Journal of Mathematical Psychology 76 (2017) 25–50

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Tutorial

# Parameter recovery for the Leaky Competing Accumulator model

Steven Miletić [a], Brandon M. Turner [b], Birte U. Forstmann [a], Leendert van Maanen [a,*]

[a] Department of Psychology, University of Amsterdam, Netherlands
[b] Department of Psychology, The Ohio State University, United States

## HIGHLIGHTS

- There is no closed-form solution to the likelihood function of the Leaky Competitive Accumulator (LCA) model.
- There are strong trade-offs between accumulation rate, leakage, and inhibition in the LCA model.
- Therefore, it is extremely difficult to faithfully recover the parameters of the LCA model.

## ARTICLE INFO

*Article history:*
Received 13 May 2016
Received in revised form
19 December 2016
Available online 17 January 2017

*Keywords:*
Leaky competitive accumulator model
Decision making
Parameter recovery
Simulation-based fitting

## ABSTRACT

The Leaky Competitive Accumulator (LCA) model for perceptual discrimination is rapidly growing in popularity due to its neural plausibility. The model assumes that perceptual choices and associated response times are the consequence of the accrual of evidence for the various response alternatives up to a certain predetermined threshold. In addition, accrual of evidence is influenced by temporal leakage of information and mutual inhibition between the accumulators. In this paper we provide an overview of fitting routines that may be used to identify the parameter values used for generating data under the LCA assumptions. We find that because (a) there is no closed-form solution to the likelihood function of the LCA model, and (b) there are strong trade-offs between accumulation rate, leakage, and inhibition, it is extremely difficult to faithfully recover the parameters of the LCA model. To minimize these problems, we recommend to use DE-MCMC sampling, collect very large amounts of data, and constrain the parameter space where possible.



## 1. Introduction

The ability to quickly and accurately make decisions about the environment is of crucial importance to survival. Whether an animal near you is predator or prey means the difference between being lunch and having lunch.

To understand how perceptual decision-making works, formal sequential sampling models (SSMs) are often used. The fundamental idea of SSMs is that, for each choice option (e.g., predator or prey), there is a latent accumulator that gathers noisy evidence from sensory input. Evidence accumulation continues until an accumulator reaches a level of activation that exceeds a certain threshold, and a decision is made.

The most influential and benchmark SSM is the diffusion decision model (DDM, Ratcliff & McKoon, 2008). The DDM proposes that evidence accumulation in a two-choice decision is characterized by Eq. (1):

$$dx = Adt + cdW \tag{1}$$

where $dx$ is the change in activation of an accumulator, $A$ is the average speed of evidence accumulation (drift rate), and $dW$ is Gaussian noise with standard deviation $c$. At the start of each trial, the accumulator starts at point 0 (i.e., no evidence) between $[-a, a]$. During a trial, the accumulator collects evidence, and the gathered amount of evidence moves towards $a$ or $-a$. The stochastic nature of the accumulation process makes it possible to make an incorrect decision (e.g., hit the lower threshold with a positive drift rate), and causes variation in the speed of the decision process, resulting in distributions of reaction times. Once one of the boundaries is reached, a response is initiated corresponding to the boundary that is reached. The actual reaction time is the decision time, plus a non-decision time to account for early perceptual processes and motor processes. Extensions of the standard DDM include between-trial variability in drift rates and starting point (Ratcliff, 1978), and sometimes variability in the non-decision time (Ratcliff & Tuerlinckx, 2002).

* Corresponding author.
*E-mail address:* lvmaanen@gmail.com (L. van Maanen).

http://dx.doi.org/10.1016/j.jmp.2016.12.001
0022-2496/

The use of SSMs is appealing for both practical and theoretical reasons. Practically, SSMs explain both reaction time distributions and accuracy data at the same time very well, unlike, for example, signal detection theory (Ratcliff & McKoon, 2008). In addition, SSMs can account for a wide range of findings in reaction time data, including the speed–accuracy trade-off (e.g., Boehm, Van Maanen, Forstmann, & Van Rijn, 2014; Forstmann et al., 2008; Mulder et al., 2013), and Hick's law (Usher, Olami, & McClelland, 2002; Van Maanen et al., 2012). Theoretically, SSMs are appealing because they provide information about latent processes instead of raw behavioral measures, and as such, they are informative for the mechanistic processes underlying behavior (Anders, Alario, & Van Maanen, 2016; Forstmann, Ratcliff, & Wagenmakers, 2015; Mulder, Van Maanen, & Forstmann, 2014; Van Maanen, Grasman, Forstmann, & Wagenmakers, 2012).

Although the DDM and related SSMs were primarily developed to model cognitive processes, several lines of research have shown that evidence accumulation happens on a neuronal level as well. For example, animal studies have reported single cell recordings showing evidence accumulation (e.g. in the LIP, FEF, Churchland & Ditterich, 2012; Gold & Shadlen, 2007; Purcell et al., 2010; Shadlen & Newsome, 2001, but see Latimer, Yates, Meister, Huk, & Pillow, 2015; Shadlen et al., 2016). In addition, EEG-studies have shown ERP-components that correlate remarkably well with drift rate (O'Connell, Dockree, & Kelly, 2012) or boundary (Boehm et al., 2014). This work is accompanied by more detailed biophysical models of neuronal networks that on a relatively gross level are consistent with the sequential sampling framework as well (Lo & Wang, 2006; Wang, 2008).

It has been argued (Usher & McClelland, 2001) that SSMs make assumptions about the evidence accumulation process that are implausible when considering the behavior of pools of neurons. In particular, the assumption of full evidence integration is argued to be problematic: neural excitatory input decays within a very short time period (5–10 ms) (Abbott, 1991). A mechanism of recurrent self-excitation delays the decay of information to some extent, but not perfectly (Amit, Brunel, & Tsodyks, 1994). Hence, information decays over time, so accumulators in a SSM should be 'leaky' integrators. To reconcile SSMs with findings in neurophysiology, neurally plausible models have been proposed.

### 1.1. The leaky, competing accumulator model

One particular such model is the leaky, competing accumulator model (LCA, Usher & McClelland, 2001). The LCA (Fig. 1) is an $n$-choice version of the DDM, including two neuroscientifically motivated parameters: leakage (sometimes called decay) and lateral inhibition. Evidence accumulation in the LCA is described by:

$$\begin{aligned} dx_i &= \left[ I_i - \kappa x_i - \beta \sum_{i' \neq i} x_{i'} \right] \frac{dt}{\tau} + \xi_i \sqrt{\frac{dt}{\tau}} \\ x_i &= \max(x_i, 0) \end{aligned} \tag{2}$$

where $dx_i$ is the change in activation of accumulator $i$, $I$ is input (c.f. drift rate in the DDM), $\kappa$ is leakage, $\beta$ is inhibition, $\frac{dt}{\tau}$ is the time step size, and $\xi$ is noise which is drawn from a Gaussian distribution when the model is simulated (Fig. 1). The second equation is a non-linear component preventing the activation of an accumulator to drop below 0. The idea behind this is that because neurons cannot have 'negative' activation, neither should accumulators (but see below). Response times are obtained through the first-passage time distribution with a common threshold level of activation $Z$, and non-decision time $NDT$.

