in the empirical study by Voss et al. (2004): We applied the Kolmogorov–Smirnov optimization criterion implemented in fast-dm-30 (Voss et al., 2015) and reestimated the parameters separately for each data set and each condition. More specifically, we estimated the threshold

separation, starting point, two drift rates (one for each response alternative), nondecision time, and the intertrial variabilities of drift rate and starting point. The intertrial variability of nondecision time was set to 0, as in the generation of the data sets.

Results

For each condition and each of the four main diffusion model parameters⁵, we computed the difference between the parameter estimates of the accuracy and speed conditions. Boxplots of these differences are depicted in Fig. 2 for the manipulation of the threshold separation, and in Fig. 3 for the manipulation of the nondecision time. Positive values indicate that the parameter estimates of the accuracy condition are larger than the estimates obtained for the speed condition. The boxplots show the first quartile, the median, and the third quartile. Additionally, the gray lines display the means of the different conditions. The black horizontal lines indicate the true differences between the two conditions (e.g., 0.19 for a medium-sized difference in threshold separation). Finally, the red lines show the differences reported by Voss et al. (2004) for the parameters that have not been (intentionally) manipulated. The red lines are plotted at the number of trials (n = 40) that was also employed by Voss et al. (2004), to enhance comparability with their study.

Figure 2 shows that the differences in threshold separation between the two conditions were recovered very well. Unsurprisingly, with an increasing number of trials, the estimation grows more precise. However, even for the lower trial numbers the estimations are unbiased, with both the medians and means of the differences very close to the true differences. Importantly, the nondecision times were also estimated very well. Only in the condition with the largest effect size and the smallest trial number (i.e., 40 trials) does a noticeable positive difference between the two conditions show up: In the accuracy condition, the mean nondecision times were longer than in the speed condition, even though the data were generated on the assumption of no difference in this parameter.

Notably, the mean difference found in the simulation study for the medium effect size is smaller than the mean difference observed by Voss et al. (2004)—indicated by the red line—as a one-sample t test revealed (p < .001, $d_z = 0.48$). Moreover, the mean difference in the simulation study deviated from 0 to only a small degree (p < .001, $d_z = 0.12$). Accordingly, the effect on nondecision time reported by Voss et al. (2004) is probably mainly attributable to the manipulation and not to the estimation procedure.

In terms of drift rates, there were no systematic differences between the two conditions. Note that also in the study by Voss et al. (2004), no significant differences in drift rates

⁵ The intertrial variabilities are usually not the focus of empirical studies. However, we also analyzed the results for these parameters and found no systematic differences in s_{zr} or s_v between the two conditions.



 $[\]overline{^4}$ In more detail, first the difference between the larger (para_{lg,orig}) and the smaller (para_{sm,orig}) value of the original parameter was computed (*diff*). Then, the new parameter values were computed according to the following formula:

Condition with the smaller difference: para_{sm,new} = para_{sm,orig} + $\frac{diff}{4}$; para_{lg,new} = para_{lg,orig} - $\frac{diff}{4}$;

Condition with the larger difference: $para_{sm,new} = para_{sm,orig} - \frac{diff}{2}$; $para_{lg,new} = para_{lg,orig} + \frac{diff}{2}$).

 Table 1
 Study 1: Parameter sets for the data generation

	а		t_0		ν_0	ν_1	z_r	S_{t0}	s_{γ}	S_{ZJ}
Manipulated Parameter	Small	Large	Small	Large						
Threshold separation	1.23 1.18	1.32 1.37	0.50		- 1.11	0.90	0.51	0.00	0.83	0.07
Nondecision time	1.09 1.47		0.48 0.47 0.45	0.51 0.52 0.55						

The diffusion coefficient is set to 1 in fast-dm. For comparability with studies that use a coefficient of .1, the parameters a, ν , z_r , s_v , and s_{zr} have to be multiplied by .1

between the two conditions appeared. Finally, regarding the starting point, no differences between the accuracy and speed

conditions were present, in the data from either our simulation study or the study by Voss et al. (2004).

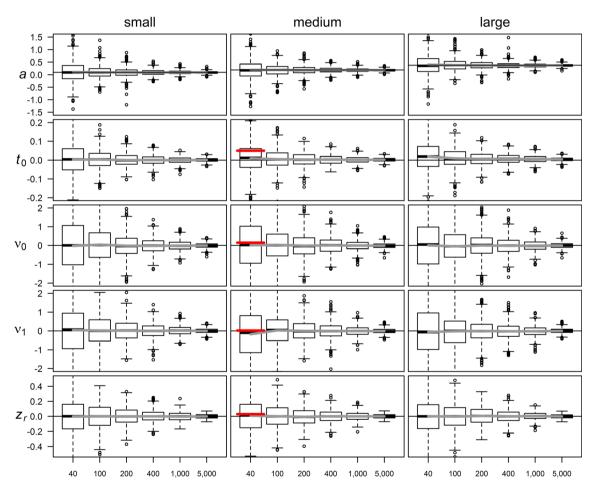


Fig. 2 Results of the manipulation of threshold separation in Study 1: Boxplots of the differences between parameter estimates from the condition with large versus small threshold separation, as a function of the size of the true effect (small–medium–large) and the number of trials. Positive values indicate larger values in the high than in the low threshold

separation condition. The gray lines display the means, the black horizontal lines the true differences between the two conditions, and the red lines the differences reported by Voss et al. (2004). For better comparability with ν_1 , ν_0 was multiplied by -1



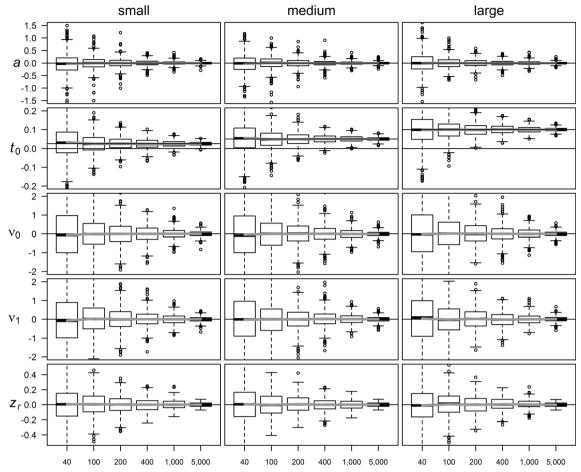


Fig. 3 Results of the manipulation of nondecision time in Study 1: Boxplots of differences between the parameter estimates from the conditions with slow versus fast nondecision time, as a function of the size of the true effect (small-medium-large) and the number of trials.

