Factoring out non-decision time in choice RT data: Theory and implications

Stijn Verdonck and Francis Tuerlinckx
Faculty of Psychology and Educational Sciences
KU Leuven, University of Leuven

Choice RT experiments are an invaluable tool in psychology and neuroscience. A common assumption is that the total choice response time is the sum of a decision and a non-decision part (time spent on perceptual and motor processes). While the decision part is typically modeled very carefully (commonly with diffusion models), a simple and ad hoc distribution (mostly uniform) is assumed for the non-decision component. Nevertheless, it has been shown that the misspecification of the non-decision time can severely distort the decision model parameter estimates. In this paper, we propose an alternative approach to the estimation of choice RT models that elegantly bypasses the specification of the non-decision time distribution by means of an unconventional convolution of data and decision model distributions (hence called the D*M approach). Once the decision model parameters have been estimated, it is possible to compute a non-parametric estimate of the non-decision time distribution. The technique is tested on simulated data, and is shown to systematically remove traditional estimation bias related to misspecified non-decision time, even for a relatively small number of observations. The shape of the actual underlying non-decision time distribution can also be recovered. Next, the D*M approach is applied to a selection of existing diffusion model application papers. For all of these studies, substantial quantitative differences with the original analyses are found. For one study, these differences radically alter its final conclusions, underlining the importance of our approach. Additionally, we find that strongly right skewed non-decision time distributions are not at all uncommon.

Keywords: choice RT, diffusion model, bias, non-decision time, misspecification, D*M

Introduction

In the last decade, the use of evidence accumulation models for choice RT experiments has revealed fundamental insights into the process of elementary decision making (Basten, Biele, Heekeren, & Fiebach, 2010; DasGupta, Ferreira, & Miesenböck, 2014; Forstmann et al., 2008; Krajbich & Rangel, 2011; Polanía, Krajbich, Grueschow, & Ruff, 2014; Ratcliff & Dongen, 2011; Resulaj, Kiani, Wolpert, & Shadlen, 2009). A common assumption in all these studies, is that choice RT can be additively decomposed into a decision time (taken up by the process responsible for the actual choice) and a residual non-decision time. The latter is considered to be the contribution of encoding and motor response execution (Luce, 1986). In contrast to the intricate evidence accumulation or diffusion models for the decision part of choice RT (Brown & Heathcote, 2008; Ratcliff & Rouder, 1998; Usher & McClelland, 2001; Verdonck & Tuerlinckx, 2014; Wong & Wang, 2006), the models traditionally used for the non-decision time are surprisingly basic: adding a simple time constant, or, at best, adding a simple parametric distribution (mostly uniform, see Ratcliff & Tuerlinckx, 2002). However, choosing the wrong model for the non-decision time is known to be a cause for bias for the decision model's parameter estimates, especially when data originating from a skewed non-decision time distribution are fitted under the assumption of a non-skewed non-decision time distribution (Ratcliff, 2013). Seeing a (right) skewed non-decision time distribution is at least as plausible as the current non-skewed default. Therefore, today's golden standard for the non-decision time may be systematically delivering biased results for the decision process parameters.

In this paper, we develop a novel estimation approach that

The research leading to the results reported in this paper was sponsored in part by Belgian Federal Science Policy within the framework of the Interuniversity Attraction Poles program (IAP/P7/06), by the grant GOA/15/003 from the KU Leuven and by the grant G.0806.13 from the Fund of Scientific Research Flanders. For the computations, we used the infrastructure of the VSC–Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government– Department of Economy, Science, and Innovation (EWI). We would also like to thank Martijn Mulder, Gilles Dutilh and Florian Schmitz for kindly sharing their data and Irene Gijbels for advice.

bypasses the specification of the non-decision time distribution and allows us to estimate the decision model's parameters without any non-decision time misspecification bias. Moreover, if desired, a non-parametric estimate of the residual non-decision time can be estimated.

Attempts to disentangle non-decision from decision time have been proposed before (Smith, 1990), but from a different perspective: the non-decision time distribution was assumed to coincide with the distribution of simple (i.e., one-choice) response times, observed in a comparable experimental setup. This distribution was then deconvolved from the distribution of total choice response times (from the original two-choice version of the experiment) in order to isolate the decision part. The resulting distributions could then be modeled with a decision model. The approach we propose in this paper, does not use information of secondary experiments nor does it presume any particular parametric distribution (shape) for the non-decision time distribution.

In what follows we will first explain the method, then illustrate its performance based on simulated data, and finally use it to re-analyze data from three different diffusion model application papers.

The D*M method

Assume a two-choice RT experiment with I different conditions ($i = 1, ..., I$). These conditions can be stimuli of different difficulty (e.g., percentages coherently moving dots in a random dots motion task) but also other manipulations (e.g., prior expectations). During the course of the experiment, there are several trials within a condition, and each trial can results in either a correct ($c = 1$) or an error choice response ($c = 0$). For the sake of simplicity, we assume an equal number of N trials. Let us start with focusing on the theoretical probability density function (pdf) of the total response time for condition i and choice response c. For convenience, we introduce an index $p = 2i - c$, running over all $2I$ condition-response pairs ic: correct choice responses have odd p, incorrect choice responses have even p. This total response time pdf can be seen as a convolution of a decision pdf (denoted as $m_p(t; \boldsymbol{\theta}_0)$) and a non-decision pdf (denoted as $r(t)$):

$$f_p(t) = (m_p * r)(t) = \int_0^\infty m_p(t - x; \boldsymbol{\theta}_0) r(x) dx,$$

where $*$ stands for the convolution operator and $\boldsymbol{\theta}_0$ is the true parameter vector of the decision pdf. Obviously, also $f_p(t)$ depends on $\boldsymbol{\theta}_0$, but we suppress this dependence for simplicity. In what follows, we will denote the convolution product between densities $m(t; \boldsymbol{\theta}_0)$ and $r(t)$ as $m(\boldsymbol{\theta}_0) * r$. Note also that the integral over t for $f_p(t)$ is equal to the probability of condition-response pair p, thus $f_p(t)$ is a defective pdf. However, because a choice response for condition i is either correct or incorrect, $\int_0^\infty f_{2i-1}(t)dt + \int_0^\infty f_{2i}(t)dt = 1$. For convenience, an overview of the notation used in this paper is provided in Table 1.

Table 1
List of symbols used.

Variable	Description
I	number of conditions (e.g., stimuli)
p	index used to indicate condition-response pairs ($0 < p \leq 2I$)
m_p	(parametric) decision model distribution
r	non-decision time distribution
f_p	total choice response time distribution
$\hat{g}_p$	kernel density estimate of f_p

Because the convolution operator is commutative, for any two non-identical condition-response pairs p and p', it holds exactly that

$$\begin{aligned} r * m_p(\boldsymbol{\theta}_0) * m_{p'}(\boldsymbol{\theta}_0) &= r * m_{p'}(\boldsymbol{\theta}_0) * m_p(\boldsymbol{\theta}_0) \\ f_p * m_{p'}(\boldsymbol{\theta}_0) &= f_{p'} * m_p(\boldsymbol{\theta}_0). \end{aligned} \quad (1)$$

This convolution identity is the fundament of our method: The non-decision time distribution is factored out.

The idea is to use Equation 1 to find an estimate of $\boldsymbol{\theta}_0$, given observed choice RT samples from f_p and $f_{p'}$ on the one hand, and well-specified decision model pdfs m_p and $m_{p'}$ on the other. The data enter the equality by approximating f_p and $f_{p'}$ through nonparametric estimates based on the observations, denoted as $\hat{g}_p$ and $\hat{g}_{p'}$. The estimated densities $\hat{g}_p$ and $\hat{g}_{p'}$ are kernel density estimators, with $\hat{g}_p$ based on the N_p observed RTs t_p^j ($j = 1 \ldots N_p$) for condition-response pair p:

$$\hat{g}_p(t) = \frac{1}{N_p} \sum_{j=1}^{N_p} K\left(\frac{t - t_p^j}{h}\right).$$

As a smoothing kernel function $K(t)$, we use a uniform distribution from 0 to 1. The bandwidth is chosen to be $h = 1$.

When replacing f_p and $f_{p'}$ by $\hat{g}_p$ and $\hat{g}_{p'}$, the equality of Equation 1 is no longer exact. However, the discrepancy (or objective function)

$$D_{pp'}(\boldsymbol{\theta}) = d\left(\hat{g}_p * m_{p'}(\boldsymbol{\theta}), \hat{g}_{p'} * m_p(\boldsymbol{\theta})\right), \quad (2)$$

where $d(\cdot, \cdot)$ is a distance defined on the space of defective pdfs, should be small for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Note that in Equation 2, a convolution appears between the data from condition-response pair p', represented by the smoothed estimated density $\hat{g}_{p'}$, and the model's density for condition-response pair p, that is $m_p(\boldsymbol{\theta})$. Therefore we refer to our technique as the D*M method. To evaluate this crucial convolution both model and data densities are discretized using an equally

spaced grid and the convolution integral is then approximated by a finite sum over this grid.

In this paper we opt for a chi-square type of distance between the compared distributions

$$d(a,b) = \int_0^\infty \frac{[a(t) - b(t)]^2}{a(t) + b(t)} dt. \tag{3}$$

For a conventional chi-square distance, where typically an observed distribution is compared to a model pdf, the integrand's denominator is the pdf of the model. Because in Equation 2, both compared distributions are a convolution of a data and a model pdf, we take the sum of the two compared distributions for the integrand's denominator in Equation 3. This distance is called the triangular discrimination (Topsoe, 2000) and can be considered as a symmetrized version of the chi-square distance (a "chi-square like distance"; Le Cam, 1986).

It is obvious that for $\hat{g}_p \to K * f_p$ and $\hat{g}_{p'} \to K * f_{p'}$, $D_{pp'}(\boldsymbol{\theta})$ reaches its minimal value of zero, when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (because of Equation 1). In Appendix A it is shown that if the decision model does not contain a non-decision component of its own and the total response time pdfs f_p underlying the data are not all equal across condition-response pairs, this minimum is unique, at least for the parameters pertaining to m_p or $m_{p'}$. In this sense, $D_{pp'}(\boldsymbol{\theta})$ can be used to estimate the parameters pertaining to m_p and $m_{p'}$ without explicitly solving the underlying inverse convolution problem leading to an estimate of $r(t)$. Because we are interested in the parameters pertaining to all condition-response pairs, we use the total objective function

$$T(\boldsymbol{\theta}) = \sum_{p=2}^{2I} \left(\sum_{p'=1}^{p-1} D_{pp'}(\boldsymbol{\theta}) \right), \tag{4}$$

where the sum runs over all possible unique combinations of two different condition-response pairs. The estimated parameter vector $\hat{\boldsymbol{\theta}}$ is then found as: $\hat{\boldsymbol{\theta}} = \text{argmin} T(\boldsymbol{\theta})$.

If the true parameter vector $\boldsymbol{\theta}_0$ is the absolute and unique minimum of all separate terms $D_{pp'}$ (at least for all parameters pertaining to m_p and $m_{p'}$), the same will hold for the sum $T(\boldsymbol{\theta})$ (for all parameters). For $\boldsymbol{\theta}_0$, the non-decision pdf $r(t)$ clearly exists and can be retrieved as the deconvolution of each condition-response pair p's total response time and decision pdfs:

$$r \quad = \quad f_p * m_p(\boldsymbol{\theta}_0)^{-1}, \tag{5}$$

where -1 in the exponent refers to the deconvolution operation.