A few mechanistic differences between the LCA and the canonical DDM should be mentioned. The DDM is only applicable in two-choice situations (but see Leite & Ratcliff, 2010; McMillen & Holmes, 2006; Ratcliff & Starns, 2013; Smith, 2016), as it assumes one accumulator that gathers evidence for both choice options. The sign of the accumulator (positive or negative) determines whether there is more evidence for choice option one or two. The DDM includes competition: every piece of evidence for choice option 1 always counts as evidence against choice option 2. The LCA assumes one accumulator for each choice option, and these accumulators race towards the threshold. The LCA also includes competition, but proposes a different competition mechanism: the inhibition parameter determines the proportion of one accumulator's activation that inhibits the other accumulators (see bottom right panel in Fig. 1).

The inclusion of the leakage and inhibition parameters is theoretically appealing, but needs to be validated by reaction time data that can only be explained by these parameters (Pitt & Myung, 2002). There have been some studies into the effect of inhibition (Churchland & Ditterich, 2012; Teodorescu & Usher, 2013; Tsetsos, Usher, & McClelland, 2011), but the importance of including a leakage parameter remains debated (Ossmy et al., 2013; Purcell et al., 2010; Ratcliff & Starns, 2013; Turner, Sederberg, & McClelland, 2016 but see Ratcliff, 2006; Winkel, Keuken, van Maanen, Wagenmakers, & Forstmann, 2014).

An important reason for the sparse amount of research into the LCA is that the model does not have a (known) likelihood function (Turner & Sederberg, 2014). The only methods available to fit the model to data are simulation-based, meaning that, in order to calculate any measure of fit, the LCA has to generate simulated data for each proposal set of parameters. As a consequence, model-fitting procedures are slow—so slow, that a thorough investigation of model fitting procedures, recoverability, and identifiability of the LCA has, as far as the authors are aware, not been performed.

To circumvent fitting a model without a known likelihood function, studies in which leakage and inhibition may be of relevance have often used the Ornstein–Uhlenbeck process (OU, Busemeyer & Townsend, 1993). The OU can be considered as a linear simplification of the LCA with a known likelihood function. The use of the OU instead of the full LCA is, in case of a two-choice paradigm, validated if we drop the assumption of non-linearity. Without the non-linearity, it can be shown that the evidence accumulation process as described by the LCA Eqs. (2) is equivalent to an OU-process (ignoring differences in the threshold parameter, see also van Ravenzwaaij, van der Maas, & Wagenmakers, 2012). If we define $x = x_2 - x_1$ (a single accumulator representing the difference in activation between the two accumulators in the LCA, c.f. the DDM), the two LCA-equations of the two accumulators:

$$\begin{aligned} dx_1 &= [I_1 - \kappa x_1 - \beta x_2] \frac{dt}{\tau} + \xi \sqrt{\frac{dt}{\tau}} \\ dx_2 &= [I_2 - \kappa x_2 - \beta x_1] \frac{dt}{\tau} + \xi \sqrt{\frac{dt}{\tau}} \end{aligned} \tag{3}$$

can be rewritten as one simple equation:

$$dx = [(I_2 - I_1) + (\beta - \kappa)x] \frac{dt}{\tau} + \xi \sqrt{2 \frac{dt}{\tau}}. \tag{4}$$

The major disadvantage of the use of the OU is that the leakage and inhibition parameters reduce to one 'difference' parameter, often named $\lambda$. Fitting an OU-process can thus only tell something about whether the difference value between leakage and inhibition is positive or negative; which means that we can only know whether there was more leakage than inhibition or vice versa.

This problem only occurs when the assumption of non-linearity is dropped. Excluding the non-linearity has been defended with theoretical and practical reasons. Theoretically, the motivation behind the assumption has been questioned: although the activation of neurons can indeed not decrease below 0, this argument might

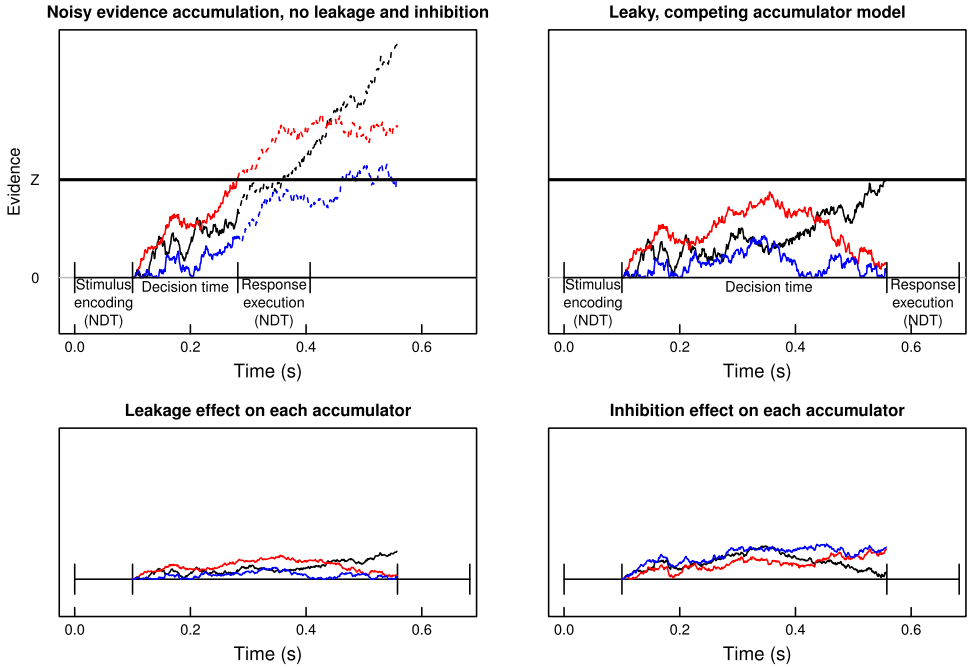


**Fig. 1.** Illustration of the main parameters in the decision process as proposed by the leaky competitive accumulator model (LCA), for a three-alternative choice. Top left panel: Noisy evidence accumulation, similar to an independent, non-leaky race model of choice. One trial is depicted, in which each accumulator gathers noisy evidence over time, until one accumulator reaches a common threshold $Z$. The time when this happens is called the first-passage time. The bottom panels show the potential effects of leakage ($\kappa$, left) and inhibition ($\beta$, right) on each accumulator in the same trial. The leakage panel shows the amount of activation that each accumulator would lose on every point in time due to leakage. The inhibition panel shows the amount of activation that each accumulator would lose on every point in time due to inhibition by the other accumulators. Finally, the top right panel shows the evidence accumulation process as proposed by the LCA. Note that, in this trial, the first passage time and choice differ from the independent race due to the influence of leakage and inhibition.

be based on the false assumption that 'inactive' neurons are not firing. In contrast, inactive neurons fire at a baseline rate, and it is known that activation can drop below baseline by inhibitory input (see, for this line of reasoning, Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). The practical reason for dropping the assumption of non-linearity is that it has been shown that, under reasonable parameter ranges (i.e., when all accumulators receive a substantial amount of input), the non-linear process can be mimicked very well by a linear process (Brown & Holmes, 2001; McMillen & Holmes, 2006; Usher & McClelland, 2001, 2004; van Ravenzwaaij et al., 2012).

Thus, in part of the parameter space, it will be impossible to disentangle the leakage and inhibition parameters in a two-choice LCA model. To ensure that these crucial parameters can be disentangled, we focus on three-choice settings, where the leakage and inhibition parameters can be disentangled in the whole parameter space. A thorough analysis of fitting the LCA on three-accumulator data has, as far as the authors are aware, not been performed (although fits to multi-alternative data have been reported, e.g., Teodorescu & Usher, 2013; Tsetsos et al., 2011). Further studies can use our findings to study the two-choice LCA.