Positive values indicate larger values in the slow than in the fast nondecision time condition. The gray lines display the means, and the black horizontal lines, the true differences between the two conditions. For better comparability with ν_1 , ν_0 was multiplied by -1

The parameter estimation was also very effective if nondecision time was varied between conditions. Figure 3 reveals that the effect on nondecision time was captured exclusively by the nondecision time parameter. Even for small trial numbers and for small to large effects, the effect was attributed to the correct parameter.

Discussion

In the experimental validation study by Voss et al. (2004), the speed–accuracy manipulation influenced both threshold separation and nondecision time. As expected, the threshold separation was higher in the accuracy than in the speed condition. Additionally, however, the nondecision time was higher in the accuracy than in the speed condition. We wanted to investigate whether this effect on nondecision time was mainly attributable to trade-offs in parameter

estimation. If there are no severe trade-offs, the effect is likely to be attributable to a lack of discriminant validity of the experimental manipulation. The speed-accuracy manipulation might influence not only the accumulation of information, but also nondecisional components such as motoric or encoding processes. Thus, in our simulation study we aimed to disentangle these two possible accounts.

We generated data based on the assumption that differences between the two conditions were based exclusively on either threshold separation or nondecision time. The parameter sets for the data simulation were based on the estimates reported by Voss et al. (2004). In addition to the trial number and effect size from the empirical study, we further examined larger trial numbers and smaller and larger effect sizes. Notably, the parameter estimation worked well: In almost all conditions, the true differences showed up only in the parameter actually manipulated, and in none of the other parameters.



A comparison of the mean difference reported by Voss et al. (2004) with the one in our study further revealed that it is very unlikely that the difference in nondecision time reported by Voss et al. (2004) is attributable to trade-offs in the parameter estimation procedure. Only in the condition with the largest effect and the smallest trial number did the nondecision time estimate increase notably in the accuracy condition. Conversely, we observed no difference in the condition with a medium-sized effect that was based on the effect size observed in the empirical data. Accordingly, we assume that the effect in the empirical study was due to true differences in nondecision time between the speed and accuracy conditions.

Note that in the interpretation of our results, we make the assumption that the empirical data on which we based our simulation study were generated by diffusion processes. Theoretically, differences in nondecision time between the speed and accuracy conditions might have resulted from contaminated data—that is, from trials in which decisions were not (purely) based on a diffusion process but, for example, on fast guesses or attention lapses. If there were more fast contaminants in the speed condition and more slow contaminants in the accuracy condition, this could explain the differences in nondecision time. Thus, effects on nondecision time could emerge, even though there were no real differences in either motoric response execution or the encoding of information, but rather in the contamination between the two conditions.

To examine whether contaminants were responsible for the effect on nondecision time, we "contaminated" 4% of trials (a typical value that has also been used, e.g., by Ratcliff & Childers, 2015; Ratcliff & Tuerlinckx, 2002) by changing the simulated data. We assumed an extreme case in which, in the low-threshold condition 100% of the contaminants were fast, and in the high-threshold condition 100% of the contaminants were slow. Given this extreme case, the contaminants should be most likely to impact nondecision time. For the generation of contaminants, we used the same strategy as in Lerche et al. (2017). Fast contaminants simulated fast guesses with a random response (either upper or lower threshold) and the RTs randomly drawn from a uniform distribution ranging from $t_0 - 100$ ms to $t_0 + 100$ ms. Thus, the responses in these fast-guess trials were located at the lower edge of the RT distribution. For the slow contaminant trials, the simulated response latencies were replaced by higher values without changing the actual response. The slow RTs were drawn from a uniform distribution ranging from 1.5 to 5 interquartile ranges above the third quartile of the RT distribution. The results of the analyses with contaminated data are presented in Fig. 4. Importantly, despite the presence of contaminants, condition still had no effect on nondecision time. Rather, the presence of contaminants further increased the effect on threshold separation. Thus, the presence of contaminants can result in an overestimation of the effect on threshold separation. In a further set of analyses, we added fast contaminants to the condition with low t_0 and slow contaminants to the condition with high t_0 . Again, the contaminants affected threshold separation rather than nondecision time (Fig. 5). In sum, our additional analyses support the view that the effect on nondecision time typically observed in empirical validation studies cannot be explained by estimation problems, but rather is attributable to true effects of the speed–accuracy manipulation on nondecisional components.

In the following section, we present a further simulation study that was conducted to test the generalizability of our results to a different method of data generation (using a multivariate normal distribution) and different underlying parameter values.

Study 2: Simulation study based on Dutilh et al. (2018)

For Study 2, we relied on data from Experiment 3 of Dutilh et al. (2018).⁶ Whereas some experiments from this large-scale validation project contained several manipulations, the data of Experiment 3 were based solely on a speed–accuracy manipulation. In terms of convergent validity, the results from the different research teams were promising: 82% (i.e., 14 out of 17 teams) detected the expected effect in the threshold separation parameter. In addition, all teams that employed a diffusion model (whether the simple or the full diffusion model) found the effect. Additionally, 41% (seven out of 17 teams) detected an effect on nondecision time, with a higher estimate in the accuracy than in the speed condition. More specifically, 60% (six out of ten groups) of the diffusion model teams observed this effect, in contrast to only one out of five LBA teams.