However, for an arbitrary parameter vector $\boldsymbol{\theta}$, it is not sure that these deconvolutions exist. Conditions for the existence of a deconvolution are described in Appendix B and it is shown that, up to a practical degree of accuracy when dealing with noisy data, they boil down to the following condition on the variances σ^2:

$$\sigma^2_{f_p} \geq \sigma^2_{m_p(\boldsymbol{\theta})}, \tag{6}$$

basically avoiding a negative variance for $r(t)$. Because we use a smoothing kernel K to estimate f_p, the kernel needs to be taken into account and effectively we will use the equivalent constraint:

$$\sigma^2_{K * f_p} \geq \sigma^2_{K * m_p(\boldsymbol{\theta})} = \sigma^2_{m_p(\boldsymbol{\theta})} + \sigma^2_K,$$

where the last step follows from the additive property of variances of independent random variables. In terms of the data, the constraint is:

$$\sigma^2_{\hat{g}_p} \geq \sigma^2_{K * m_p(\hat{\boldsymbol{\theta}})} = \sigma^2_{m_p(\hat{\boldsymbol{\theta}})} + \sigma^2_K. \tag{7}$$

This condition has to be met for all condition-response pairs p. Although these conditions have to be met for all condition-response pairs p, the validity of each of these constraints is only as good as the estimate $\sigma^2_{\hat{g}_p}$. For this reason, densities $\hat{g}_p(t)$ based on only a few data points should not be included. If these constraints are imposed while minimizing $T(\boldsymbol{\theta})$, the existence of a common $r(t)$ for all pairs p is ensured. For the final estimation of $r(t)$, we use the following expression:

$$\begin{aligned} r \quad &= \quad r * \left(\frac{1}{I} \sum_{p=1}^{2I} m_p(\boldsymbol{\theta}_0) \right) * \left(\frac{1}{I} \sum_{p=1}^{2I} m_p(\boldsymbol{\theta}_0) \right)^{-1} \\ &= \quad \left(\frac{1}{I} \sum_{p=1}^{2I} r * m_p(\boldsymbol{\theta}_0) \right) * \left(\frac{1}{I} \sum_{p=1}^{2I} m_p(\boldsymbol{\theta}_0) \right)^{-1} \\ &= \quad \left(\frac{1}{I} \sum_{p=1}^{2I} f_p \right) * \left(\frac{1}{I} \sum_{p=1}^{2I} m_p(\boldsymbol{\theta}_0) \right)^{-1}, \end{aligned} \tag{8}$$

where both sums runs over all $2I$ pairs p. Having obtained an estimate $\hat{\boldsymbol{\theta}}$ of the true parameter values $\boldsymbol{\theta}_0$, this results in an estimator

$$\hat{r} = \left(\frac{1}{I} \sum_{p=1}^{2I} \hat{g}_p \right) * \left(\frac{1}{I} \sum_{p=1}^{2I} K * m_p(\hat{\boldsymbol{\theta}}) \right)^{-1}, \tag{9}$$

where the same smoothing kernel K is used for both factors of the deconvolution. The constraints in Equation 7 ensure the existence of this deconvolution. If the deconvolution exists, then clearly for $\hat{g}_p \to K * f_p$, and therefore $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$, $\hat{r}(t) \to r(t)$. Note that the smoothing kernel K used in this final step can differ from the kernel used in the minimization procedure, but has to be applied in a recalculation of $\hat{g}_p(t)$ as well (for the deconvolution, we will use a uniform distribution from 0 to 0.01 for K). In practice, we solve the deconvolution problem in Equation 9 by defining a grid (we

take a grid spacing of 0.01s) and minimizing the triangular discrimination distance (see Equation 3) between the data-based distribution $\left(\frac{1}{I}\sum_{p=1}^{2I}\hat{g}_p\right)$ and the model-based distribution $r * \left(\frac{1}{I}\sum_{p=1}^{2I} K * m_p(\hat{\boldsymbol{\theta}})\right)$ with respect to the probability weights assigned to the grid points. We use the same global optimizer as before to tackle this high dimensional minimization problem: for a grid spacing of 0.01s and $\hat{r}$ clipped at 1.5s, $\hat{r}$ consists of 150 grid points whose weights have to be estimated. We again use the triangular discrimination distance for consistency, but for this estimation other distances or could be used as well.

In the theory developed above, we have assumed that all $2I$ condition-response pairs share the same non-decision time distribution r. Note however that the total objective function $T(\boldsymbol{\theta})$ in Equation 4 can be easily changed to allow for multiple non-decision time distributions, each shared within a subset (with 2 elements or more) of condition-response pairs. This setup is the analogue of a traditional diffusion model analysis in which the non-decision time parameter T_{er} is allowed to vary across (some of the) conditions, while the other parameters are constrained to be equal. In order to implement such a situation, it suffices to limit the double sum in Equation 4 to terms produced within subsets of condition-response pairs with the same non-decision time distribution. Each non-decision time distribution can then be estimated separately by limiting the sums in Equation 9 to the subset of condition-response pairs corresponding to that non-decision time distribution.

Performance on simulated data

In this section we show how well the D*M technique recovers a standard parameter set of the Ratcliff diffusion model (Ratcliff & Tuerlinckx, 2002), in conjunction with different non-decision time distributions. The quality of the estimates is systematically compared to the results obtained with a standard approach in which it is assumed that the non-decision time is uniformly distributed.

Data simulation

We simulated data from a typical Ratcliff diffusion model parameter set (Table 2) in conjunction with three different non-decision pdfs (see the righthandside figures of the three panels in Figure 1): one right skewed, one uniform and one bimodal distribution. For all three parameter configurations, we look at data sets of 300, 1,000 and 1,000,000 RTs per stimulus. For each combination of sample size and non-decision time distribution, we simulated 100 data sets.

Standard and D*M estimation procedures

For the standard estimation procedure we used both the standard maximum likelihood method and the quantile likelihood method (Heathcote & Brown, 2004; Heathcote, Brown, & Mewhort, 2002). As the results are very comparable, we only show results of the latter. As is customary, a uniform distribution for the non-decision time distribution was used and the location and spread of this uniform distribution was estimated alongside the pure decision model parameters.

Table 2
Ratcliff diffusion model parameters used for the decision part of the simulated data.

Parameter	Description	Value
a	boundary separation	0.08
η	inter-trial variability of drift rate	0.08
zr	relative bias	0.5
sz	uniform inter-trial variability of bias $z = a \cdot zr$	0.02
v_i $(i = 1, \ldots, 4)$	drift rates	$[0.4, 0.25, 0.1, 0]$

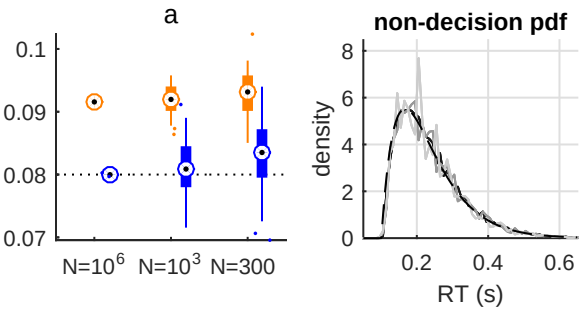
In the D*M procedure no non-decision pdf has to (nor should) be included. To evaluate the convolution integral, we choose a grid ranging from 0 to 5s (well beyond any observed RT) with equally spaced nodes at every 0.01s.

For both methods (standard and D*M), the model density $m(\boldsymbol{\theta})$ needs to be calculated. To do this, we use code from the fast-dm project (Voss & Voss, 2007). Also in both methods, an objective function has to be minimized (the negative quantile likelihood for the standard method and Equation 4 for D*M). To find this minimum, we use a global optimizer (i.e., differential evolution; Storn & Price, 1997) .

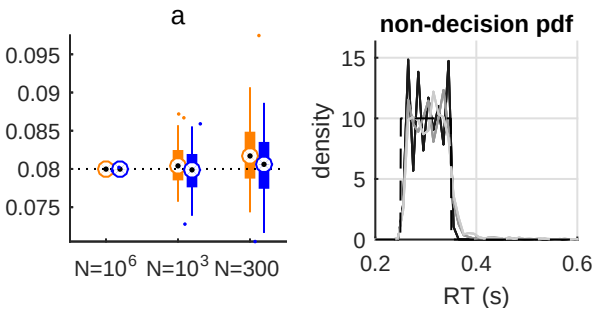
Results

Figure 1 shows the estimates for the boundary separation a (results for the other parameters are offered as supplemental material), in conjunction with three different shapes of non-decision time densities. The traditional approach results in a systematic estimation bias, except for the uniform non-decision pdf, in which case the non-decision model is perfectly specified for the data. This systematic bias is resolved with the D*M method, which gives better estimates (or comparable in the unlikely case that the actual non-decision time density is a uniform distribution), even for a number of data points as low as 300 per condition. In addition, the average recovered non-decision time densities match remarkably well with the true non-decision time densities.

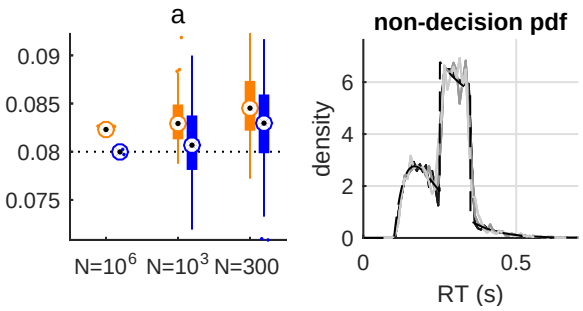
As an additional test, we have repeated this analysis for different values of boundary separation a. Increasing values of a correspond to a larger proportion of the total response time (and its variance) being accounted for by the decision model (keeping everything else constant). The results are presented in Figure 2. For the right skewed non-decision time distribution, the D*M method always outperforms (or matches) the traditional method, regardless of the value of a or the number of observations. For the uniform and bi-



(a) right skewed non-decision pdf

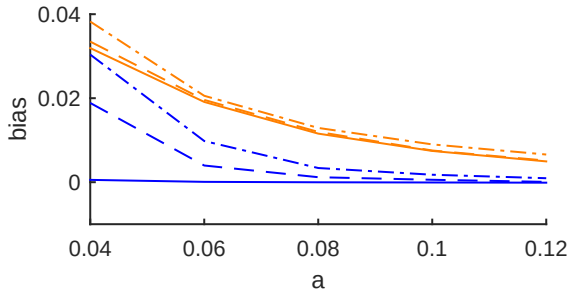


(b) uniform non-decision pdf

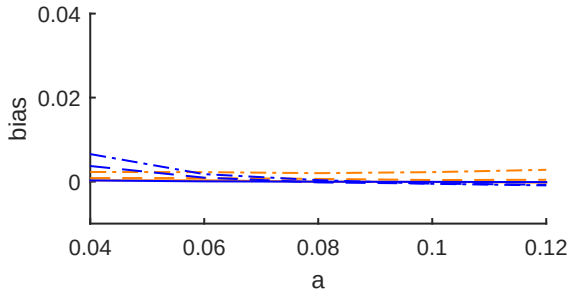


(c) bimodal non-decision pdf

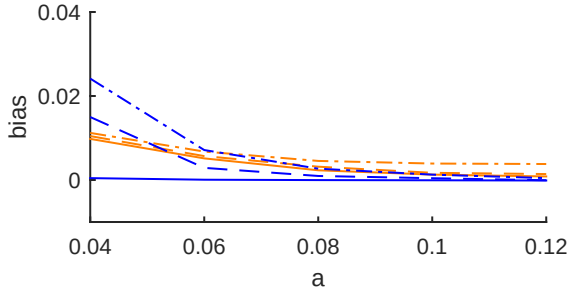
Figure 1. In the left panels, the recovery of the Ratcliff diffusion model's boundary separation parameter a is shown, for data simulated with either a right skewed non-decision pdf (a), a uniform non-decision pdf (b) or a bimodal combination of both (c). Estimates are shown for data sets with subsequently $N = 10^6$, $N = 10^3$ and $N = 300$ observations per condition. The orange box plots show the distributions of the estimates obtained with the traditional method, the blue box plots those obtained with the D*M method. (The recovery results for the other diffusion model's parameters yield very similar conclusions and are offered as supplemental material.) In the right panels, we show for each non-decision time density, the average of the non-parametric estimates based on the D*M decision model parameter estimates as given by Equation 9. The average is represented by solid lines varying from light to dark with the increasing number of observations and compared to the original shape (dashed black line).