The goal of the present article is twofold. Firstly, we want to share best practices in fitting likelihood-free cognitive models such as the LCA. In our attempts to perform parameter recovery for this model, we found that not all methods are equally suitable to handle the increased uncertainty that is a direct consequence from the lack of likelihood function. In particular, we focus on the Probability Density Approximation method (PDA, Turner & Sederberg, 2014). This method is the most suitable for obtaining a reliable estimate of the likelihood function, and we will demonstrate its use in the context of maximum likelihood estimation as well as a Bayesian parameter estimation procedure (the supplemental materials describe two additional methods).

Secondly, our aim is to study the ability to recover the data-generating parameters of the LCA model. The ability to recover parameters is of crucial importance when model fitting is used as a method to gain insight into latent cognitive processes.

By studying the recoverability, we indirectly also investigate the *identifiability* of the LCA. That is, if the LCA is identified, there is only one unique set of parameters that predict a certain outcome (e.g., a probability density; see also Moran, 2016). The ability to recover data-generating parameters implies that the model is identified (although the reverse inference does not necessarily hold). In the context of the current research, this exercise will also allow us to learn which model fitting routine should be preferred for fitting the LCA model.

## 2. Maximum likelihood estimation

In general, maximum likelihood estimation (MLE) is the standard method of estimating the parameters of a model (e.g., Myung, 2003). MLE uses the likelihood function to describe the likelihood of a set parameters $\theta$ given the data set $D$:

$$\mathcal{L}(\theta|D) = \prod_{i=1}^{N} Model(D_i|\theta) \tag{5}$$

where $Model(D_i|\theta)$ is the probability density function of the model, describing the relative probability of observation $D_i$ under parameters $\theta$. The likelihood is the product of the probabilities of the individual observations $D_i$ under parameters $\theta$. In models with a known likelihood function, estimating the best set of parameters is a matter of maximizing this equation (Myung, 2003).

Maximum likelihood estimation is often preferred to methods based on summary statistics (e.g., quantile-based difference minimizing) because it avoids the problem of sufficiency (but see Anders et al., 2016). The problem of sufficiency entails that, without a likelihood function, there is no guarantee that a summary statistic such as a set of quantiles includes all the information of the original data. Consequently, there is a possibility

that information about the exact shape of the true distribution is lost when summarizing the distribution in a set of quantiles. As a consequence, parameter estimates might be inaccurate.

### 2.1. Probability density approximation

However, the LCA does not have a known likelihood function, which makes application of MLE impossible. To overcome this problem, recent studies have developed a method to perform approximate maximum likelihood estimation without the use of a likelihood function (Turner, Dennis, & Van Zandt, 2013; Turner & Sederberg, 2012, 2014; Turner et al., 2016; Turner & Van Zandt, 2014; Turner & Zandt, 2012). This method entails that we can approximate the likelihood function using kernel density estimation (KDE). By performing KDE on simulated reaction time data, we obtain a probability density distribution; i.e., a distribution that describes the likelihood of a certain reaction time and choice under that model. Such a likelihood is exactly what is needed to perform maximum likelihood estimation. To avoid confusion, the term 'pseudo-likelihood' is used to refer to likelihood estimates obtained by the PDA-method, which approximate the true likelihood.

Mathematically, the PDA method requires a simulated proposal data set $X$ under a proposal parameter set $\theta^*$. A kernel density estimate of $X$ can be obtained using (Silverman, 1986):

$$f(x|X) = \frac{1}{hJ}\sum_{j=1}^{J} K\left(\frac{x - X_j}{h}\right) \tag{6}$$

where $\int f(x|X)\,dx = 1$. The kernel density estimate requires a kernel function $K(\cdot)$ and a bandwidth function $h(\cdot)$. The kernel function used in these analyses is an Epanechnikov kernel (Eq. (7)), as suggested by Turner and Sederberg (2014).

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{if } |x| \le 1 \\ 0 & \text{if } |x| \ge 1. \end{cases} \tag{7}$$

However, initial explorations showed that the use of different kernels does not alter the performance of the PDA method substantively. For a bandwidth function, Silverman's rule of thumb can be used:

$$h = 0.9 \min\left(\mathrm{SD}(X), \frac{\mathrm{IQR}(X)}{1.34}\right) n^{-\frac{1}{5}} \tag{8}$$

where SD is the standard deviation, IQR the interquartile range, and $n$ the number of observations. Other options in choosing a bandwidth are available (Sheather, 2004, see also below). To increase the accuracy of the kernel, all data were log transformed before estimating the density.

The kernel density estimate $f(D_i|X)$ provides a pseudo-likelihood of experimental data point $D_i$, given proposal data set $X$. The likelihood of a set of parameters $\theta$ given the full data set $D$ is approximated with the product of the likelihoods of individual observations $D_i$ given the proposal data set $X$:

$$\mathcal{L}(\theta|D) = \prod_{i=1}^{N} Model(D_i|\theta) \approx \prod_{i=1}^{N} f(D_i|X)\,. \tag{9}$$

The product of all the individual data points is a number very close to 0 (i.e., in the order of $10^{-1000}$ or below). Statistics software such as R cannot handle such numbers very well. Instead, a transformation of likelihoods to log-likelihoods is used:

$$\log \mathcal{L}(\theta|D) \approx \sum_{i=1}^{N} \log f(D_i|X). \tag{10}$$

When estimating log-pseudo-likelihoods of reaction time data with correct and incorrect answers (as is usually the case), kernel density estimation is done separately for correct and incorrect answers, and multiplied by the proportion of correct or incorrect answers:

$$\log \mathcal{L}(\theta|D) \approx \sum_{i=1}^{N_1} \log\left(f(D_{i,cor}|X_{cor}) * \frac{N_1}{N}\right) + \sum_{i=1}^{N_2} \log\left(f(D_{i,incor}|X_{incor}) * \frac{N_2}{N}\right) \tag{11}$$

where $D_{i,cor}$ and $D_{i,incor}$ are individual reaction times for correct and incorrect answers, $X_{cor}$ and $X_{incor}$ are the distributions of correct and incorrect reaction times observed under parameter set $\theta$, and $N_1$ and $N_2$ are the number of correct and incorrect responses, respectively.

### 2.2. Maximum likelihood estimation: simulation study

In this section we report the results of a simulation study aimed at recovering the data-generating parameters of LCA using MLE. The general procedure to perform a single parameter recovery is the following:

1. Generate a set of parameters $\theta$.
2. Simulate an 'experimental' data set $D$ using the LCA equation.
3. Generate a proposal set of parameters $\theta^*$.
4. Simulate a proposal data set $X$ under $\theta^*$.
5. Calculate the log-pseudo-likelihood of the proposal set of parameters $\theta^*$ given experimental data set $D$ (i.e., $\log \mathcal{L}(\theta^*|D)$).
6. Repeat steps 3–5 until the best set of parameters is found.

In order to understand the generalizability of the parameter recovery method to the full parameter space, we repeat steps 1–6. Ideally, the best fitting set of parameters $\theta^*_{best}$ is identical to the data-generating parameters with each recovery. In practice, there will always be some small deviations from this set of parameters due to noise both in the simulated experimental data set and the proposal data sets, so a more realistic aim is that $\theta^*_{best}$ should approximate $\theta$.