Again, essentially two main explanations for the unexpected effect on nondecision time can be put forward: The effect might be due to (1) trade-offs in the parameter estimation or (2) the insufficient validity of the experimental manipulation. It is possible that the participants in the speed condition not only reduced the time spent on accumulating information, but also hastened their motoric responses and information encoding. In our simulation study, we aimed to disentangle the two possible explanations.

Method

Data generation Each experiment by Dutilh et al. (2018) consisted of 20 participants with 400 trials each (200 trials per condition). We used the parameter estimates obtained by the Voss and Lerche (VL) modeling team for Experiment 3. VL estimated the parameters separately for the two conditions



⁶ The data can be retrieved from the Open Science Framework project page: https://osf.io/jy7pi/.

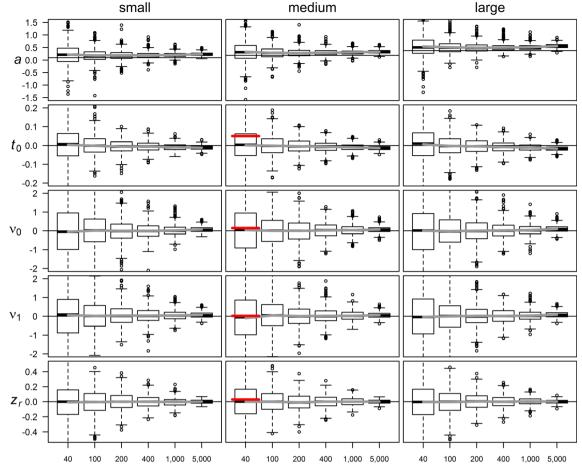


Fig. 4 Results of the manipulation of threshold separation with the addition of contaminated trials in Study 1: Boxplots of differences between the parameter estimates from the conditions with large versus small threshold separation, as a function of the size of the true effect (small-medium-large) and the number of trials. Positive values indicate

larger values in the high than in the low threshold separation condition. The gray lines display the means, the black horizontal lines the true differences between the two conditions, and the red lines the differences reported by Voss et al. (2004). For better comparability with ν_1 , ν_0 was multiplied by -1

of the experiment, using the Kolmogorov–Smirnov optimization criterion and a full diffusion model (i.e., including intertrial variability parameters). More specifically, their model included the following parameters: threshold separation, starting point, two drift rates (one for each response alternative), nondecision time, and the intertrial variabilities of drift rate, starting point, and nondecision time. For the data from this experiment, VL obtained significant effects on threshold separation (p < .001, $d_z = 1.76$), and nondecision time (p < .001, $d_z = 1.27$).

Akin to Study 1, we simulated data sets that differed between the two conditions in either threshold separation or nondecision time. For all parameters that were not manipulated, we used the mean parameter values across the two conditions. All means, standard deviations, and intercorrelations between the parameters can be found in Table 2. On the basis of these values, we generated parameter sets assuming a

multivariate normal distribution. More specifically, for each type of manipulation (threshold separation or nondecision time), we simulated 100 experiments that each included 20 parameter sets (i.e., 20 participants). For each parameter set, we then created data sets with varying numbers of trials. The trial numbers were identical to those in Study 1 (i.e., 40, 100, 200, 400, 1,000, and 5,000). When threshold separation and

⁷ Studies 1 and 2 were based on different simulation approaches. Whereas in Study 1 we simulated 1,000 identical "participants" on the basis of a certain parameter set, in Study 2, we assumed different parameter sets for the different "participants" (based on a multivariate normal distribution). In both studies we used all information that was available about the original data (in Study 1, only the means of the parameters were available). Both types of simulation approaches have been used in the diffusion model literature (identical data sets: e.g., Ratcliff & Tuerlinckx, 2002; distinct data sets based on a multivariate normal distribution: e.g., van Ravenzwaaij & Oberauer, 2009). Because the findings of our two studies did not differ, we assume that the specific simulation approach does not make a relevant difference.



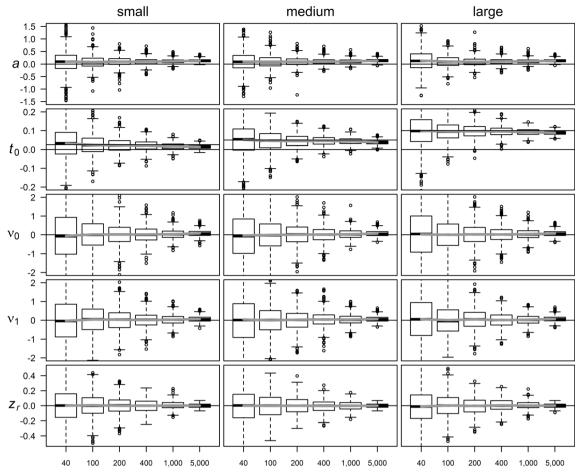


Fig. 5 Results of the manipulation of nondecision time with the addition of contaminated trials in Study 1: Boxplots of differences between the parameter estimates from the conditions with slow versus fast nondecision time, as a function of the size of the true effect (small—medium—large) and the number of trials. Positive values indicate larger

values in the slow than in the fast nondecision time condition. The gray lines display the means, and the black horizontal lines, the true differences between the two conditions. For better comparability with ν_1 , ν_0 was multiplied by -1

Table 2 Study 2: Means, standard deviations, and intercorrelations of the parameters on the basis of which the parameter sets were created

		a	$ u_0$	ν_1	t_0	z_r	S_{t0}	$S_{\mathbf{v}}$	$S_{\mathcal{I}'}$
	Small	0.90 (0.12)			0.36 (0.06)				
M(SD)	Mean	1.10 (0.13)	- 1.70 (0.74)	1.70 (0.55)	0.40 (0.05)	0.50 (0.06)	0.20 (0.05)	0.74 (0.22)	0.25 (0.11)
	Large	1.29 (0.21)			0.44 (0.07)				
a		_	05	.32	.13	22	.22	.21	.09
ν_0			_	50	.22	55	.20	54	.12
ν_1				_	16	31	28	.25	06
t_0					_	.03	.36	32	04
z_r						_	02	.23	16
S_{t0}							_	49	05
s_{ν}								_	32
S_{ZF}									_