(a) right skewed non-decision pdf



(b) uniform non-decision pdf

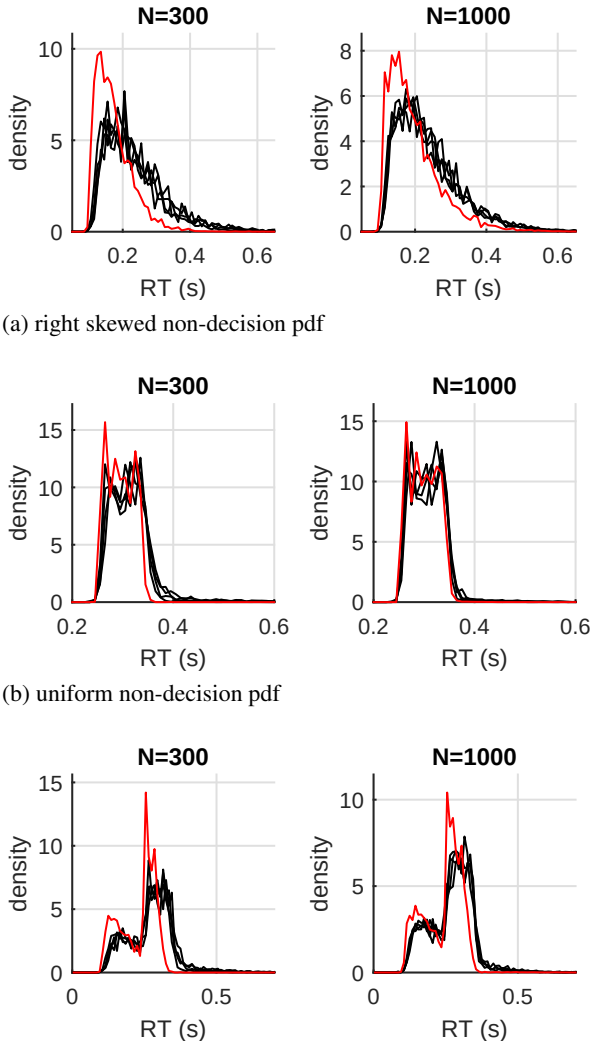


(c) bimodal non-decision pdf

Figure 2. Estimation biases of boundary separation a, for varying a. The same non-decision time distributions were used as in Figure 1: a right skewed non-decision pdf (a), a uniform non-decision pdf (b) or a bimodal combination of both (c). Orange lines correspond to traditional estimates, blue lines to D*M estimates. The estimates are based on data sets with subsequently $N = 10^6$ (solid lines), $N = 10^3$ (dashed lines) and $N = 300$ (dotted-dashed lines) observations per condition. If the number of trials is high (solid lines), the D*M method always outperforms (or matches) the traditional method, no matter what the value is of a. For less data (dashed lines, dotted-dashed lines), D*M still performs better for normal and high values of a. For a small amount of data and small values of a, however, and if the non-decision time distribution is more in line with the traditional uniform assumption like it is in (b) and (c), the D*M method shows a higher bias than the traditional method. Similar patterns are found for the other parameters.

modal non-decision time distributions, D*M still performs better or equal for normal and high values of a, but can have higher biases for small values of a, if there is a limited number of observations. This is what one would expect: As the proportion of total response time variance generated by the non-decision time process gets larger (the non-decision time distribution proportions of total variance at $a = 0.04$ are 0.91,0.47 and 0.87, for right-skewed, uniform and bimodal, respectively), the decision part becomes increasingly obfuscated by the non-decision part. As the D*M method only uses a parametric model for the decision part, more observations are needed to correctly disentangle the two components. Thus, if (1) the uniform distribution is a good enough proxy for the actual non-decision time distribution, (2) the actual non-decision time distribution contributes enough to the total response time and (3) there is not a lot of data, the traditional method can outperform the D*M method. It has to be noted however that, unless we have some prior information about the real shape of the non-decision time distribution, the first condition can never be checked in a real world problem. If one wants to avoid results that depend on any particular presupposed shape of the non-decision time distribution, the D*M approach is by definition the better choice. Figure 3 shows the non-parametric estimates of the non-decision time distributions following the D*M estimates shown in Figure 2. Only for the smallest boundary separation $a = 0.04$ and a limited number of observations, the non-parametric estimates of the non-decision time distribution is somewhat biased, but even then the main features of the distributions are recovered.

For the estimation of the decision model parameters, the repeated calculation of the complete decision model pdfs on a reasonably detailed grid is by far the most computationally expensive part for both the traditional and the D*M method. In our implementation, a single evaluation of the D*M or traditional objective function takes up a comparable amount of time. We opted for a global minimization routine to reduce potential problems concerning local minima, but this resulted in optimization times that are probably a lot slower than strictly necessary. Running single-threaded on an i7 core clocked at 3.60GHz, both a single D*M estimate and traditional estimate take around 5 minutes to complete. The estimation of the non-decision time (once D*M parameter estimates have been obtained) takes about 30 minutes for a grid with 0.01s spacing but only about 15 seconds for a grid with 0.05s spacing. To be confident we reached convergence, we repeated every estimation multiple times (with a different population of starting values). The D*M estimates seemed somewhat more robust than the traditional estimates: for the D*M objective function almost every repetition resulted in the same global minimum; for the traditional objective function, ending up in a local minimum was a bit more frequent. By repeating the minimization procedure multiple times (5 times for D*M and 10 times for the traditional method), we repeatedly found a lowest minimum and were convinced of convergence. An alternative for the global optimizer may be a local optimizer in combination with a rational starting point (e.g., using EZ diffusion, Wagenmakers, Maas, & Grasman, 2007). Such a routine may yield equally good results and will be much faster. However, because of the novelty of the D*M method, we were more concerned with accuracy (i.e., avoiding local optima) than with speed.



(a) right skewed non-decision pdf

(b) uniform non-decision pdf

(c) bimodal non-decision pdf

Figure 3. Recovery of the respective non-decision time distributions based on data sets generated with varying boundary separation a, with $N = 300$ (left) and $N = 1000$ (right) observations per condition. Red lines indicate the non-decision time distributions obtained with the lowest value of boundary separation $a = 0.04$; the rest of the non-decision time distributions (with a somewhere between 0.06 and 0.12) are in black.

Existing diffusion model analyses revisited

To illustrate how the D*M parameter estimation method can lead to fundamentally different findings than the traditional method, we re-analyze the data of three choice RT studies in which a diffusion model is used. These papers can be seen as typical examples of today's common practice of using diffusion models parameters to explain differences between (groups of) participants and/or experimental conditions. We only present the main results; more detailed results are offered as supplemental material.

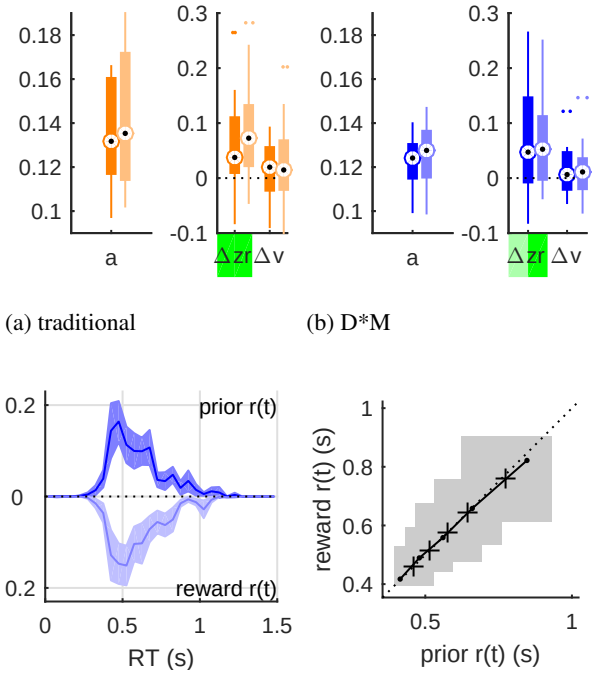
Application 1: A diffusion model analysis of choice bias (Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012)

Mulder et al. (2012) investigate, using a moving-dot perceptual decision making experiment, how different types of pre-trial information biases people towards one of two choice alternatives. One type of pre-trial information concerns the elevated prior likelihood of the occurrence of one of two choices, the other type involves a larger potential pay-off for one of the two choices. The diffusion model is used to study which decision process aspects are affected by each form of pre-trial information (elevated prior likelihood or larger pay-off, each requiring a separate diffusion model analysis): One possibility is that the bias is caused by a shift Δv in the diffusion model's drift rate parameter, the other possibility is that the bias is caused by a shift Δzr in the starting position parameter. The results are shown in Figure 4.

The authors concluded that it is mainly the starting position parameter that is responsible for the bias introduced by both elevated prior likelihood and larger potential pay-off conditions. Upon re-analysis we come to the same qualitative conclusion, for both the traditional and D*M procedure. Quantitatively, however, there are clear differences between the D*M and the traditional estimates. From Figure 4, it can be seen that the boundary separation parameter a is systematically lower for the D*M estimates and the resulting between-person variance (as indicated by the box plots) is considerably smaller than that of the traditional estimates. Clearly, the assumption of a uniform non-decision time density is not valid; instead, the non-decision time density estimates indicate a strong right skew. However, in this case, this misspecification does not yield a different qualitative conclusion.

Application 2: A diffusion model analysis of post-error slowing (Dutilh et al., 2011)

A well documented observation in choice RT experiments is that the response time increases on a trial immediately following an incorrect choice. This phenomenon is called post-error slowing (PES). Dutilh et al. (2011) have attempted to isolate the diffusion model parameters that can account



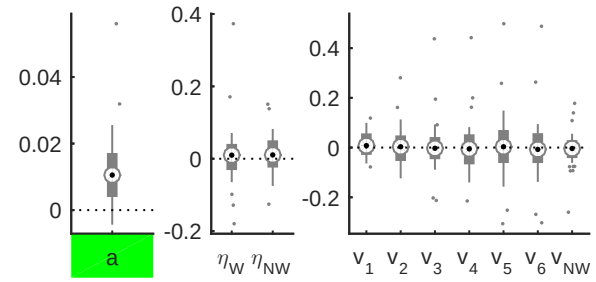
(a) traditional (b) D*M

(c) non-decision pdfs (d) Q-Q plot non-decision pdfs

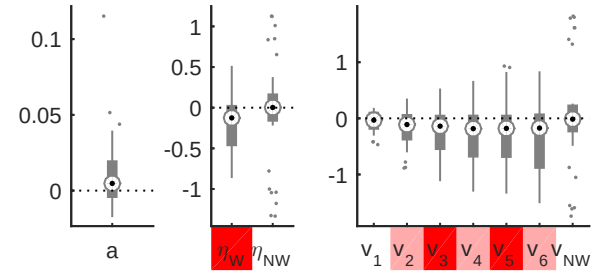
Figure 4. Application 1: A selection of parameter estimates (boundary separation a and effects of bias on starting point Δzr and drift rate Δv) of a diffusion model analysis of choice bias for 20 participants (for a complete parameter overview please see supplemental material). The first two panels show the estimates of a selection of decision model parameters, obtained with either the traditional (a, in orange) or D*M method (b, in blue). For each of the two panels, the darker box plots show the estimates for the elevated prior likelihood condition, the lighter plots for the larger potential pay-off condition. If for the group of participants Δzr or Δv is significantly different from 0 (two-sided sign test), this is indicated with a green ($p < 0.001$) or light green ($p < 0.01$) marking of the label. Panel (c) shows the non-decision time densities inferred from the D*M estimates in panel (b). The densities for the elevated likelihood condition are shown in the upper half of the plot and those for the larger potential pay-off condition are shown, mirrored, in the lower half. The solid lines show the mean non-decision pdfs across participants, the lighter areas display the double standard error interval. Panel (d) is a quantile-quantile plot of the data in panel (c), and is better suited to look at the differences between the non-decision pdfs from the two conditions. The grey area represents a 95% confidence interval of the mean quantile-quantile values (black crosses).

for the differences between post-correct and post-error trials of a lexical decision task. Participants have to classify a string of letters as a word or non-word. Besides the obvious word vs. non-word manipulation, there are six different word frequency types. In their diffusion model analyses, the authors estimate a separate non-decision parameter T_{er} for every stimulus condition. Analogously, the D*M method is set-up to allow for a separate non-decision time distribution for each stimulus condition (grouping choice RT distributions that share a non-decision time distribution) as explained in the last paragraph of the theory section. The results are shown in Figure 5. Based on their analysis, which involves the estimation of two separate diffusion models (one for the post-error and one for the post-correct condition), the authors concluded that post-error slowing is very much associated with an increase in response caution. A similar analysis with the D*M method, however, shows no association between PES and response caution, but reveals other associations, namely with η_W and most word drift rates v_i ($i = 2, \ldots, 6$). More specifically, after an error, the drift rates for words become smaller (i.e., closer to zero) and the trial-to-trial variability of word drift rates also decreases. The reason for the difference between the original and the D*M analysis lies with the specification of the non-decision pdf. The authors allow a different mean non-decision time for all stimuli, but assume an equal, uniform width. As can be seen in Figure 5, panel (d), the D*M method suggests both mean and variance of the non-decision time increase after an error trial when judging a word stimulus (shown in black in panel (d) in Figure 5), but the non-decision time does not change after an error trial when judging a non-word stimulus (shown in red in panel (d) in Figure 5). These effects could not be accommodated by the author's particular specification of the non-decision time distribution, so other parameters had to compensate, with different results and conclusions as a consequence. Based on our analysis, we have to conclude that in the context of this particular diffusion model, there is a fundamental difference in the processing of word and non-word stimuli following an error trial.