The experimental paradigm simulated here is a three-choice experiment with one correct answer (cf. Keuken et al., 2015; Van Maanen et al., 2012). The three accumulators receive a 'shared' amount of input $I$, the correct accumulator receives additional input $\Delta I$. Thus, there is an equal amount of evidence for both incorrect choice options. This way, only two input parameters need to be estimated; furthermore, it offers the constraint that $\Delta I$ should always be higher than 0. Note that this paradigm violates the sum-to-one input/drift assumption that is sometimes used in modeling studies, which may influence fitting performance (see Teodorescu & Usher, 2013). Hence, the parameters that need to be recovered are input difference $\Delta I$, input $I$, leakage $\kappa$, inhibition $\beta$, threshold $Z$, and non-decision time $NDT$. The variance of noise was fixed to $\xi = 0.1$ to avoid scaling problems.

#### 2.2.1. Parameter generation

All sets of parameters $\theta$ were sampled from uniform distributions:

$\Delta I \in \mathcal{U}(0.05, 0.3)$
$I \in \mathcal{U}(0.8, 1.2)$
$\kappa \in \mathcal{U}(1, 8)$
$\beta \in \mathcal{U}(1, 8)$
$Z \in \mathcal{U}(0.05, 0.25)$
$NDT \in \mathcal{U}(0.200, 0.500)$.

Although these distributions do not cover the full range of possible LCA-parameters, they do include the most reasonable

ranges (see also the discussion section). These parameter ranges are also used in previous studies that used a fixed variance of noise to avoid scaling problems (e.g., Bogacz et al., 2006; McMillen & Holmes, 2006; van Ravenzwaaij et al., 2012).

Some combinations of parameters generated unrealistic reaction time distributions. Parameter sets were rejected if they created too slow reaction times (i.e. decision times longer than 5 s), or if the range of observed reaction times was smaller than 400 ms.

#### 2.2.2. Data generation

A set of data was generated using the LCA (Eq. (2), including the non-linearity) with a high temporal resolution, $\frac{dt}{\tau} = \frac{1}{1000}$, corresponding to the temporal resolution of 1 ms that experimental data offer. Three sizes of data sets were considered: a data set consisting of 250 trials, one of 1000 trials, and one of 10,000 trials. 250 trials are a set size that is highly reasonable given practical experimental constraints. 1000 trials would still be practically possible, but would require either longer experimental sessions, or multiple sessions per subject. 10,000 trials constitute a very large data set, requiring multiple long experimental sessions. Although it is not very common to do so, these amounts of data are sometimes collected (see, e.g., Noorbaloochi, Sharon, & McClelland, 2015). For the purposes of simulation, such a large data set would provide a practical upper bound on the precision of parameter recovery. We generated and estimated 59 data sets with 250 trials, 47 with 1000 trials, and 37 with 10,000 trials.

#### 2.2.3. Proposal generation and simulation

Proposal parameter sets $\theta^*$ were generated by an optimization algorithm. The differential evolution algorithm used (see below) requires no starting set of parameters, only upper and lower bounds. The bounds chosen were $[0, 0.5]$ for $\Delta I$, $[0, 2]$ for $I$, $[0, 14]$ for $\kappa$ and $\beta$, $[0, 1]$ for $Z$, and $[0, 1]$ for $NDT$. Although it is possible to have parameters of leakage and inhibition higher than the maximum values chosen, the reaction time distributions produced in those regions are generally[1] too slow to be likely in real experimental data. Moreover, as the boundaries of the algorithm are always wider than the ranges of parameter values, these limits should not decrease parameter recovery performance.

Proposal data sets $X$ were simulated with 20,000 trials with $\frac{dt}{\tau} = \frac{3}{1000}$ to accurately represent the reaction time distribution.[2] Note that proposal data set simulation is the bottleneck of speed in the procedure, as for every $\theta^*$, a large data set needs to be simulated.

#### 2.2.4. Pseudo-likelihood estimation

The PDA method described above was used to obtain pseudo-likelihood estimates.

#### 2.2.5. Optimization

It is not feasible to try all possible sets of parameters $\theta^*$ to find the best-fitting parameters; instead, we use efficient optimization algorithms. In this study, a differential evolution algorithm was used, applying principles of evolution to optimization to quickly find a best possible solution and to avoid getting stuck in local minima (Ardia, Arango, & Gomez, 2011a; Ardia, Boudt, Carl, Mullen, & Peterson, 2011b; Ardia, Mullen, Peterson, & Ulrich, 2015; Conn, Gould, & Toint, 1991, 1997; Goldberg, 1989; Mullen, Ardia, Gil, Windover, & Cline, 2011; Price, Storn, & Lampinen, 2006; Turner & Sederberg, 2012; Turner, Sederberg, Brown, & Steyvers, 2013).

As we know the true underlying set of parameters $\theta$, we can reduce computation time by stopping the optimization algorithm once it has found a log-pseudo-likelihood as good as or higher than the log-pseudo-likelihood of the true set of parameters $\theta$: stop if $\log \mathcal{L}(D|\theta^*) \geq \log \mathcal{L}(D|\theta)$. Because of their simulation-based nature, there is unavoidably some noise in the log-pseudo-likelihood estimates. To ensure that the stopping rule does not bias against recovery, the log-pseudo-likelihood estimate of the data-generating parameter set was first estimated 1000 times, and the highest value was selected as the true log-pseudo-likelihood estimate.

The algorithm was stopped after 500 iterations if the true log-pseudo-likelihood was not found. This occurred in only 5% of the data sets, indicating that the optimization procedure had no trouble finding parameter sets that fitted the data at least as well as the true parameter sets. Therefore, it is highly unlikely that the results would benefit from using a different optimization algorithm.

To further ensure that this stopping rule did not bias against recovery, we also performed maximum likelihood estimation with a different stopping rule. In this procedure, the differential evolution algorithm was stopped if it was not able to improve its best known parameter set for 100 iterations, with a maximum of 10,000 iterations. The results of maximum likelihood estimation with this second stopping rule, shown in Fig. A.9, are comparable to the results with the first stopping rule.

### 2.3. Results

The results of the simulation study are summarized in Fig. 2, which depicts the individual true parameters of sets $\theta$ on the $x$-axes, and best fitting parameters $\theta^*$ on the $y$-axes. Every point in a graph represents a fit of an individual data set. Ideally, each point should be on the dashed diagonal line, meaning that the fitted parameter estimate is identical to the true parameter. This, however, is not the case. The difference in evidence accumulation and the non-decision time can be estimated very well. The threshold estimate is slightly less accurate, but the procedure performs worst in estimating input, leakage, and inhibition.

To quantify whether the recovery of the input, leakage, and inhibition parameter increases with increasing data set size, we applied a bootstrapping procedure. In this procedure, we estimated 20,000 correlation coefficients by repeatedly sampling from every data set size, and calculating the correlation between the true and fitted parameter values. The number of samples was equivalent to the number of recoveries for each data set size (i.e., 59, 47, and 37 for data set sizes of 250 trials, 1000 trials, and 10,000 trials, respectively). Then, we calculated the difference between the estimated correlation coefficients of 10,000 trials and 250 trials, between estimated correlations of data set sizes 10,000 trials and 1000 trials, and between estimated correlations of data set sizes 1000 trials and 250 trials. The probability that the correlation between true and fitted parameter values is higher for the larger data set size than for the lower data set size, is equal to the proportion of differences in estimates that is larger than 0.