Either parameter a or parameter t_0 was varied (small vs. large value). When varying a, the mean value of t_0 was used and when varying t_0 , the mean value of a was used. The diffusion coefficient is set to 1 in fast-dm. For comparability with studies that have used a coefficient of .1, the parameters a, ν , z_r , s_v , and s_v have to be multiplied by .1



nondecision time were manipulated, the mean effect sizes (i.e., the effect sizes averaged across the true parameter values of the 100 experiments) were 1.67 and 0.98, respectively.

Parameter estimation We used the KS optimization criterion implemented in fast-dm-30 and estimated a full diffusion model. This strategy was used for best comparability with the results obtained by the VL team, who used the same estimation procedure. Furthermore, in an additional set of analyses we employed the maximum likelihood (ML) criterion and estimated a five-parameter model (with the intertrial variabilities of drift rate and starting point fixed at zero). This estimation procedure produces reliable results and has been found to be superior to other estimation procedures in both empirical test-retest and simulation studies (Lerche & Voss, 2016). The patterns of results from the two estimation approaches (KS with seven-parameter model vs. ML with five-parameter model) were basically identical. Like VL, we estimated the parameters separately for each data set and each of the two conditions.

Results

For each type of manipulation (threshold separation vs. nondecision time), each trial number, and each diffusion model parameter, we computed the proportions of significant results (paired t tests, two-sided, p < .05) across the 100 experiments. The findings for the four main diffusion model parameters are presented in Fig. 6. Positive (vs. negative) proportions indicate cases in which a parameter was significantly larger (vs. smaller) in the accuracy than in the speed condition. If a parameter was at times significantly smaller and at other times significantly larger in the accuracy than in the speed condition, two symbols are displayed (positive proportions in black and negative proportions in gray).

First, it is notable that from 100 trials onward, the proportions of significant results for the respective manipulated parameters are very close or equal to 1. Thus, the true effects on either threshold separation or nondecision time were detected very well. Only for the lowest number of trials (i.e., 40 trials per condition) was this proportion of "hits" slightly lower.

Concerning discriminant validity, the manipulation of threshold separation resulted in only small percentages of effects on other parameters. Yet the highest numbers of significant effects were observed for nondecision time (12%) and for the intertrial variability of nondecision time (11%) in the condition with 40 trials. For higher trial numbers and for the other diffusion model parameters, the proportions were even smaller.

We also examined effect sizes (see Fig. 7). In the condition with the manipulation of threshold separation, the mean effect sizes (Cohen's d_z) for threshold separation, averaged across the 100 experiments, were 0.81 in the condition with 40 trials and 1.66 in the condition with 5,000 trials. As we mentioned earlier, the effect size underlying the true parameter sets was 1.67. From around 200 trials on, the effect size of the estimated values ($d_z = 1.33$) was close to the effect size of the true parameter values. For parameter t_0 , the effect sizes range from 0.18 (n = 40) to -0.01 (n = 5,000). Notably, from 100 trials onward the effect size is already smaller than 0.10. Thus, only for very small numbers of trials are the true effect sizes overestimated. The effect sizes of the remaining parameters amount to a maximum absolute value of 0.19 (s_{t0} , n = 40).

The manipulation of nondecision time, akin to Study 1, did not lead to an effect in any of the other parameters. In the condition with 40 trials, the effect size based on the t_0

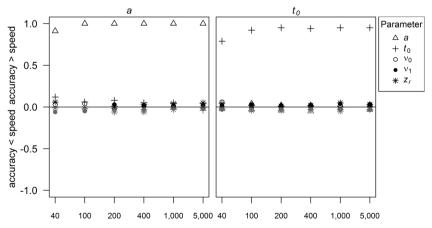


Fig. 6 Relative proportions of significant effects (p < .05) in Study 2, depending on the manipulated parameter (left plot, a; right plot, t_0), the estimated parameter, and the number of trials. Positive (vs. negative) proportions indicate cases in which the parameter was significantly

larger (vs. smaller) in the accuracy than in the speed condition. Positive proportions are depicted in black, and negative proportions in gray. For better comparability with ν_1 , ν_0 was multiplied by -1



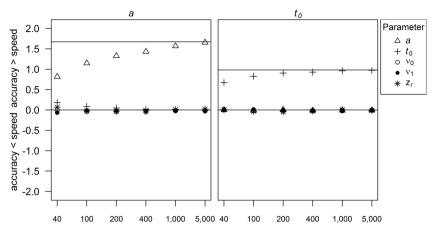


Fig. 7 Effect sizes (Cohen's d_z) in Study 2, depending on the manipulated parameter (left plot, a; right plot, t_0), the estimated parameter, and the number of trials. Positive (vs. negative) values indicate cases in which the

parameter was larger (vs. smaller) in the accuracy than in the speed condition. The black lines depict the true effect sizes. For better comparability with ν_1 , ν_0 was multiplied by -1

estimates was 0.67, and in the condition with 5,000 trials it reached 0.97. For comparison, the effect size based on the true parameter values was 0.98. From around 200 trials on, the effect size ($d_z = 0.90$) was similar to the effect size of the true parameter values. The effect sizes of the other parameters had a maximum absolute value of 0.04.