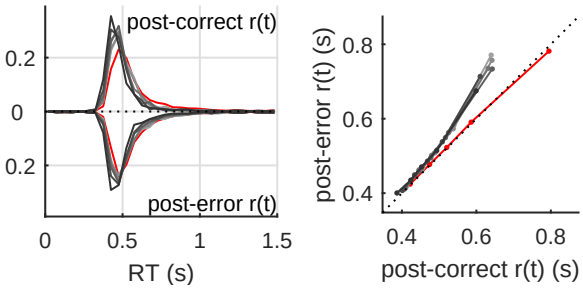
As for all our re-analyses, we rigorously implemented the model assumptions of the original paper. However, it is tempting to wonder if just allowing the width of the uniform non-decision time distribution to vary across stimuli in the traditional approach, would be sufficient to get traditional results comparable to the D*M analysis. To answer this question, we performed the extra analysis and a similar picture as Figure 5, panel (a) was obtained (see supplemental material), showing that the absence of skew in the non-decision time distribution specification is an essential component of the misspecification. As a possibly viable traditional alternative to our method, one could now suggest to have mean, skew and width as parameters of some new non-decision specification, but this would mean estimating 35 parameters, (21



(a) traditional



(b) D*M



(c) non-decision pdfs (d) Q-Q plot non-decision pdfs

Figure 5. Application 2: A selection of estimation results of a diffusion model analysis of post-error slowing for 39 participants (for a complete parameter overview please see supplemental material). Panel (a) shows the within-person differences in traditional parameter estimates between post-error and post-correct conditions for boundary separation a, inter-trial variability in drift rate for words and non-words (η_W and η_{NW}), and the drift rates for six different word types and non-words ($v_1, ..., v_6$ and v_{NW}). Statistically significant effects for the differences (two-sided sign test) are indicated with a green/red ($p < 0.001$) or light green/light red ($p < 0.01$) marking of the label (green means a positive effect or a larger value post-error compared to post-correct, red a negative effect or a smaller value post-error compared to post-correct). Panel (b) is the D*M version of panel (a). Panel (c) shows the participant averaged non-decision time densities inferred from the D*M estimates, separately for all six different word types (black) and non-words (red). The non-decision time densities of the post-correct condition are shown in the upper half, those of the post-error condition are shown, mirrored, in the lower half. Panel (d) is a quantile-quantile plot of the data in panel (c), and is better suited to look at the differences between the non-decision time densities from the two conditions. (Because of the many non-decision time pdfs, no confidence intervals are shown in panels (c) and (d).)

non-decision parameters on top of the 14 decision model parameters), while D*M can handle the problem with only the 14 decision model parameters. Also, there is no guarantee that this alternative parametric non-decision time distribution will (always) suffice.

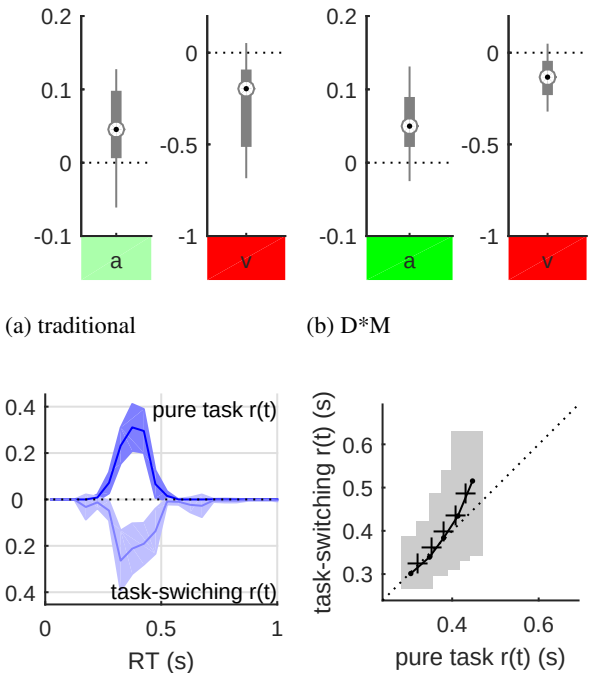
Application 3: A diffusion model analysis of task switching costs (Schmitz & Voss, 2012)

Schmitz and Voss (2012) investigate which diffusion model parameters can best explain task-switching costs. We limit our analyses to the first experiment in the paper, which compares task-switching and task-repeating trials from a classical alternating runs paradigm to each other and to pure task trials (no task-switching within an experimental block). In these analyses, three separate diffusion models are estimated (with both methods): one for the pure task trials, one for the task-repeating trails and one for the task-switching trials. Partial results are shown in Figure 6. The main qualitative findings of the original paper pertaining to this experiment, namely a positive change in boundary separation a (or caution) and a negative change in drift rate between task-switching and pure task trials, were confirmed by our re-analysis, for both the traditional and the D*M method. Participants are more cautious in the task switch condition and at the same time, they process the information less well. Quantitatively, the parameter estimates again differ considerably from the original study. It can also be seen that the non-decision time distribution in the task switching condition has a much more outspoken skew to the right (compared to the pure task non-decision time pdf).

Discussion

Traditional parameter estimates of decision models to choice RT data, have been shown to be vulnerable to the misspecification of the extra non-decision component (Ratcliff, 2013). In this paper, we have proposed a solution to this problem. By means of a handy convolution between data and decision model distributions, hence called the D*M method, we were able to factor out the non-decision time distribution from the estimation procedure. Through theory and a simulation study, the method was shown to remediate the traditional bias related to the misspecification of the non-decision time distribution, even for a limited number of data points. Additionally, the actually simulated non-decision time distributions could systematically be recovered. To illustrate the method's relevance, we applied it to three existing diffusion model application papers. For all studies, we found substantial differences with the traditional parameter estimates; in one case, using the D*M method radically altered the conclusions of the original paper, clearly demonstrating the necessity of the method.

In two of the three applications we re-analyzed, non-parametric estimates of the non-decision time distributions



(a) traditional (b) D*M

(c) non-decision pdfs (d) Q-Q plot non-decision pdfs

Figure 6. Application 3: A selection of estimation results of a diffusion model analysis of task switching costs for 24 participants (for a complete parameter overview please see supplemental material). Panel (a) shows the within-person differences in traditional parameter estimates between task-switching trials and pure task blocks for boundary separation a and drift rate v. Statistically significant effects for the differences (two-sided sign test) are indicated with a green/red ($p < 0.001$) or light green/ light red ($p < 0.01$) indicator (green means a larger parameter value in the task switch condition compared a pure task condition, red means smaller values). Panel (b) is the D*M version of panel (a). Panel (c) shows the participant averaged non-decision time densities inferred from the D*M estimates in panel (b). The non-decision time densities of the pure task condition are shown in the upper half of the plot, those of the task-switching trials are shown, mirrored, in the lower half. The solid lines show the mean non-decision pdfs across participants, the lighter areas display the double standard error interval. Panel (d) is a quantile-quantile plot of the data in panel (c), and is better suited to look at the differences between the non-decision pdfs from the two conditions. The grey area represents a 95% confidence interval of the mean quantile-quantile values (black crosses).

revealed a clear right skew (application 1 and 2). In our simulation study we have shown that ignoring such a right skew may cause severe biases in the diffusion model parameter estimates. Therefore, in the absence of a good model for non-decision time, we strongly advise the use of the D*M method for estimating choice RT models. Additionally, the non-parametric estimates of the resulting non-decision time distributions give a detailed picture of what non-decision time can look like (assuming the decision model itself is well-specified that is), which could in turn be used as a guide or template for constructing more intricate models for non-decision time.

References

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010, December). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, *107*(50), 21767–21772. doi: 10.1073/pnas.0908104107

Brown, S. D., & Heathcote, A. (2008, November). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Cressie, N., & Holland, P. W. (1983, March). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, *48*(1), 129–141. doi: 10.1007/BF02314681

DasGupta, S., Ferreira, C. H., & Miesenböck, G. (2014, May). FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*, *344*(6186), 901–904. doi: 10.1126/science.1252114

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2011, November). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454–465. doi: 10.3758/s13414-011-0243-2

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Cramon, D. Y. v., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008, November). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*(45), 17538–17542. doi: 10.1073/pnas.0805903105

Heathcote, A., & Brown, S. (2004, June). Reply to Speckman and Rouder: A theoretical basis for QML. *Psychonomic Bulletin & Review*, *11*(3), 577–578. doi: 10.3758/BF03196614

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002, June). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*(2), 394–401. doi: 10.3758/BF03196299

Karlin, S., & Studden, W. J. (1966). *Tchebycheff systems: with applications in analysis and statistics*. Interscience Publishers.

Krajbich, I., & Rangel, A. (2011, August). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857. doi: 10.1073/pnas.1101328108

Kreyszig, E. (2010). *Advanced Engineering Mathematics*. John Wiley & Sons.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York, NY: Springer New York.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012, February). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience*, *32*(7), 2335–2343. doi: 10.1523/JNEUROSCI.4156-11.2012

Nobile, A. G., Ricciardi, L. M., & Sacerdote, L. (1985, September). Exponential Trends of First-Passage-Time Densities for a Class of Diffusion Processes with Steady-State Distribution. *Journal of Applied Probability*, *22*(3), 611–618. doi: 10.2307/3213864

Polanía, R., Krajbich, I., Grueschow, M., & Ruff, C. C. (2014, May). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, *82*(3), 709–720. doi: 10.1016/j.neuron.2014.03.014

Ratcliff, R. (2013, January). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*, *120*(1), 281–292. doi: 10.1037/a0030775

Ratcliff, R., & Dongen, H. P. A. V. (2011, July). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*(27), 11285–11290. doi: 10.1073/pnas.1100483108

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347 –356. doi: 10.1111/1467-9280.00067

Ratcliff, R., & Tuerlinckx, F. (2002, September). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. doi: 10.3758/BF03196302

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009, September). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. doi: 10.1038/nature08275

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 222–250. doi: 10.1037/a0026003

Smith, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time data. *Psychological Bulletin*, *108*(3), 533–550. doi: 10.1037/0033-2909.108.3.533

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341–359.