Further, we applied the same bootstrapping procedure for the RMSE of the parameter estimates. In this case, the probability that the RMSE is higher for the larger data set size than for the smaller, is equal to the proportion of differences that is smaller than 0. All obtained probabilities can be found in Table 1.

[1] This is not always the case; depending on the values of input and threshold, higher leakage and inhibition parameters can generate realistic reaction time distributions.

[2] Although a 3-millisecond resolution is coarser than the resolution of the true data, the used temporal resolution is comparable to or higher than previous research (cf. Ratcliff & McKoon, 2008; Tsetsos et al., 2011; Turner et al., 2016; Usher & McClelland, 2001). Initial simulations showed that the difference in resolution between the experimental and proposal data sets did not influence the recovery performance.

**Table 1**
Probabilities of increasing correlations and decreasing RMSE-values with increasing data set size.

| Parameter | Data set sizes compared | Probability of increase in correlation | Probability of decrease in RMSE |
|---|---|---|---|
| Input | 1000–250 | 0.81 | 0.62 |
| | 10,000–1000 | 0.83 | 0.44 |
| | 10,000–250 | 0.90 | 0.75 |
| Leakage | 1000–250 | 0.94 | 0.85 |
| | 10,000–1000 | 0.76 | 0.51 |
| | 10,000–250 | 0.99 | 0.83 |
| Inhibition | 1000–250 | 0.88 | 0.85 |
| | 10,000–1000 | 0.77 | 0.95 |
| | 10,000–250 | 0.96 | 1.00 |

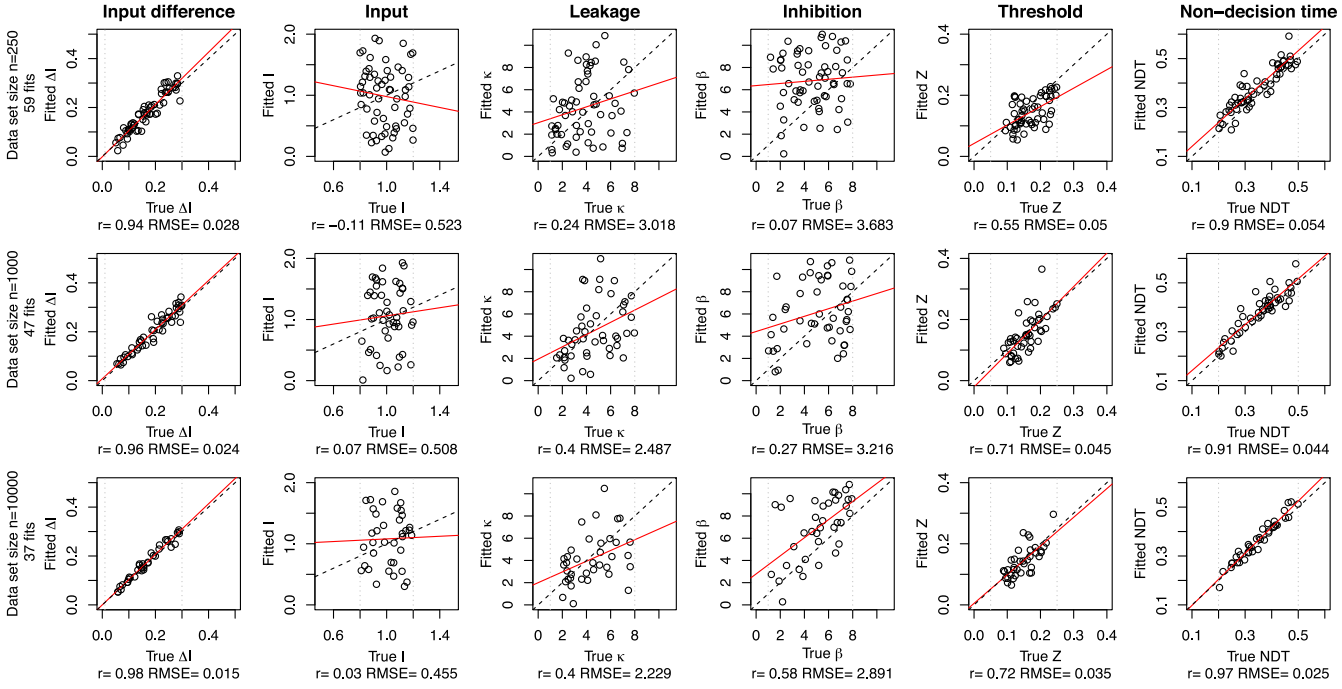


**Fig. 2.** Results of the parameter recovery study with maximum likelihood estimation for the three data set sizes (top row $N = 250$, middle row $N = 1000$, bottom row $N = 10{,}000$). Each point's position on the $x$-axis represents the parameter value used to generate the data set, the position on the $y$-axis the parameter estimate obtained by fitting. The vertical dashed lines correspond to the ranges of parameters used to generate data. The red line is the actual correlation between the data-generating parameter values and the fitted estimates. The correlation coefficients (Pearson's $r$) and RMSEs are shown underneath each panel. Ideally, all points in the graphs should be positioned on the diagonal line; i.e., each parameter estimate obtained by fitting should correspond to the true parameter value. The red and dashed line would coincide, with a correlation coefficient of $r = 1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Using correlation coefficients as a metric of recovery accuracy, for all parameters, there is a high probability (0.76–0.99) that accuracy improves with larger data set sizes. Using the RMSE as a metric of recovery accuracy, there is a high probability (0.85–1.00) that the inhibition parameter is recovered better with higher data set sizes. However, the recovery of the input and leakage parameters only seem to benefit from increasing data set size from 250 trials to 1000 trials, but increasing the number of trials to 10,000 does not lead to better estimates than using a data set size of 1000 trials.

### 2.3.1. Noise in pseudo-likelihood estimates

How good are the individual fits obtained using MLE with PDA? Fig. 3 shows a typical fit obtained using MLE on a 10,000 trial data set. The top panels depict the density of the data set $D$ (black line) with the density of a proposal data set $X$ of the LCA with the true parameters $\theta$ (red line). The bottom panels depict the density of the data and the density of the best fitting parameters $\theta^*$ on this data set.

A possible concern about the densities of the proposal and true data sets is noise in the log-likelihood estimates. If the amount of noise in the pseudo-likelihood estimate is large, using a point estimate of the pseudo-likelihood can have a significant influence on the ability to estimate parameters. The pseudo-likelihood of some $\theta^*$ can be higher than the pseudo-likelihood of $\theta$ due to chance, and consequently, an optimization procedure will rarely (if ever) find $\theta$. To explore how noisy log-pseudo-likelihood estimates are, we estimated the log-pseudo-likelihood of true parameters $\theta$ of a randomly generated 10,000-trial data set $D$ repeatedly for 1000 times. A scatterplot of these log-pseudo-likelihood estimates, shown in Fig. 4 (left panel), reveals an interesting pattern.

Intuitively, one would expect to find the log-pseudo-likelihoods to be normally distributed. Instead, repeatedly estimating the log-pseudo-likelihood of the true parameter values results in skewed clusters of likelihoods.