Finally, we also analyzed the data of Study 2 with the same approach as in Study 1, examining biases in the parameter estimations. More specifically, we computed difference scores between the two conditions across all 2,000 participants (100 experiments \times 20 participants). Analogous to Study 1, when threshold separation was manipulated, the boxplots for non-decision time were centered around zero. Only in the condition with 40 trials was t_0 slightly higher in the accuracy than in the speed condition. Importantly, for the condition with 400 trials (i.e., the trial number used by Dutilh et al., 2018), there was no bias. If nondecision time was manipulated, we only found an effect on this parameter.

Discussion

In Study 2, using a different simulation strategy based on a multivariate normal distribution and on different underlying parameter values, we generated data sets with a difference between two conditions in either threshold separation or nondecision time. More specifically, the data were generated on the basis of the parameters estimated by one of the modeling teams (VL) for the data from Experiment 3 of Dutilh et al. (2018).

Importantly, we found that the generated effects were correctly captured by the parameters that were manipulated for the simulation. Moreover, since we based our data generation and parameter estimation on the findings and procedure used by VL in the Dutilh et al. (2018) study, we were able to

compare the results from the simulation with the results from the real data. Thus, we could test whether the unexpected effect on nondecision time is attributable to the parameter estimation procedure or rather to a lack of discriminant validity of the experimental manipulation. Toward this aim, let us consider the effect sizes in the original study and those in the simulation study. In the analyses of the empirical data, for threshold separation an effect size of 1.76 emerged, as compared to an effect size of 1.27 for nondecision time. Thus, the effect size of parameter a was 1.4 times as large as the one observed for t_0 . In our simulation study, on the other hand, for 400 trials (as employed by Dutilh et al., 2018), the effect size for parameter a was 1.43, and the one for t_0 was 0.02. Thus, the effect size of the threshold separation was about 70 times as large as the effect size of parameter t_0 . This indicates that the number of trials in the study by Dutilh et al. was sufficiently large to avoid false alarms in nondecision time as a consequence of trade-offs in the parameter estimation. In turn, this implies that the nondecision time effect observed in the real data was caused by true differences in motoric or encoding processes between the two conditions.

In addition to the unexpected effect on nondecision time in Experiment 3 of Dutilh et al. (2018), 65% of the teams (ten out of 17) also found an effect on drift rate, with a higher drift rate in the accuracy than in the speed condition. This was the case for all analyses based on the linear ballistic accumulator (LBA) model, but was only present in 40% of the diffusion model analyses. VL were among the teams who did not find a significant effect. In our simulation study, we also had a proportion of only 5% of significant effects (for n = 400) on the drift rate, and the effect sizes had a maximum absolute value of $d_z = 0.05$.

As for Study 1, we also examined the influence of contaminants. We used the same procedure as in Study 1 for the



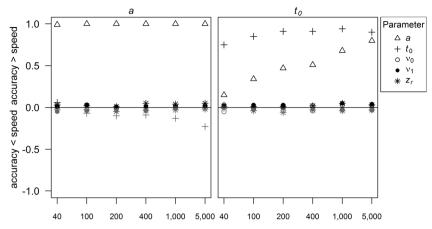


Fig. 8 Relative proportions of significant effects (p < .05) in Study 2 with the addition of contaminated trials, depending on the manipulated parameter (left plot, a; right plot, t_0), the estimated parameter, and the number of trials. Positive (vs. negative) proportions indicate cases in

which the parameter was significantly larger (vs. smaller) in the accuracy than in the speed condition. Positive proportions are depicted in black, and negative proportions in gray. For better comparability with ν_1 , ν_0 was multiplied by -1

generation of contaminated trials. 8 The results are presented in Fig. 8 (as proportions of significant effects) and Fig. 9 (as effect sizes). The results are compatible with those obtained for Study 1. Again, the effects on threshold separation were larger when contaminants were introduced. At 200 trials, the estimated effect size was comparable to the true effect size. For even higher trial numbers, contaminants caused an overestimation of the true effect size. At the same time, higher nondecision times were found in the speed than in the accuracy condition. If nondecision time was manipulated and contaminants were added, again threshold separation was affected by the contaminants, with a higher value in the condition with slow contaminants than in the condition with fast contaminants. The effect on nondecision time was slightly smaller than without contaminants. Thus, as these additional analyses show, as in Study 1, contaminants were not responsible for the higher nondecision times in the accuracy condition typically observed in parameter validation studies.

General discussion

To test the validity of the diffusion model parameter measuring decisional caution (i.e., the threshold separation parameter *a*), speed–accuracy manipulations have frequently been used. However, to provide a convincing test of validity, the manipulation must address the cognitive process in question *effectively* and *exclusively*. A typical finding reported in experimental studies that utilize the diffusion model (e.g., Arnold et al., 2015; Voss et al., 2004) is that speed–accuracy manipulations not only result in the expected effect on threshold

separation (i.e., higher threshold separation in the accuracy than in the speed condition) but also influence nondecision time, with nondecision times being found to be higher in the accuracy than in the speed condition.

Effect of speed–accuracy manipulations on nondecision time

Theoretically, this effect might be attributable to trade-offs in the parameter estimation procedure. In this case, the manipulation would—as intended—exclusively impact decision style, but the effect would spread to other parameters (i.e., nondecision time) in the parameter estimation procedure. This would indicate a general problem with the validity of diffusion model parameters, and thus make it problematic to interpret diffusion model findings. Alternatively, the effect of speed-accuracy manipulations might be due to a lack of discriminant validity of the experimental manipulation. Following this idea, the manipulation might influence not only decision style, but also the speed of nondecisional processes such as motoric response execution (Rinkenauer et al., 2004). It is plausible that participants who are instructed to speed up their decision will also try to press the respective key as fast as possible, whereas they might take more time for the key press if the accuracy of the response is emphasized. In addition, the instructions might also impact the encoding of information: Participants might more efficiently focus their encoding on relevant information if they are told to be as fast as possible. This second explanation thus challenges the validity of the manipulation and assumes that the estimated parameters are valid measures for the actual cognitive processes.