Topsoe, F. (2000, July). Some inequalities for information divergence and related measures of discrimination. *IEEE Transactions on Information Theory*, *46*(4), 1602–1609. doi: 10.1109/18.850703

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. doi: 10.1037/0033-295X.108.3.550

Verdonck, S., & Tuerlinckx, F. (2014). The Ising Decision Maker: A binary stochastic network for choice response time. *Psychological Review*, *121*(3), 422–462. doi: 10.1037/a0037012

Voss, A., & Voss, J. (2007, November). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775. doi: 10.3758/BF03192967

Wagenmakers, E.-J., Maas, H. L. J. V. D., & Grasman, R. P. P. P. (2007, February). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. doi: 10.3758/BF03194023

Wong, K.-F., & Wang, X.-J. (2006, January). A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *The Journal of Neuroscience*, *26*(4), 1314–1328. doi: 10.1523/JNEUROSCI.3733-05.2006

Appendix A
Unique global optimum

In this appendix, we show under which conditions the objective function $D_{pp'}(\boldsymbol{\theta})$ of Equation 2, for $\hat{g}_p \rightarrow K * f_p$ and $\hat{g}_{p'} \rightarrow K * f_{p'}$, has a unique global optimum at the true parameter vector $\boldsymbol{\theta}_0$. We will start from the reasonable assumption that the decision model does not contain a non-decision component of its own at the true parameter vector $\boldsymbol{\theta}_0$, or, more specifically, that for the true parameter vector $\boldsymbol{\theta}_0$, there exists no distribution $u(t)$ and parameter vector $\boldsymbol{\theta}_u \neq \boldsymbol{\theta}_0$ such that for every p:

$$m_p(\boldsymbol{\theta}_0) = m_p(\boldsymbol{\theta}_u) * u. \tag{10}$$

Imagine for a moment that such a $u(t)$ and $\boldsymbol{\theta}_u \neq \boldsymbol{\theta}_0$ do exist. The fundamental Equation 1 may then be transformed as (for every p, p'):

$$\begin{aligned}
f_p * m_{p'}(\boldsymbol{\theta}_0) &= f_{p'} * m_p(\boldsymbol{\theta}_0) \\
f_p * m_{p'}(\boldsymbol{\theta}_u) * u &= f_{p'} * m_p(\boldsymbol{\theta}_u) * u \\
f_p * m_{p'}(\boldsymbol{\theta}_u) &= f_{p'} * m_p(\boldsymbol{\theta}_u) \\
m_p(\boldsymbol{\theta}_0) * r * m_{p'}(\boldsymbol{\theta}_u) &= m_{p'}(\boldsymbol{\theta}_0) * r * m_p(\boldsymbol{\theta}_u) \\
m_p(\boldsymbol{\theta}_0) * m_{p'}(\boldsymbol{\theta}_u) &= m_{p'}(\boldsymbol{\theta}_0) * m_p(\boldsymbol{\theta}_u),
\end{aligned}$$

where we make use of the fact convolution is a commutative operator and that all factors have a Laplace transform (see below in section A.2 for more information) such that they can be safely deconvolved from both sides. Thus, the final identity is a direct consequence of the presence of a non-decision component inside the decision model at $\boldsymbol{\theta}_0$. Note that another way this final identity can be reached, is when for all p, p': $m_p(\boldsymbol{\theta}_0) = m_{p'}(\boldsymbol{\theta}_0)$ and, as a consequence of Equation 10, $m_p(\boldsymbol{\theta}_u) = m_{p'}(\boldsymbol{\theta}_u)$. This would mean that all total response time pdfs f_p would be equal across condition-response pairs. Now we can move to the actual proposition.

Proposition

If the decision model itself does not contain a common non-decision component that can be factored out, or more specifically, if there does not exist a parameter $\boldsymbol{\theta}_u \neq \boldsymbol{\theta}_0$ such that for every p, p':

$$m_p(\boldsymbol{\theta}_0) * m_{p'}(\boldsymbol{\theta}_u) = m_{p'}(\boldsymbol{\theta}_0) * m_p(\boldsymbol{\theta}_u), \tag{11}$$

the minimum of

$$D_{pp'}(\boldsymbol{\theta}) = d\left(\hat{g}_p * m_{p'}(\boldsymbol{\theta}), \hat{g}_{p'} * m_p(\boldsymbol{\theta})\right),$$

for $\hat{g}_p \rightarrow K * f_p$ and $\hat{g}_{p'} \rightarrow K * f_{p'}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is unique.

Proof

Assume there does exist another global minimum at $\boldsymbol{\theta}_u \neq \boldsymbol{\theta}_0$ for which

$$D_{pp'}(\boldsymbol{\theta}_u) = d\left(\hat{g}_p * m_{p'}(\boldsymbol{\theta}_u), \hat{g}_{p'} * m_p(\boldsymbol{\theta}_u)\right) = 0$$

for $\hat{g}_p \rightarrow K * f_p$ and $\hat{g}_{p'} \rightarrow K * f_{p'}$.

Then, for every p, p'

$$\begin{aligned}
K * f_p * m_{p'}(\boldsymbol{\theta}_u) &= K * f_{p'} * m_p(\boldsymbol{\theta}_u) \\
K * r * m_p(\boldsymbol{\theta}_0) * m_{p'}(\boldsymbol{\theta}_u) &= K * r * m_{p'}(\boldsymbol{\theta}_0) * m_p(\boldsymbol{\theta}_u)
\end{aligned}$$

Because every factor of this equation has a Laplace transform (see below in Appendix B for more information), we can safely deconvolve K and r from both sides:

$$m_p(\boldsymbol{\theta}_0) * m_{p'}(\boldsymbol{\theta}_u) = m_{p'}(\boldsymbol{\theta}_0) * m_p(\boldsymbol{\theta}_u) \tag{12}$$

This is in contradiction with the initial assumption of Equation 11, which states that there is no such parameter $\boldsymbol{\theta}_u \neq \boldsymbol{\theta}_0$. This proves the proposition.

Appendix B
Conditions for the existence of a deconvolution

In order to formulate conditions for the existence of the deconvolution in Equation 5, we make use of cumulant-generating functions. The cumulant-generating function $[\mathcal{G}(d)](s)$ of a probability density function $d(t)$ is defined as

$$[\mathcal{G}(d)](s) = \log\left(\int_0^\infty e^{st} d(t) dt\right).$$

Thus, the cumulant-generating function is the logarithm of the moment generating function. Because the range of integration is the positive half of the real line, we can write the cumulant-generating function also as the logarithm of the Laplace transform Feller (1971) evaluated at $-s$. Because of this link with the Laplace transform, we can use the existence conditions for the Laplace transform (Kreyszig, 2010). It follows that for a probability density function $d(t)$ that is piecewise continuous and of exponential order, i.e. $d(t) \leq Me^{-kt}$ with M and k real positive numbers, the cumulant generating function $[\mathcal{G}(d)](s)$ exists, at least for $s < k$.[1] When we do not need the argument s, we will also write the cumulant-generating function in short as $\mathcal{G}(d)$.

The cumulant-generating function $\mathcal{G}(d)$ is determined by the coefficients κ_n^d $(n > 0)$ of its MacLaurin expansion, and these are called the cumulants:

$$[\mathcal{G}(d)](s) = \sum_{n=1}^{\infty} \kappa_n^d \frac{s^n}{n!}.$$

Note that κ_2^d is the variance of $d(t)$, denoted as σ_d^2 in the main text.

The cumulant-generating function is a convenient tool when dealing with convolutions because it treats them additively. Consider another pdf $e(t)$, with cumulant generating function $\mathcal{G}(e)$, then

$$\mathcal{G}(d * e) = \mathcal{G}(d) + \mathcal{G}(e).$$

Obviously, the additivity of the cumulant-generating function is transferred to the cumulant coefficients:

$$\kappa_n^{d*e} = \kappa_n^d + \kappa_n^e,$$

which holds for all n $(n > 0)$.

In our approach, we assume that for both the model $m_p(t; \boldsymbol{\theta}_0)$ and the non-decision time distribution $r(t)$, the corresponding cumulant-generating function exists. Most choice RT models are diffusion models and have an exponential decay (Nobile, Ricciardi, & Sacerdote, 1985), thereby satisfying the exponential order condition. In these cases, $m_p(t; \boldsymbol{\theta}_0)$ always has a corresponding cumulant-generating function $[\mathcal{G}(m)](s; \boldsymbol{\theta}_0)$ or $\mathcal{G}(m; \boldsymbol{\theta}_0)$ in short. For the unknown non-decision time distribution $r(t)$ we will simply assume the existence of a cumulant generating function $\mathcal{G}(r)$ implying the existence of a cumulant-generating function $\mathcal{G}(f_p)$ as well.

Because the cumulants, and therefore also the cumulant-generating function, can be expressed in terms of moments, sufficient conditions for a valid cumulant-generating function can be seen as sufficient conditions for a set of moments to lead to a valid pdf. The latter conditions, for a non-negative random variable, are a well known set of inequalities (Cressie & Holland, 1983; Karlin & Studden, 1966). More specifically, given a set of $J + 1$ numbers $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_J)$, if

$$\Delta_{2k} = \det\begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & & \vdots \\ \mu_k & \mu_{k+1} & \cdots & \mu_{2k} \end{bmatrix} \geq 0$$

(for $2k \leq J$) and

$$\Delta_{2k+1} = \det\begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \mu_2 & \mu_3 & \cdots & \mu_{k+2} \\ \vdots & \vdots & & \vdots \\ \mu_{k+1} & \mu_{k+2} & \cdots & \mu_{2k+1} \end{bmatrix} \geq 0$$

(for $2k + 1 \leq J$), there exists a density function $d(t)$, so that $\int_0^\infty t^j d(t) dt = \mu_j$ for $j = 0, .., J$.

For the first three inequalities (i.e., $J = 2$) this means:

$$\Delta_0 = \mu_0 \geq 0$$

$$\Delta_1 = \mu_1 \geq 0$$

$$\Delta_2 = \mu_0\mu_2 - \mu_1\mu_1 = \mu_2 - \mu_1^2 = \kappa_2 \quad \geq \quad 0.$$

The first two inequalities are trivial and always fulfilled in this context. The third inequality reduces to a constraint on the second order cumulants that are being subtracted:

$$\kappa_2^r = \kappa_2^{f_p} - \kappa_2^{m_p} \geq 0,$$

or

$$\kappa_2^{f_p} \geq \kappa_2^{m_p}.$$

In terms of variances σ^2 (the terminology and notation used in the main text), this inequality becomes:

$$\sigma_{f_p}^2 \geq \sigma_{m_p}^2.$$

[1] The importance of the exponential order condition can be shown as follows: $[\mathcal{G}(d)](s) = \log\left(\int_0^\infty e^{st} d(t) dt\right) \leq \log\left(\int_0^\infty e^{st} M e^{-kt} dt\right) = \log\left(M \int_0^\infty e^{-(k-s)t} dt\right) = \log M - \log(k - s)$, which shows that the integral exists for all $s < k$.

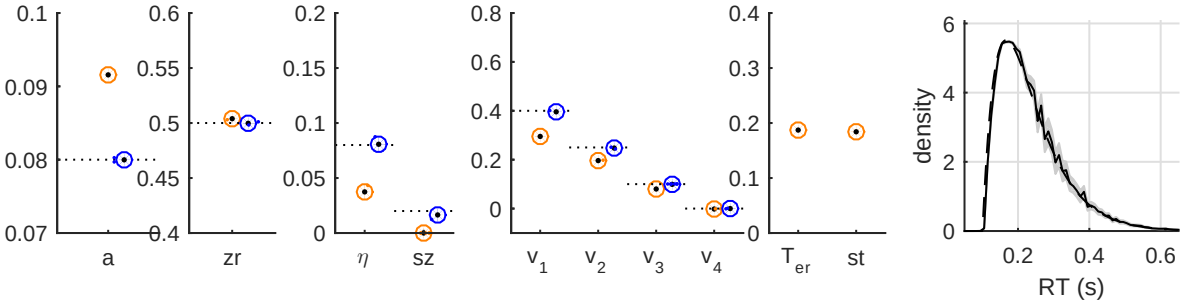
More intricate constraints apply to higher moments and cumulants. For our purposes, however, the higher order criteria are ignored during the minimization process, as their estimates become increasingly volatile for a finite number of data points.

Factoring out non-decision time in choice RT data: supplemental information

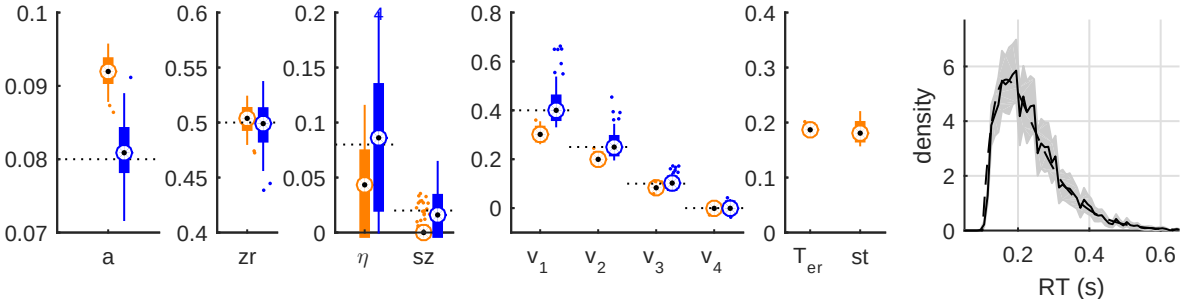
Stijn Verdonck and Francis Tuerlinckx
Faculty of Psychology and Educational Sciences
KU Leuven, University of Leuven

Simulation study

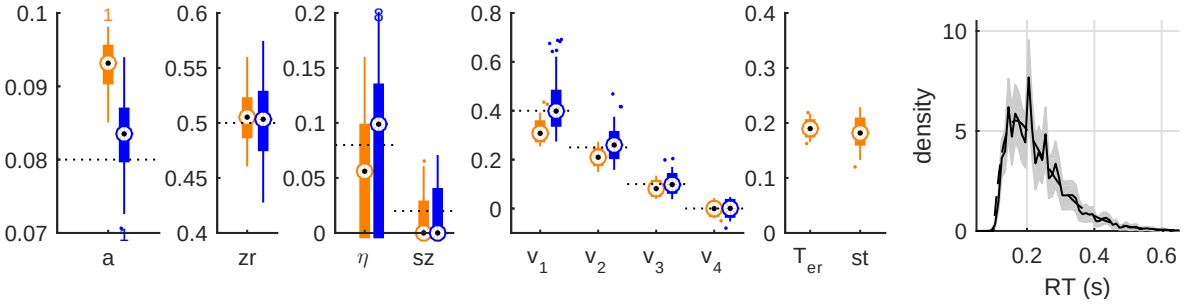
For the details on the simulation study, we refer to the main text. In Figures 1,2,3 below we present the recovery results of all diffusion model parameters as listed in Table 1. In each figure we show the results for a single non-decision pdf scenario's, and three different sample sizes.