The cause of this cluster-pattern lies in the right tail of the reaction time distribution. To illustrate what happens, Fig. 4 (right panels) shows the right tail of data set $D$ (corresponding to the log-pseudo-likelihoods in the left panel). Points on the $x$-axis mark observations in $D$. The two lines are the kernel density estimates of a proposal data set $X$ with a high log-pseudo-likelihood (red line) and a proposal data set $X$ with a low log-pseudo-likelihood (black line). There are five observations in $D$ at which the density of the black-lined simulated set $X$ is 0. For practical reasons,

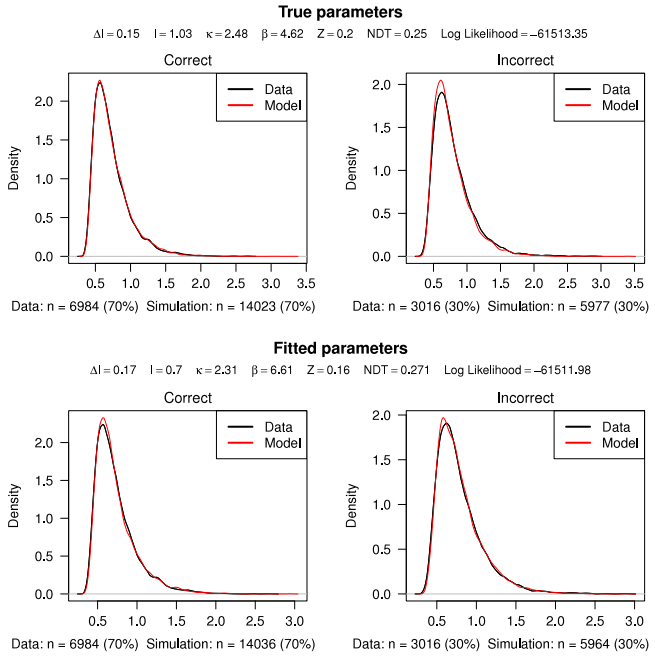


**Fig. 3.** Comparison of the fit of the true parameter values $\theta$ (top row) and the fit of the best-fitting estimate of the parameters $\theta^*$ on a randomly generated 10,000-trial data set. In every panel, the $x$-axis depicts the time in seconds. Density of the reaction time data is indicated by the black lines, the red lines indicate the density of a proposal data set generated using $\theta$ (top) or $\theta^*$ (bottom). For example noticeable is an overestimation of the density of the mode in the bottom panels. However, the log-pseudo-likelihood of $\theta^*$ is higher. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

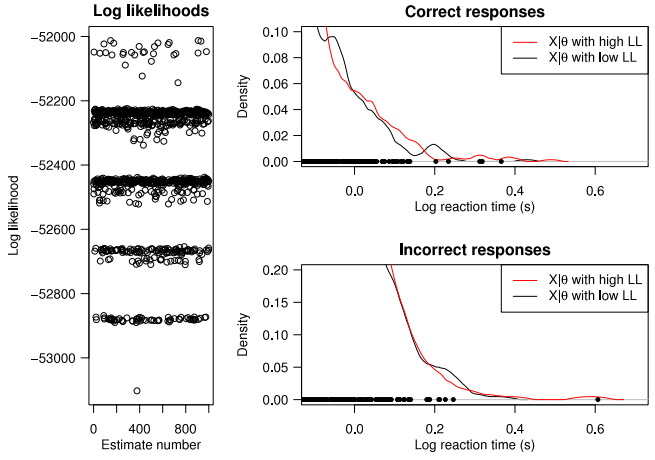


**Fig. 4.** Left panel: scatterplot of 1000 calculated log-pseudo-likelihoods of $\theta$ on a randomly generated data set $D$. Log-pseudo-likelihoods were obtained using the PDA method with $J = 20{,}000$ simulations with $\frac{dt}{\tau} = \frac{3}{1000}$. Right panels: Right tail of the density of the correct responses (top) and incorrect responses (bottom). Reaction times were log-transformed to increase the accuracy of the Epanechnikov kernel. Points on the $x$-axis mark observations in the data set $D$. Lines are densities of proposal data sets $X$ using the true parameter values. The red proposal data set has density at all (extreme) values in data set $D$, while the black proposal data set has a density of 0 at five observations in $D$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

densities of 0 were recoded to $10^{-99}$; a log-likelihood of 0 would be $-\inf$, a value that optimizers cannot handle. The log of $10^{-99}$ is approximately $-227$, corresponding to the average distance in log-pseudo-likelihood clusters. For every data point $D_i$ that, under a certain proposal data set $X$, has a density of 0, the log-pseudo-likelihood decreases with 227. Hence, the cluster-pattern is caused by observations in $D$ of which the density estimate is 0.

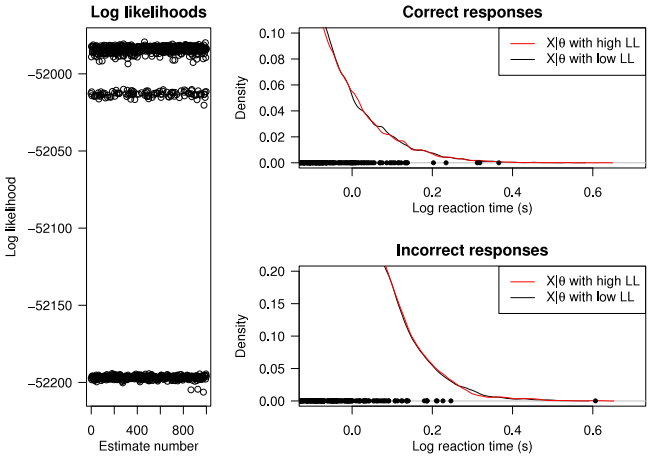


**Fig. 5.** Left panel: scatterplot of 1000 calculated log-pseudo-likelihoods of $\theta$ on a randomly generated data set $D$. Log-pseudo-likelihoods were obtained using the PDA method with $J = 1{,}000{,}000$ simulations with $\frac{dt}{\tau} = \frac{1}{1000}$. Right panels: Right tail of the density of the correct responses (top) and incorrect responses (bottom). Points on the $x$-axis mark observations in the data set $D$. The lines are densities of proposal data sets $X$ using the true parameter values. Note that reaction times are on a log-scale.

Whether or not this pattern will occur depends on the data set (the length of the right tail). Also of importance are the amount of trials simulated in the proposal data set relative to the simulation time step size, which determine the bandwidth relative to the minimal distance between data points in the proposal set. The consequence of these inaccuracies is that, when an optimizer generates proposal set $\theta^*$, there is a chance that the log-pseudo-likelihood greatly underestimates the true log-likelihood, and hence, the proposal data set $\theta^*$ will be wrongfully dismissed.

Is there a way around this issue? Recent studies have made advances in density estimation of heavy-tailed distributions (e.g., Buch-larsen, Nielsen, Guillen, & Bolance, 2005), but as far as the authors are aware, there is no definitive way to accurately estimate densities in both the tail and the body of a distribution at the same time. Increases in accuracies of the tail come at the cost of a decrease of accuracy in the rest of the distribution.

A brute-force solution would be to (greatly) increase the number of observations $J$ in the proposal data sets $X$. The reasoning is that increasing $J$ automatically increases the number of observations in the right tail, and thereby the accuracy of density estimation in the right tail. Exploratory simulations showed that proposal set sizes of up to $J = 1{,}000{,}000$ with a time step size of $\frac{1}{1000}$ indeed greatly decreased, but did not completely solve this issue (Fig. 5).