In several previous studies in which speed-accuracy manipulations have been employed, only an effect on threshold



The only difference was that the interval for the fast contaminants reached from $(t_0 - s_{t0}/2) - 100$ ms to $(t_0 - s_{t0}/2) + 100$ ms (in Study 1 s_{t0} was 0).

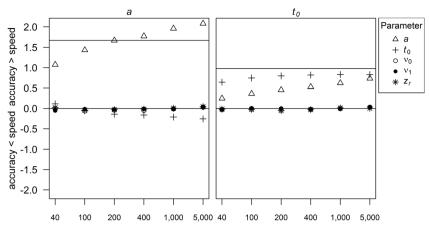


Fig. 9 Effect sizes (Cohen's d_z) in Study 2 with the addition of contaminated trials, depending on the manipulated parameter (left plot, a; right plot, t_0), the estimated parameter, and the number of trials. Positive (vs. negative) values indicate cases in which the parameter was

larger (vs. smaller) in the accuracy than in the speed condition. The black lines depict the true effect sizes. For better comparability with $\nu_{\rm l},\,\nu_{\rm 0}$ was multiplied by - 1

separation was reported (e.g., Ratcliff et al., 2003; Thapar et al., 2003). However, in these studies the parameter estimation was based on strong restrictions: Only the threshold separation parameter, and occasionally the starting point, were allowed to vary between conditions, with all other parameters being forced to be equal across conditions. This restriction may have been based on incorrect assumptions, because studies that have not constrained nondecision time have consistently observed an effect of this parameter (Arnold et al., 2015; Rae et al., 2014; Rinkenauer et al., 2004; Voss et al., 2004).

Summary of the method and main results

In two simulation studies, we examined whether the effect on nondecision time is traceable to problems in parameter estimation. If trade-offs in parameter estimation are not the underlying problem, the effect on nondecision time is likely to be attributable to a lack of discriminant validity of the experimental manipulation. Data sets were generated on the basis of parameter values estimated from previous experimental validation studies. More specifically, in Study 1 the data generation was based on the parameter estimates reported by Voss et al. (2004), and in Study 2 on the estimates found by VL for Experiment 3 of the large collaborative validation project by Gilles Dutilh and Chris Donkin (Dutilh et al., 2018). Drawing on the reported parameter values, we simulated data for two conditions that differed either exclusively in threshold separation or exclusively in nondecision time, and fitted the complete diffusion model to the simulated data. Thus, we could analyze in which of the diffusion model parameters the simulated effect was captured. If the parameter estimation worked perfectly, the effect should be reflected solely in the manipulated parameter and in none of the other parameters.

We examined results from simulations using different numbers of trials (ranging from 40 to 5,000) so that we could also test how many trials are required to clearly separate the effects of the manipulations.

We found that a true effect on nondecision time was exclusively picked up by the nondecision time parameter, even for small numbers of trials and for small to large effects. Thus, if in real data a difference in nondecision time is present, it is likely to be detected. For the manipulation of threshold separation, the findings were similar: Again, the separation of the true effect from effects on other parameters was successful. As Study 1 showed, only for a condition with a very large difference in threshold separation and a very small number of trials (n = 40) was the nondecision time estimate slightly increased in the accuracy condition. Thus, in experiments in which speed-accuracy manipulations are extreme, the experiments are more likely to falsely detect an effect on nondecision time. Prominently, by far the largest effect was always present in the threshold separation parameter. If an experimental study produces very high differences in threshold separation and also small effects on nondecision time, we advise that the nondecision time effects be interpreted cautiously. These effects might arise from trade-offs in the parameter estimation procedure and might not reveal effects that truly exist.

In additional analyses, we examined whether the effect on nondecision time is attributable to differences in the percentages of contaminants in the speed and accuracy conditions. We added fast contaminants to the speed and slow contaminants to the accuracy condition. Importantly, this extreme distribution of contaminants did not result in an effect on nondecision time. Therefore, we are confident that the effects on nondecision time found in previous studies are attributable to true effects on this parameter.



Influence of the number of trials

A further noteworthy finding regards the number of trials. Not surprisingly, with an increase in the number of trials, the estimation became more precise. However, the trial number had a negligible effect on the detection of differences in the parameters between conditions. Only for the smallest trial number of n = 40 was the detection rate slightly reduced. From 100 or 200 trials onward, the performance of the estimation procedures was impressively stable. Thus, whether, for example, 200 or 1,000 trials are used will not likely affect the pattern of results. For very high numbers of trials, however, participants may lose motivation or concentration, causing contamination from nondiffusion processes. Accordingly, we consider it inadvisable to use more than several hundred trials. These findings are in line with other studies that have shown further increase in the number of trials to be of limited utility (Lerche & Voss, 2017b; Lerche et al., 2017).

Since the data of our simulation studies were generated on the basis of empirical studies (Dutilh et al., 2018; Voss et al., 2004), a comparison of the findings in the empirical and simulation studies was possible. This comparison showed that the effects on nondecision time in the simulation studies were clearly smaller than the effects observed in the experimental studies. This suggests that the effects on nondecision time in the empirical data are not (or, if at all, are only to a small part) attributable to problems with the parameter estimation.

Interpretation of the results by Dutilh et al. (2018)

Importantly, the results from the present studies can help us interpret the findings reported by Dutilh et al. (2018). The authors proposed different "scoring keys" to award the best parameter estimation performance. According to their standard scoring key, speed—accuracy manipulations should selectively affect the threshold separation. Relying on this key, EZ2 (Grasman et al., 2009) performs best of all methods. Dutilh et al. further evaluated the performance of estimation methods based on two alternative scoring keys. In one of these keys, they assumed an additional effect on nondecision time, and in the other key an effect on drift rate.