(a) recovery diffusion model parameters for 10^6 observations per condition

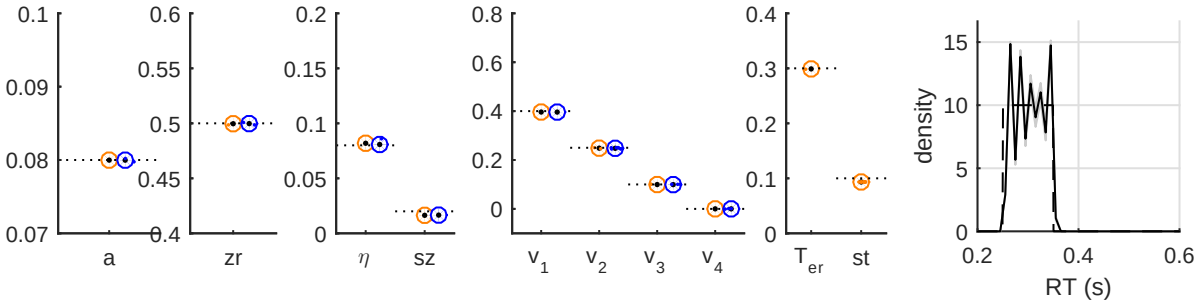


(b) recovery diffusion model parameters for 1000 observations per condition

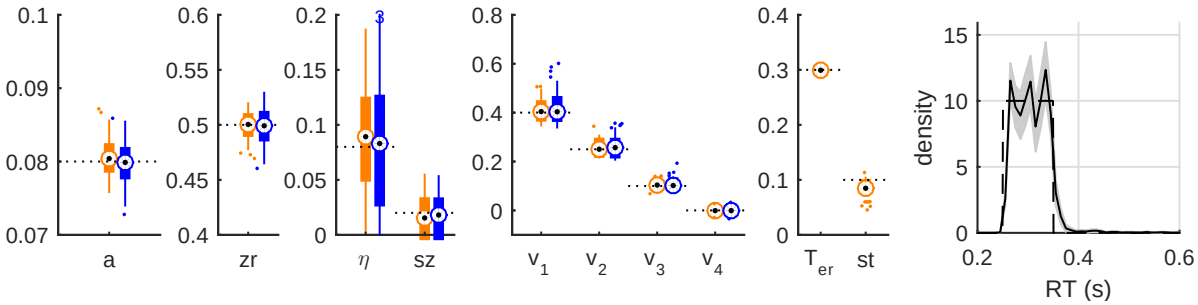


(c) recovery diffusion model parameters for 300 observations per condition

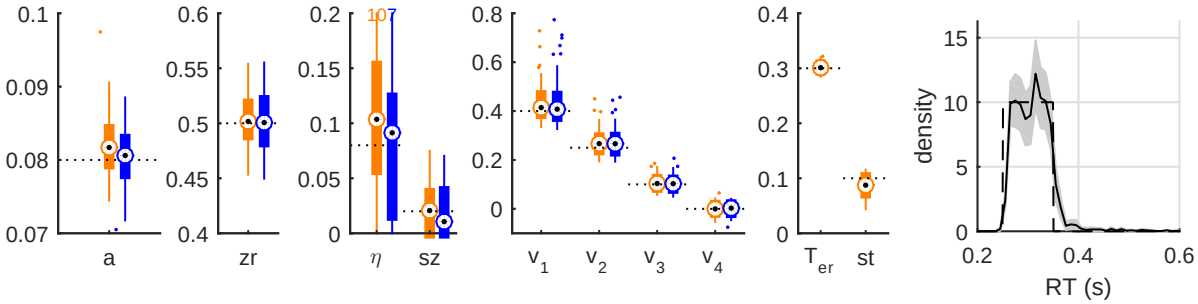
Figure 1. Recovery results for a right skewed non-decision pdf scenario. The three panels (a), (b), and (c) refer to three different sample sizes: 10^6, 1000 and 300 per condition. Each panel shows the estimates of the diffusion model parameters common to both the classical analysis (orange) and the D*M analysis (blue): boundary separation a, relative bias zr, drift rate variability η, starting point variability sz and the four drift rates v_1, v_2, v_3, v_4. In addition, the estimated T_{er} and st (mean and width of the uniform non-decision time distribution) for the classical analysis are shown. The rightmost figure contains the true (dashed line) and estimated (solid line) non-decision time pdf based on the D*M analysis.



(a) recovery diffusion model parameters for 10^6 observations per condition

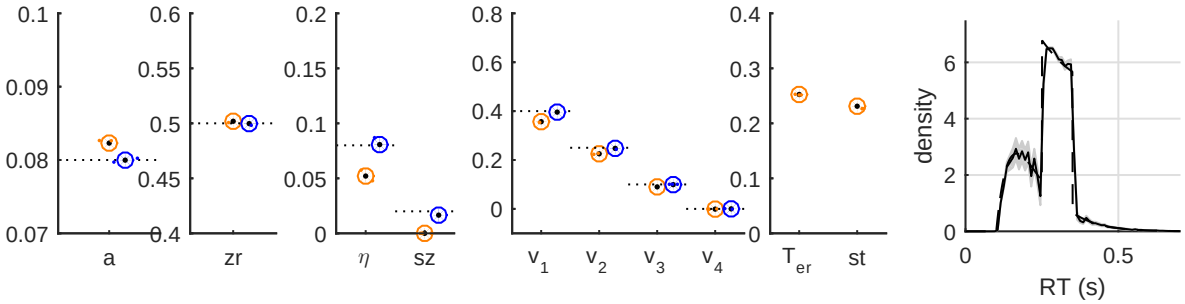


(b) recovery diffusion model parameters for 1000 observations per condition

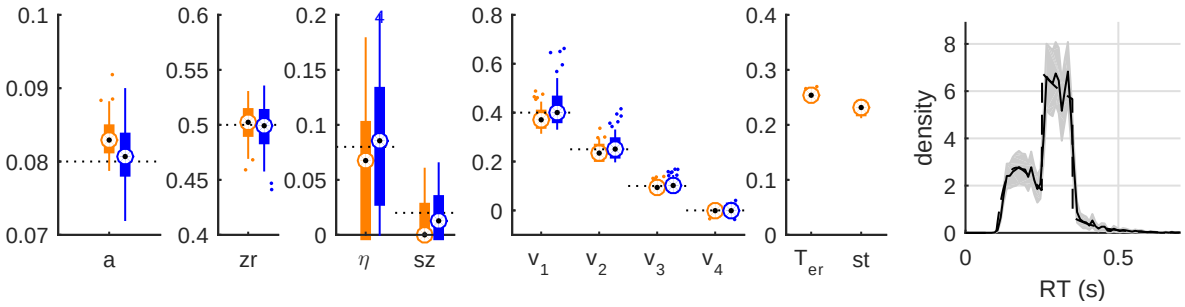


(c) recovery diffusion model parameters for 300 observations per condition

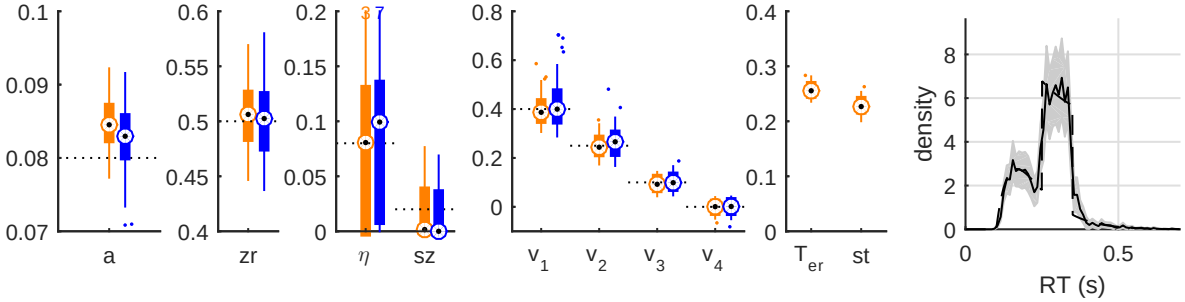
Figure 2. Recovery results for a uniform non-decision pdf scenario. The three panels (a), (b), and (c) refer to three different sample sizes: 10^6, 1000 and 300 per condition. Each panel shows the estimates of the diffusion model parameters common to both the classical analysis (orange) and the D*M analysis (blue): boundary separation a, relative bias zr, drift rate variability η, starting point variability sz and the four drift rates v_1, v_2, v_3, v_4. In addition, the estimated T_{er} and st (mean and width of the uniform non-decision time distribution) for the classical analysis are shown. The rightmost figure contains the true (dashed line) and estimated (solid line) non-decision time pdf based on the D*M analysis.



(a) recovery diffusion model parameters for 10^6 observations per condition



(b) recovery diffusion model parameters for 1000 observations per condition

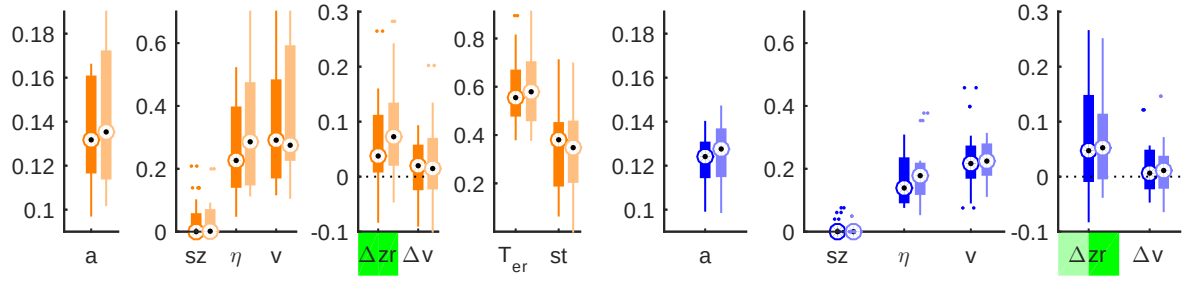


(c) recovery diffusion model parameters for 300 observations per condition

Figure 3. Recovery results for a bimodal non-decision pdf scenario. The three panels (a), (b), and (c) refer to three different sample sizes: 10^6, 1000 and 300 per condition. Each panel shows the estimates of the diffusion model parameters common to both the classical analysis (orange) and the D*M analysis (blue): boundary separation a, relative bias zr, drift rate variability η, starting point variability sz and the four drift rates v_1, v_2, v_3, v_4. In addition, the estimated T_{er} and st (mean and width of the uniform non-decision time distribution) for the classical analysis are shown. The rightmost figure contains the true (dashed line) and estimated (solid line) non-decision time pdf based on the D*M analysis.

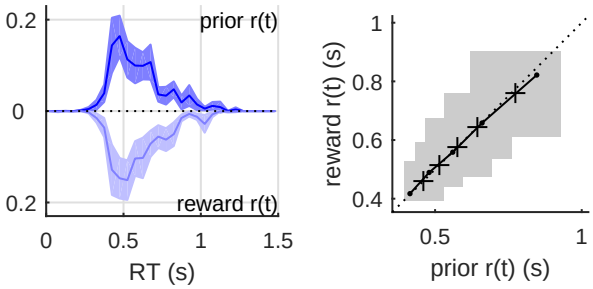
Application 1: A diffusion application of choice bias

In Figure 4 we present the full set of results (for all parameters) of the traditional and D*M diffusion model fits to the data from Mulder, Wagenmakers, Ratcliff, Boekel, and Forstmann (2012).