Would further increasing the number of simulations help? When increasing the number of simulations, one should keep in mind that as $J$ increases, the bandwidth of kernel density estimation decreases. At some $J$, the bandwidth will become too small: a kernel estimate at reaction time $t$ will not be smoothed by observations surrounding $t$, as these are outside of the bandwidth (see Supplementary Fig. A.10 for an illustration with data set $D$ with $10^8$ simulations). At which $J$ this happens depends on the data set (i.e., its standard deviation and interquartile range). Note that in the limit (infinite data points) no smoothing would be necessary, but as our Supplementary Fig. A.10 shows, smoothing is still necessary even with $10^8$ trials in a data set.

Using a different bandwidth selection method will probably not solve this problem, as other bandwidth selection methods generally produce smaller bandwidths as compared to Silverman's rule of thumb (most notably, the Sheather–Jones method; Sheather & Jones, 1991). In fact, it seems that more (instead of less) smoothing is necessary. Possibly, increasing the bandwidth is a

solution, but one should keep in mind that bandwidth selection methods available are developed to be optimal (as defined by Silverman, 1986). Increasing the bandwidth might cause more harm than good, as overall density estimates might be worse. Hence, increasing the bandwidth should be statistically motivated.

As another approach to this issue, one can slightly alter the optimization objective and use the following function:

$$\log \mathcal{L}(\theta|D) = \sum_{i=1}^{N} \log f(D_i|X) + m \log f(D_p|X) \tag{12}$$

where $N$ is the number of observations in the fastest 95% of the data, $m$ the number of observations in the slowest 5% of the data, and $D_p$ the mean reaction time of the slowest 5% of the data. In this alternative log-pseudo-likelihood function, instead of using density estimates of each individual data point, the slowest 5% of the data points in the right tail are binned. Data points in this bin are not appointed their individual pseudo-log-likelihoods (obtained using the PDA method), but the pseudo-log-likelihood of their mean reaction time. Simulations show that the zero-density problem discussed above does disappear when using this alternative log-pseudo-likelihood function: repeatedly estimating the log-pseudo-likelihood of a single parameter set on the same data result in seemingly normally distributed log-pseudo-likelihood estimates (see Supplementary Fig. A.11 for an example). Nevertheless, recovery performance did not benefit from this new log-pseudo-likelihood function, as shown in Supplementary Fig. A.12. Therefore, the zero-density problem that occurs when using the PDA method, although warranting further research, probably does not underlie the problems in recovering the input, leakage, and inhibition parameters of the LCA.

## 3. Differential evolution Markov chain Monte Carlo

Another popular method for estimating cognitive model parameters is Markov-chain Monte Carlo, typically in a Bayesian context (MCMC, e.g., Turner et al., 2013; Turner, Van Maanen, & Forstmann, 2015; Wiecki, Sofer, & Frank, 2013). In order to perform Bayesian MCMC estimation, a prior and a likelihood function are needed. Prior distributions reflect our knowledge of the parameters $\theta$ before estimation. In the current simulation study, uniform priors were used to reflect the fact that, before estimation, we know nothing of the underlying parameters except for that they should lie within the given bounds. The priors used were the following:

$\Delta I \sim \mathcal{U}(0, 3)$
$I \sim \mathcal{U}(0, 3)$
$\kappa \sim \mathcal{U}(0, 14)$
$\beta \sim \mathcal{U}(0, 14)$
$Z \sim \mathcal{U}(0, 10)$
$NDT \sim \mathcal{U}(0, 0.500)$.

Bayesian parameter estimation is not possible without an (approximate) likelihood function. The only way to do Bayesian parameter estimation with the LCA is to use the PDA method mentioned above. The posterior distributions can be approximated by combining the prior with the likelihood of a parameter set $\theta$:

$$\pi(\theta|D) \propto \pi(\theta)\mathcal{L}(\theta|D). \tag{13}$$

In differential evolution Markov-chain Monte Carlo sampling, a proposal parameter set is accepted with Metropolis Hastings probability:

$$\min(1, \pi(\theta^*|D)/\pi(\theta_t|D)). \tag{14}$$

Hence, if the probability of a proposal parameter set $\theta^*$ is higher than the probability of the current parameter set $\theta_t$, it is accepted. If the probability of $\theta^*$ is lower than the probability of $\theta_t$, the chance of acceptance is equal to the proportion of their probabilities.

Proposal sets $\theta^*$ were generated by a differential evolution Markov-chain Monte Carlo sampler (DE-MCMC; Turner et al., 2013). Conventional MCMC samplers have been shown to generate inefficient proposal sets $\theta^*$ when the parameters of a model are correlated. The DE-MCMC sampler proposes parameter sets $\theta^*$ based on the current parameter sets $\theta_t$ in other chains. This dependency increases the efficiency of the procedure, as proposals match the shape of the posterior distribution.

A downside of likelihood-free parameter estimation is that noise in the likelihood estimates is detrimental to performance. If likelihood estimates are noisy, chains get stuck: once a chain has found a likelihood that is artificially high due to noise, the probability of accepting a $\theta^*$ according to the Metropolis Hastings rule decreases. Any $\theta^*$ will only be accepted when its corresponding likelihood estimate is also artificially high to due noise.

Unfortunately, the PDA method is noisy, as discussed previously. To prevent stuck chains, the amount of noise in the log-pseudo-likelihood estimates was minimized by increasing the number of simulations per proposal data set to $J = 50{,}000$, with a time step size of $\frac{dt}{\tau} = \frac{2}{1000}$. Following Turner and Sederberg (2014), the DE-MCMC sampler was implemented using 50 chains. 1000 burn-in iterations were followed by 5000 sampling iterations, producing posteriors of 250,000 samples. Migration chance was set to 25% throughout to assist burn-in and prevent stuck chains.[3] The random perturbation factor was set to 0.001, the scaling factor $\gamma$ was randomly sampled from $\mathcal{U}(0.5, 1)$ for each proposal. Convergence of chains was assessed using Gelman's diagnostic (Brooks & Gelman, 1998; Gelman & Rubin, 1992).

### 3.1. Results

The obtained posterior distributions of all parameters and data set sizes are shown in Fig. 6. Interestingly, the pattern of results is comparable to the results obtained using maximum likelihood. The input difference, threshold, and non-decision time parameters can be recovered reasonably well, although the estimates of the input difference and non-decision time parameters are overestimated. Similar to the maximum likelihood estimates, the input, leakage, and inhibition parameters do not recover well. Because it is known from other models (e.g., the DDM and LBA) that the input or drift parameter and threshold parameters can show trade-offs, we hypothesized that part of the difficulty in recovering input, leakage, and inhibition are due to trade-offs between these three parameters and the threshold parameter.

To test this hypothesis, we repeated the DE-MCMC fitting procedure with the same data sets, but fixed the input difference, threshold, and non-decision time parameters to their true values. The results of this procedure are shown in Fig. 7. With real data, the true values of the input difference, threshold, and non-decision time parameters are not known, but it may be possible to use a two-stage fitting procedure instead. In such a two-stage procedure, the first stage should be designed to obtain highly accurate estimates of the input difference, threshold, and non-decision time parameter. It may be possible to use repeated maximum likelihood estimation (as in Appendix A.2), or perhaps even a different model (e.g., the LBA) to obtain such estimates. In the second stage, the obtained estimates of input difference, threshold, and non-decision time would be used to fix these parameters, leaving only the input, leakage, and inhibition parameters unfixed. In the

[3] Very similar results were obtained using a migration chance of 5% throughout, see Fig. A.16.