Interestingly, the EZ2 algorithm was evaluated worse if an effect of the speed–accuracy manipulations on nondecision time was considered correct (as our present results suggest). The superiority of the EZ2 method is mostly a result of the rejection of effects on nondecision time. Specifically, EZ2 did not detect an effect of nondecision time in any experiment that manipulated speed–accuracy settings. The other diffusion model accounts, on the other hand, detected such an effect, on average, 72.2% of the time. Perhaps unsurprisingly, the earlier EZ method (Wagenmakers et al., 2007) seems to behave in a manner similar to the more advanced EZ2 method. In a study by Arnold et al. (2015), in contrast to fast-dm (KS)

and DMAT, the EZ method did not find a difference in nondecision time between the speed and the accuracy conditions. We believe that the study by Dutilh et al. (2018) will be seen as an important guideline by many readers concerning which approach to use to model RT data. Thus, it is imperative that the correct scoring key be applied. On the basis of the findings of this article, a scoring key that assumes at least an additional effect in nondecision time seems more appropriate than the standard scoring key. In addition to the change in the performance of EZ2, the performance of the other methods also depends on the scoring key. Interestingly, whereas the diffusion model approaches perform better with a scoring key that assumes an influence on nondecision time (with a mean increase in percent correct classifications of 7.6%), the linear ballistic accumulator model approaches perform worse with this key (-9.6%; for further details, see Table 4 in Dutilh et al., 2018).

Directions for future research

One key avenue for future validation studies will be to analyze more complex experimental designs. In our simulation studies, we always varied only one parameter. Following the approach of Dutilh et al. (2018), it would be interesting to test whether the validity of diffusion model analyses is still high when several parameters are varied simultaneously. The study by Dutilh et al. was based on a total of 14 experiments. In some of them, two or even three of the main diffusion model parameters were manipulated experimentally. Thus, the data provided by Dutilh et al. might constitute a basis for further simulation studies. If effects on more parameters are present, more difficulties in separating effects might be encountered. In support of this hypothesis, Dutilh et al. found that a higher rate of false alarms in nondecision time was observed if two or three parameters (threshold separation, drift rate, and starting point) were tackled by manipulations simultaneously (54%) than when only the threshold separation was manipulated via a speed–accuracy manipulation (41%).

It would also be interesting to compare different estimation procedures. In this article we relied on the Kolmogorov–Smirnov optimization criterion, as this criterion was also used in the experimental manipulation studies on which our data generation was based. One might consider testing whether the findings supplied by KS are comparable with those obtained with other commonly used optimization criteria. For the data of Study 2, we tested one different estimation strategy: We used maximum likelihood and a parameter model with the intertrial variabilities of starting point and drift rate fixed. The results were comparable to those obtained by KS.

One model that was not a focus of this article but that could be examined in future studies is the LBA (Brown & Heathcote, 2008). As the project by Dutilh et al. (2018) demonstrated, that model is less likely to reveal differences



in nondecision time. However, speed–accuracy manipulations often resulted in an effect on the drift rate according to the LBA model (see also Ho et al., 2012). For example, in Experiment 3 of Dutilh et al., all five teams using LBA observed a difference in drift rate between the two conditions, with higher drift rates in the accuracy than in the speed condition. Considerations similar to those from the studies in this article could be made: Are these differences attributable to (1) trade-offs in parameter estimation or (2) real differences in the speed of information accumulation between the two conditions? Participants might put more effort into the task in the accuracy than in the speed condition. Simulation studies that follow a procedure similar to that of the present Study 2 could help disentangle the two possible accounts.

Finally, it is important to note that speed–accuracy instructions are not the only experimental manipulations that have been used with the assumption of influencing threshold settings. For example, Naefgen, Dambacher, and Janczyk (2017, Exp. 1), varied the frequency of catch trials (see also Näätänen, 1972; Seibold, Bausenhart, Rolke, & Ulrich, 2011). The diffusion model analyses of Naefgen et al. revealed that this manipulation, however, significantly influenced both nondecision time (larger with a higher frequency of catch trials) and drift rate (lower with a higher frequency of catch trials) rather than threshold separation (which had been theorized to be higher if more catch trials were present). As the authors stated, the unexpected effect on drift rate could be a result of trade-offs in the parameter estimation. It is, however, also possible that the manipulation truly influenced the speed of information accumulation. It has been shown that practice can increase drift rates (e.g., Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Lerche & Voss, 2017b), and participants are likely provided with less opportunity for practice if more catch trials are present. Simulation studies—similar to the ones in this article—could help disentangle the effects of the manipulation from trade-offs in parameter estimation.

Conclusions

Two simulation studies revealed that effects on threshold separation and nondecision time can be effectively disentangled. We concluded that the effects of speed–accuracy manipulations on each of these parameters are, therefore, likely attributable to differences in the relevant cognitive processes, rather than to trade-offs in parameter estimation. Accordingly, we question the discriminant validity of the speed–accuracy manipulations employed in a large number of studies. Our findings are relevant for the correct interpretation of experimental validation studies and for assessing the validity of published empirical findings, such as the effects on age differences.

Author note This research was supported by a grant from the German Research Foundation to A.V. (Grant No. VO1288/2-2).



- Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research*, 79, 882– 898. https://doi.org/10.1007/s00426-014-0608-y
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178. https://doi.org/10.1016/j.cogpsych.2007.12. 002
- Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*, 5–13. https://doi.org/10.5334/pb-51-1-5
- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, 18, 61–69. https://doi.org/10. 3758/s13423-010-0022-4
- Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., ... Donkin, C. (2018). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1417-2
- Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. Psychonomic Bulletin & Review, 16, 1026–1036. https://doi.org/10.3758/16.6.1026
- Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed–accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience*, 31, 17242–17249. https://doi.org/10.1523/ JNEUROSCI.0309-11.2011
- Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014). Social influence and perceptual decision making: A diffusion model analysis. *Personality and Social Psychology Bulletin*, 40, 217–231. https://doi.org/10.1177/0146167213508985
- Grasman, R. P. 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. https://doi.org/10.1016/j.jmp. 2009.01.006
- Heitz, R. P. (2014). The speed–accuracy tradeoff: History, physiology, methodology, and behavior. Frontiers in Neuroscience, 8, 150. https://doi.org/10.3389/fnins.2014.00150
- Ho, T., Brown, S., van Maanen, L., Forstmann, B. U., Wagenmakers, E.-J., & Serences, J. T. (2012). The optimality of sensory processing during the speed–accuracy tradeoff. *Journal of Neuroscience*, 32, 7992–8003. https://doi.org/10.1523/JNEUROSCI.0340-12.2012
- Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods*, 46, 702–721.
- Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology*, 7. https://doi.org/10.3389/fpsyg.2016.01324
- Lerche, V., & Voss, A. (2017a). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research*. https://doi.org/10.1007/s00426-017-0945-8
- Lerche, V., & Voss, A. (2017b). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research*, *81*, 629–652. https://doi.org/10.1007/s00426-016-0770-5
- Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of



different optimization criteria. Behavior Research Methods, 49, 513–537. https://doi.org/10.3758/s13428-016-0740-2

- Mulder, M. J., Bos, D., Weusten, J. M. H., van Belle, J., van Dijk, S. C., Simen, P., ... Durston, S. (2010). Basic impairments in regulating the speed–accuracy tradeoff predict symptoms of attention-deficit/ hyperactivity disorder. *Biological Psychiatry*, 68, 1114–1119. https://doi.org/10.1016/j.biopsych.2010.07.031
- Näätänen, R. (1972). Time uncertainty and occurrence uncertainty of the stimulus in a simple reaction time task. *Acta Psychologica*, *36*, 492–503. https://doi.org/10.1016/0001-6918(72)90029-7
- Naefgen, C., Dambacher, M., & Janczyk, M. (2017). Why free choices take longer than forced choices: Evidence from response threshold manipulations. *Psychological Research*. https://doi.org/10.1007/ s00426-017-0887-1
- Osman, A., Lou, L., Muller-Gethmann, H., Rinkenauer, G., Mattes, S., & Ulrich, R. (2000). Mechanisms of speed–accuracy tradeoff: Evidence from covert motor processes. *Biological Psychology*, *51*, 173–199. https://doi.org/10.1016/S0301-0511(99)00045-9
- Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1226–1243. https://doi.org/ 10.1037/a0036801
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108. https://doi.org/10.1037/0033-295x.85.2.59
- Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception* and Performance, 40, 870–888. https://doi.org/10.1037/a0034954
- Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, 2, 237–279. https://doi.org/10.1037/dec0000030
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for twochoice decisions. *Psychological Science*, 9, 347–356. https://doi. org/10.1111/1467-9280.00067
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, 16, 323. https://doi.org/10.1037/0882-7974.16.2.323
- Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, 65, 523–535. https://doi.org/10.3758/BF03194580
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, 50, 408–424. https://doi.org/10.1016/j.jml.2003.11.002
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, 60, 127– 157. https://doi.org/10.1016/j.cogpsych.2009.09.001
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9, 438–481. https://doi.org/10.3758/bf03196302
- Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the Locus of Speed–accuracy Trade-Off in Reaction Time: Inferences From the Lateralized Readiness Potential. *Journal* of Experimental Psychology. General, 133, 261–282. https://doi. org/10.1037/0096-3445.133.2.261
- 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. https://doi.org/10.1037/0096-3445.136.3.414
- Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., & Bergmann, K. (2015). Decomposing the relationship between mental speed and

- mental abilities. *Intelligence*, 51, 28–46. https://doi.org/10.1016/j.intell.2015.05.002
- Schulz-Zhecheva, Y., Voelkle, M., Beauducel, A., Biscaldi, M., & Klein, C. (2016). Predicting fluid intelligence by components of reaction time distributions from simple choice reaction time tasks. *Journal of Intelligence*, 4, 8. https://doi.org/10.3390/jintelligence4030008
- Seibold, V. C., Bausenhart, K. M., Rolke, B., & Ulrich, R. (2011). Does temporal preparation increase the rate of sensory information accumulation? *Acta Psychologica*, 137, 56–64. https://doi.org/10.1016/j. actpsy.2011.02.006
- Spaniol, J., Voss, A., & Grady, C. L. (2008). Aging and emotional memory: Cognitive mechanisms underlying the positivity effect. *Psychology and Aging*, 23, 859–872. https://doi.org/10.1037/a0014218
- Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, 64, 1–34. https://doi.org/10.1016/j.cogpsych.2011.10.002
- Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, 18, 415–429. https://doi.org/10.1037/0882-7974.18.3.415
- van der Lubbe, R. H. J., Jaśkowski, P., Wauschkuhn, B., & Verleger, R. (2001). Influence of time pressure in a simple response task, a choice-by-location task, and the Simon task. *Journal of Psychophysiology*, 15, 241–255. https://doi.org/10.1027/0269-8803.15.4.241
- van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: Ez, fast-dm, and DMAT. *Journal of Mathematical Psychology*, 53, 463–473. https:// doi.org/10.1016/j.jmp.2009.09.004
- Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, 14, 1011–1026. https://doi.org/10.3758/bf03193087
- Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. Behavior Research Methods, 40, 61–72. https://doi.org/10.3758/brm.40.1.61
- Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, 49, 653–673. https://doi. org/10.3758/s13428-016-0721-5
- Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology*, 44, 1048–1056. https://doi.org/10.1016/j.jesp.2007.10.009
- Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32, 1206–1220. https://doi.org/10.3758/BF03196893
- Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, 39, 767–775. https://doi.org/10.3758/bf03192967
- Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, 52, 1–9. https://doi.org/10.1016/j.jmp.2007.09.005
- Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.00336
- 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. https://doi. org/10.1016/j.jml.2007.04.006
- Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. Psychonomic Bulletin & Review, 14, 3–22. https://doi.org/10.3758/bf03194023