(a) traditional estimates

(b) D*M estimates



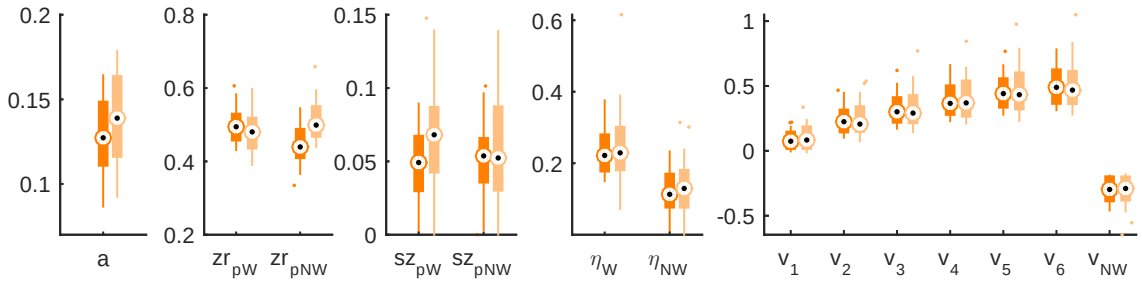
(c) non-decision pdfs

(d) Q-Q plot non-decision pdfs

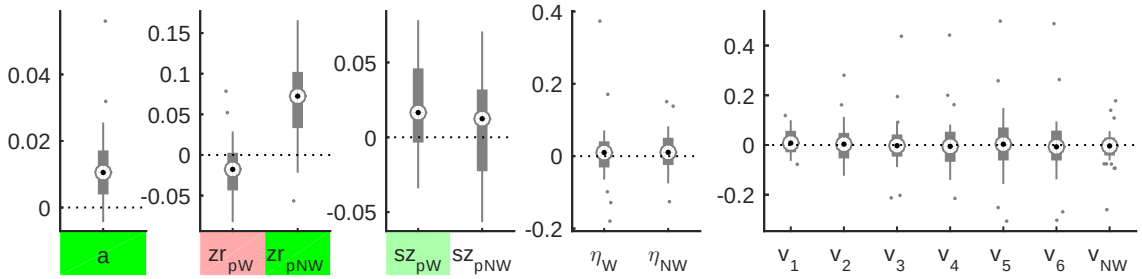
Figure 4. Application 1: Parameter estimates of a diffusion model analysis of choice bias for 20 participants. The first two panels show the estimates of all decision model parameters (boundary separation a, inter-trial variability of bias sz, inter-trial variability of drift rate η and effects of bias on starting point Δzr and drift rate Δv), obtained with either the traditional (a, in orange, with the additional non-decision time parameters mean non-decision time T_{er} and inter-trial variability of non-decision time st) or D*M method (b, in blue). For each of the two panels, the darker box plots show the estimates for the elevated prior likelihood condition, the lighter plots for the larger potential pay-off condition. If for the group of participants Δzr or Δv is significantly different from 0 (two-sided sign test), this is indicated with a green ($p < 0.001$) or light green ($p < 0.01$) marking of the label. Panel (c) shows the non-decision time densities inferred from the D*M estimates in panel (b). The densities for the elevated likelihood condition are shown in the upper half of the plot and those for the larger potential pay-off condition are shown, mirrored, in the lower half. The solid lines show the mean non-decision pdfs across participants, the lighter areas display the double standard error interval. Panel (d) is a quantile-quantile plot of the data in panel (c), and is better suited to look at the differences between the non-decision pdfs from the two conditions. The grey area represents a 95% confidence interval of the mean quantile-quantile values (black crosses).

Application 2: A diffusion application of post-error slowing

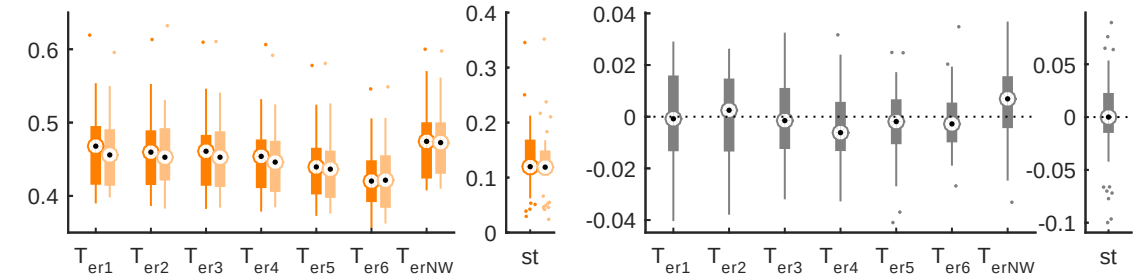
In the figures below, we present the full set of results (for all parameters) of the traditional (Figures 5,7) and D*M (Figure 6) diffusion model fits to the data from Dutilh et al. (2011).



(a) decision model estimates



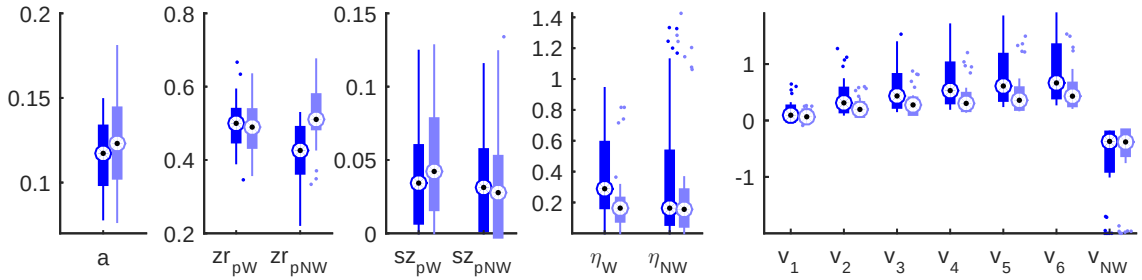
(b) decision model differences



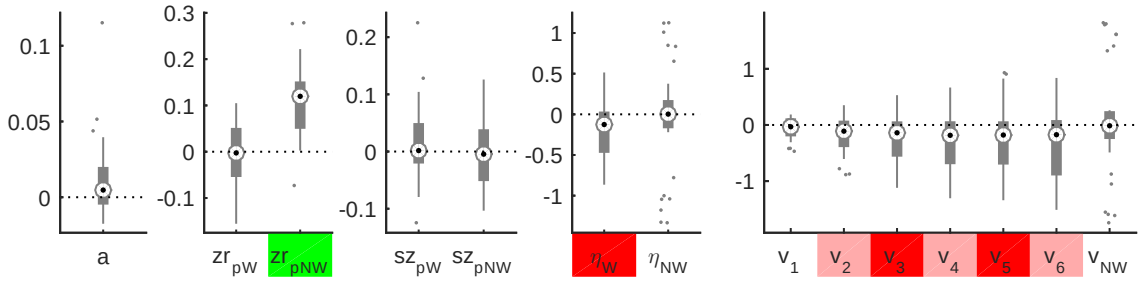
(c) non-decision estimates

(d) non-decision differences

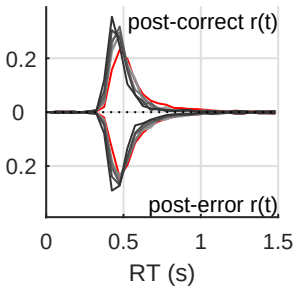
Figure 5. Application 2: Estimation results of a traditional diffusion model analysis of post-error slowing for 39 participants. Panel (a) shows the traditional estimates for boundary separation a, relative bias after a word or non-word error trial (zr_{pW} and zr_{pNW}), inter-trial variability of bias after a word or non-word error trial (sz_{pW} and sz_{pNW}), inter-trial variability of drift rate for words and non-words (η_W and η_{NW}), and the drift rates for six different word types and non-words ($v_1, ..., v_6$ and v_{NW}), for both the post-correct condition (darker box plots) and the post-error condition (lighter box plots). Panel (b) shows the within-person differences between post-error and post-correct conditions of the parameters in panel (a). Statistically significant effects for the differences (two-sided sign test) are indicated with a green/red ($p < 0.001$) or light green/light red ($p < 0.01$) marking of the label (green means a positive effect or a larger value post-error compared to post-correct, red a negative effect or a smaller value post-error compared to post-correct). Panel (c) shows the estimates of the uniform non-decision time distributions: the mean per stimulus type ($T_{er1}, T_{er2}, T_{er3}, T_{er4}, T_{er5}, T_{er6}$ and T_{erNW}) and the common width st. Panel (d) shows the within-person differences between post-error and post-correct conditions of the parameters in panel (c).



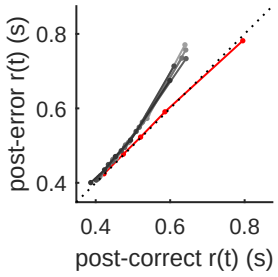
(a) decision model estimates



(b) decision model differences

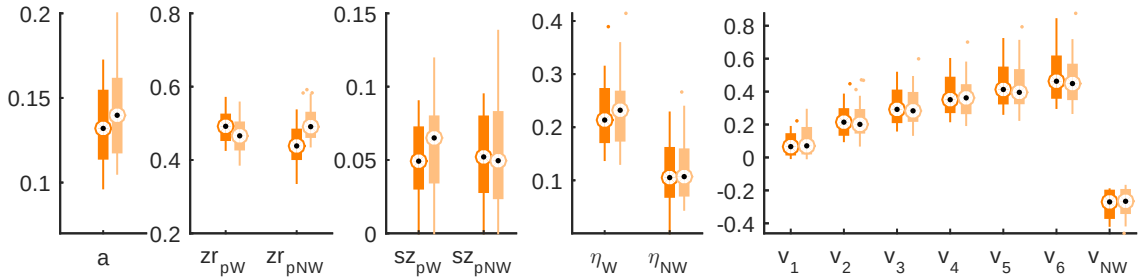


(c) non-decision pdfs

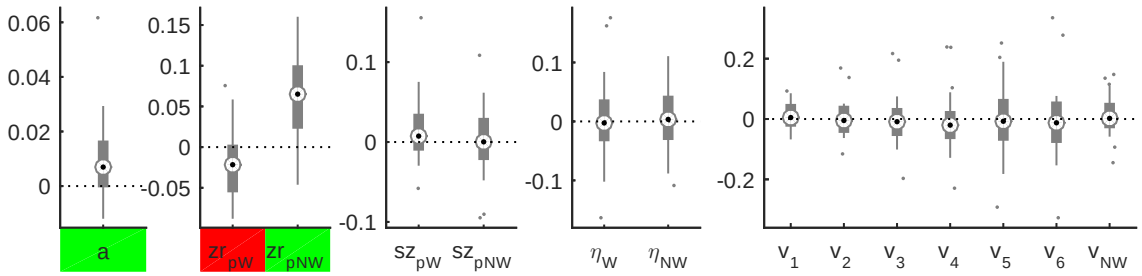


(d) Q-Q plot non-decision pdfs

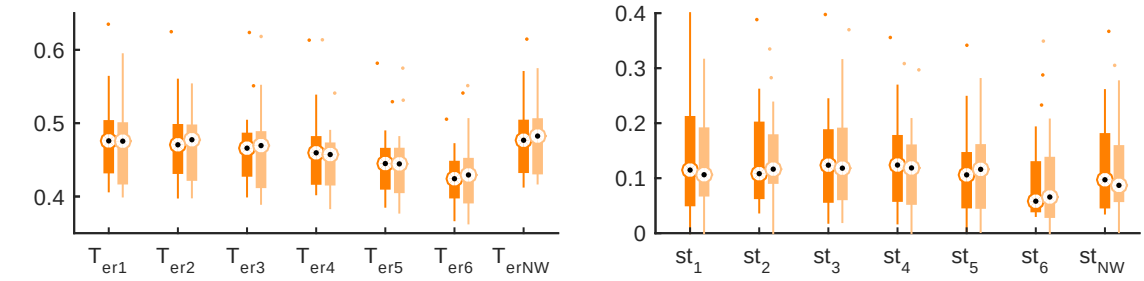
Figure 6. Application 2: Estimation results of a D*M diffusion model analysis of post-error slowing for 39 participants. Panel (a) shows the D*M estimates for boundary separation a, relative bias after a word or non-word error trial (zr_{pW} and zr_{pNW}), inter-trial variability of bias after a word or non-word error trial (sz_{pW} and sz_{pNW}), inter-trial variability of drift rate for words and non-words (η_W and η_{NW}), and the drift rates for six different word types and non-words ($v_1, ..., v_6$ and v_{NW}), for both the post-correct condition (darker box plots) and the post-error condition (lighter box plots). Panel (b) shows the within-person differences between post-error and post-correct conditions of the parameters in panel (a). Statistically significant effects for the differences (two-sided sign test) are indicated with a green/red ($p < 0.001$) or light green/light red ($p < 0.01$) marking of the label (green means a positive effect or a larger value post-error compared to post-correct, red a negative effect or a smaller value post-error compared to post-correct). Panel (c) shows the participant averaged non-decision time densities inferred from the D*M estimates, separately for all six different word types (black) and non-words (red). The non-decision time densities of the post-correct condition are shown in the upper half, those of the post-error condition are shown, mirorred, in the lower half. Panel (d) is a quantile-quantile plot of the data in panel (c), and is better suited to look at the differences between the non-decision time densities from the two conditions. (Because of the many non-decision time pdfs, no confidence intervals are shown in panels (c) and (d).)



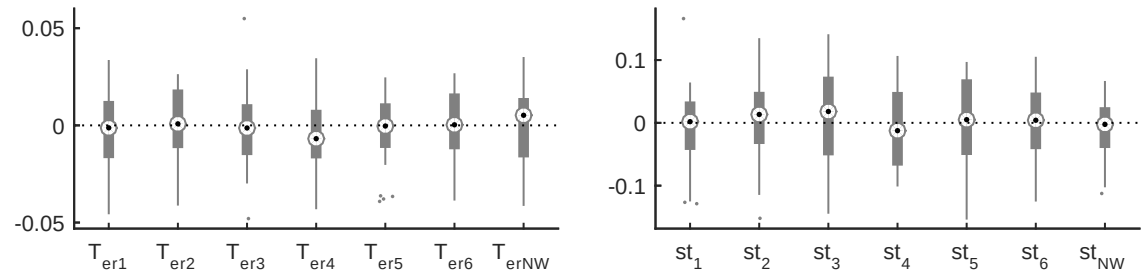
(a) decision model estimates



(b) decision model differences



(c) non-decision model estimates

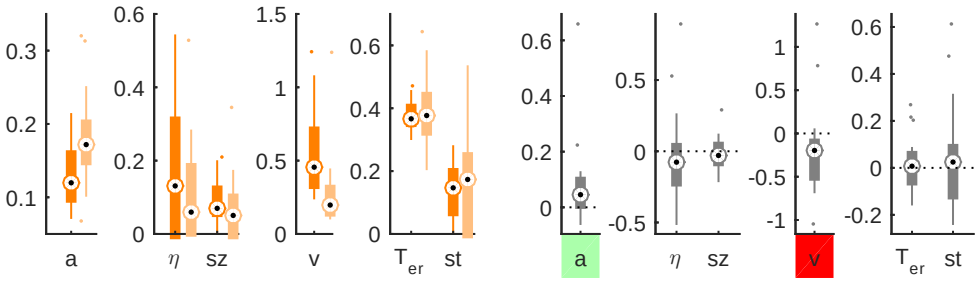


(d) non-decision model differences

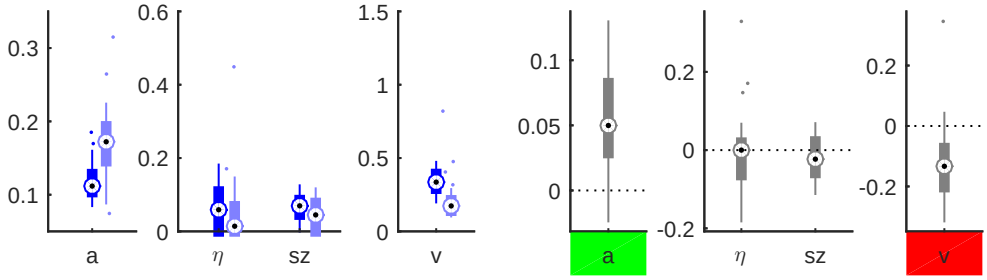
Figure 7. Application 2: Estimation results of a traditional diffusion model analysis of post-error slowing for 39 participants, with separate inter-trial variabilities of non-decision time for different stimulus types. Panel (a) shows the traditional estimates for boundary separation a, relative bias after a word or non-word error trial (zr_{pW} and zr_{pNW}), inter-trial variability of bias after a word or non-word error trial (sz_{pW} and sz_{pNW}), inter-trial variability of drift rate for words and non-words (η_W and η_{NW}), and the drift rates for six different word types and non-words ($v_1, ..., v_6$ and v_{NW}), for both the post-correct condition (darker box plots) and the post-error condition (lighter box plots). Panel (b) shows the within-person differences between post-error and post-correct conditions of the parameters in panel (a). Statistically significant effects for the differences (two-sided signtest) are indicated with a green/red ($p < 0.001$) or light green/light red ($p < 0.01$) marking of the label (green means a positive effect or a larger value post-error compared to post-correct, red a negative effect or a smaller value post-error compared to post-correct). Panel (c) shows the estimates of the uniform non-decision time distributions: the mean per stimulus type ($T_{er1}, T_{er2}, T_{er3}, T_{er4}, T_{er5}, T_{er6}$ and T_{erNW}) and the width per stimulus type ($st_1, st_2, st_3, st_4, st_5, st_6$ and st_{NW}). Panel (d) shows the within-person differences between post-error and post-correct conditions of the parameters in panel (c).

Application 3: A diffusion application of task switching costs

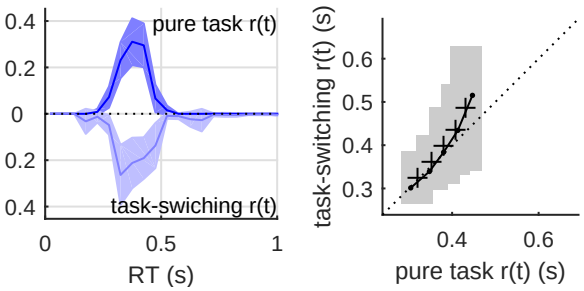
In the figures below, we present the full set of results (for all parameters) of the traditional and D*M diffusion model fits to the data from Schmitz and Voss (2012). In Figure 8, task-switching trials are compared to pure task trials; in Figure 8, task-repeating trials are compared to pure task trials.



(a) traditional estimates (b) traditional differences

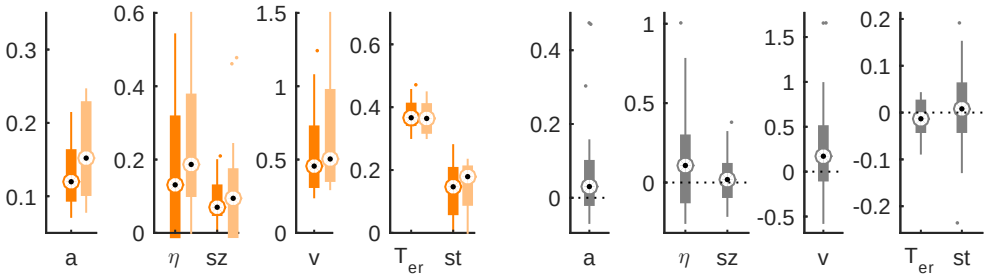


(c) D*M estimates (d) D*M differences



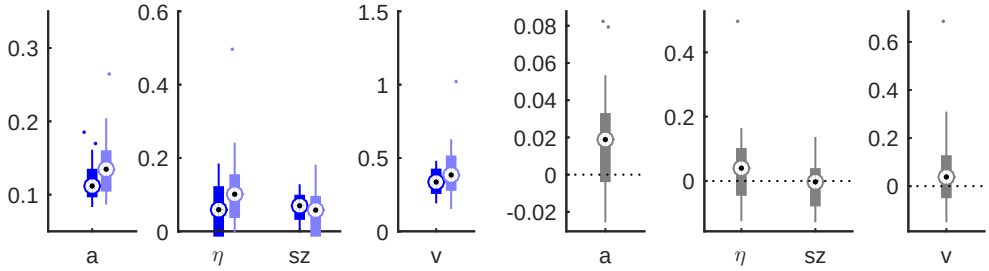
(e) non-decision pdfs (f) Q-Q plot non-decision pdfs

Figure 8. Application 3: Estimation results of a diffusion model analysis of task switching costs for 24 participants. Panel (a) shows the traditional estimates of all diffusion model parameters (boundary separation a, inter-trial variability of drift rate η, inter-trial variability of bias sz, drift rate v, uniform non-decision time distribution mean T_{er} and width st), for both the task-switching trials (lighter box plots) and the pure task trials (darker box plots). Panel (b) shows the within-person differences between task-switching and pure task conditions of the parameters in panel (a). Statistically significant effects for the differences (two-sided sign test) are indicated with a green/red ($p < 0.001$) or light green/ light red ($p < 0.01$) indicator (green means a larger parameter value in the task switch condition compared a pure task condition, red means smaller values). Panels (c) and (d) are the respective D*M versions of panels (a) and (b). Panel (e) shows the participant averaged non-decision time densities inferred from the D*M estimates. The non-decision time densities of the pure task condition are shown in the upper half of the plot, those of the task-switching trials are shown, mirrored, in the lower half. The solid lines show the mean non-decision pdfs across participants, the lighter areas display the double standard error interval. Panel (f) is a quantile-quantile plot of the data in panel (e), and is better suited to look at the differences between the non-decision pdfs from the two conditions. The grey area represents a 95% confidence interval of the mean quantile-quantile values (black crosses).



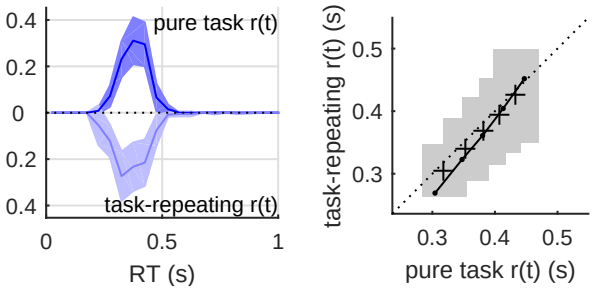
(a) traditional estimates

(b) traditional differences



(c) D*M estimates

(d) D*M differences



(e) non-decision pdfs

(f) Q-Q plot non-decision pdfs

Figure 9. Application 3: Estimation results of a diffusion model analysis of task switching costs for 24 participants. Panel (a) shows the traditional estimates of all diffusion model parameters (boundary separation a, inter-trial variability of drift rate η, inter-trial variability of bias sz, drift rate v, uniform non-decision time distribution mean T_{er} and width st), for both the task-repeating trials (lighter box plots) and the pure task trials (darker box plots). Panel (b) shows the within-person differences between task-switching and pure task conditions of the parameters in panel (a). Statistically significant effects for the differences (two-sided sign test) are indicated with a green/red ($p < 0.001$) or light green/ light red ($p < 0.01$) indicator (green means a larger parameter value in the task switch condition compared a pure task condition, red means smaller values). Panels (c) and (d) are the respective D*M versions of panels (a) and (b). Panel (e) shows the participant averaged non-decision time densities inferred from the D*M estimates. The non-decision time densities of the pure task condition are shown in the upper half of the plot, those of the task-switching trials are shown, mirrored, in the lower half. The solid lines show the mean non-decision pdfs across participants, the lighter areas display the double standard error interval. Panel (f) is a quantile-quantile plot of the data in panel (e), and is better suited to look at the differences between the non-decision pdfs from the two conditions. The grey area represents a 95% confidence interval of the mean quantile-quantile values (black crosses).

References

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2011, November). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454–465. doi: 10.3758/s13414-011-0243-2

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012, February). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience*, *32*(7), 2335–2343. doi: 10.1523/JNEUROSCI.4156-11.2012

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 222–250. doi: 10.1037/a0026003