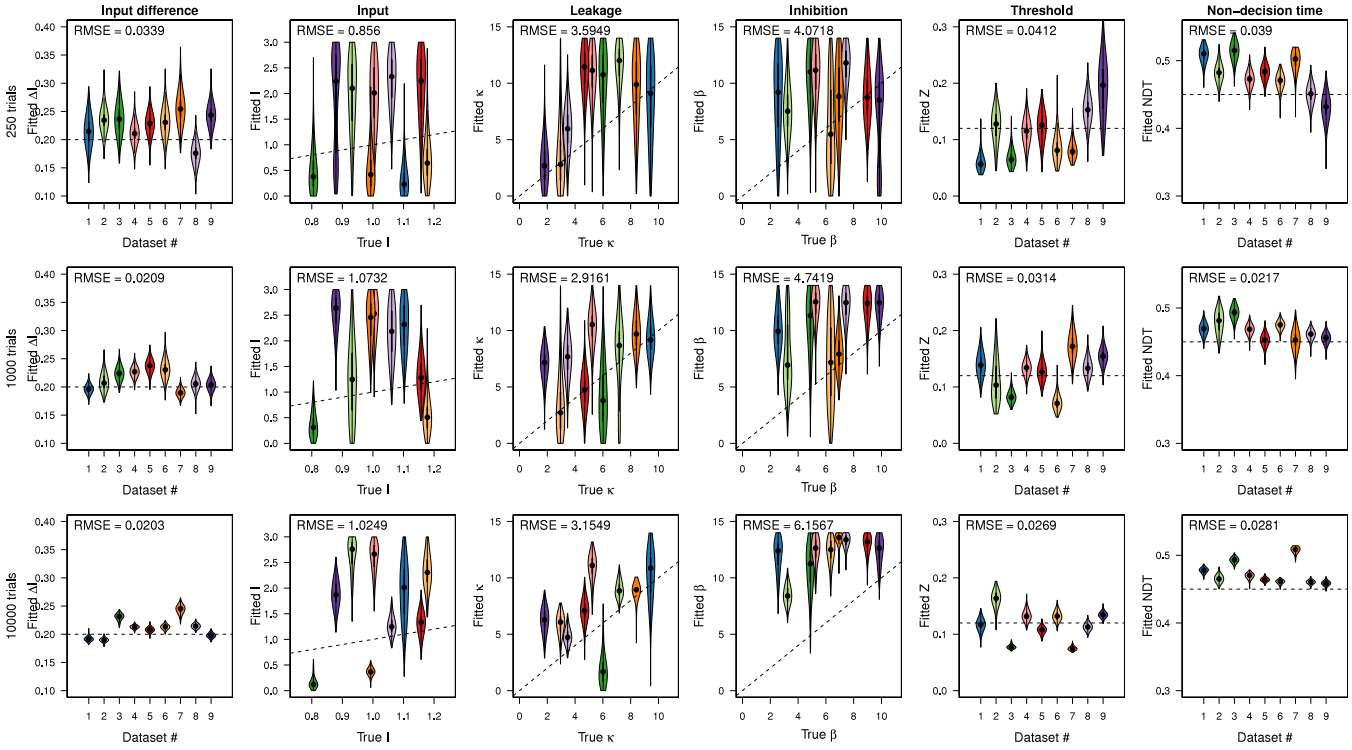


**Fig. 6.** Posterior distributions of parameter estimates obtained by using a DE-MCMC sampler on nine 250-trial (top), 1000-trial (middle), and 10,000-trial (bottom) data sets. Distributions are color coded by data set. All data sets had identical true input difference, threshold, and non-decision time parameter values, and the position on the *x*-axis in these panels is arbitrarily determined. For the input, leakage, and inhibition panels, the position on the *x*-axis is determined by the true, data-generating parameter value. The dashed line indicates the true parameter values. The mean of the root mean square errors is shown for every parameter. The pattern of results is identical to the results previously obtained using maximum likelihood: the input difference, threshold, and non-decision time parameters can be estimated reasonably accurately, but the estimates of the input, leakage, and inhibition parameters are off.

current study, we do not directly test this two-stage procedure, but instead, aim to discover whether, in the best case scenario that such a procedure works, it is possible to obtain accurate estimates of the crucial parameters in the LCA: input, leakage, and inhibition.

Using DE-MCMC with the input difference, threshold, and non-decision time parameters fixed to their true values, the amount of uncertainty in the posterior estimates in case of the 250 and 1000-trial data sets is too high to be of any practical use. The 10,000-trial data sets, however, have much less uncertainty in the obtained posteriors, and the means of the posteriors are close to the true parameter values.

### 3.1.1. Comparison of the MLE and DE-MCMC results

It appears that the DE-MCMC method slightly outperforms MLE in recovering the crucial parameters of the LCA. To study this, we applied a bootstrapping procedure to quantify whether the parameter estimates obtained using DE-MCMC are better than the estimates obtained using MLE. In this procedure, we estimated 20,000 RMSE-values by repeatedly sampling fitted data sets from both the DE-MCMC and MLE procedure, and calculating the RMSE of the sample. The number of samples was equivalent to the number of recoveries for each data set size (i.e., 59, 47, and 37 for data set sizes of 250 trials, 1000 trials, and 10,000 trials in the MLE procedure, 9 for all data set sizes in the DE-MCMC procedure). Then, we calculated the difference between the estimated RMSEs as obtained by the DE-MCMC and MLE procedure for every parameter and data set size. The probability that the RMSE is lower (i.e., more accurate recovery) using DE-MCMC than using MLE is equal to the proportion of differences in estimates that is smaller than 0. The results of this bootstrapping procedure are summarized in Table 2.

The results show that for input, even though only 9 data sets were estimated using DE-MCMC, the probability that the accuracy of the estimates using DE-MCMC is 1.00 irrespective of data set size. Similarly, the probability that the estimates of leakage are more accurate using DE-MCMC than using MLE is very high (0.84–1.00 depending on data set size). The same holds for the inhibition parameter (0.72 probability that DE-MCMC estimates are more accurate with a data set size of 250 trials, 1.00 with larger data set sizes). From these results, it can be concluded that estimates of these three crucial parameters obtained using DE-MCMC are better than the estimates obtained using MLE (see also Appendices A.3 and A.4 for similar comparisons).

**Table 2**
Probabilities, for every parameter and data set size, that RMSE obtained using DE-MCMC are lower (i.e., higher recovery accuracy) than RMSE obtained using MLE.

| Parameter | Data set size | Probability of #decrease in RMSE |
|---|---|---|
| Input | 250 | 1.00 |
| | 1,000 | 1.00 |
| | 10,000 | 1.00 |
| Leakage | 250 | 1.00 |
| | 1,000 | 0.86[a] |
| | 10,000 | 1.00 |
| Inhibition | 250 | 0.72 |
| | 1,000 | 1.00 |
| | 10,000 | 1.00 |

[a] This relatively low value, compared to the probabilities of a decrease in RMSE using DE-MCMC compared to MLE with 250 and 10,000 trials, is presumably caused by the low accuracy of the estimation of the leakage parameter of the yellow data set with 1000 trials using DE-MCMC.

## 4. Contrasts between conditions

Researchers in the cognitive neurosciences are often not primarily interested in the absolute parameter values. Instead, most experiments try to find the effect of a certain manipulation

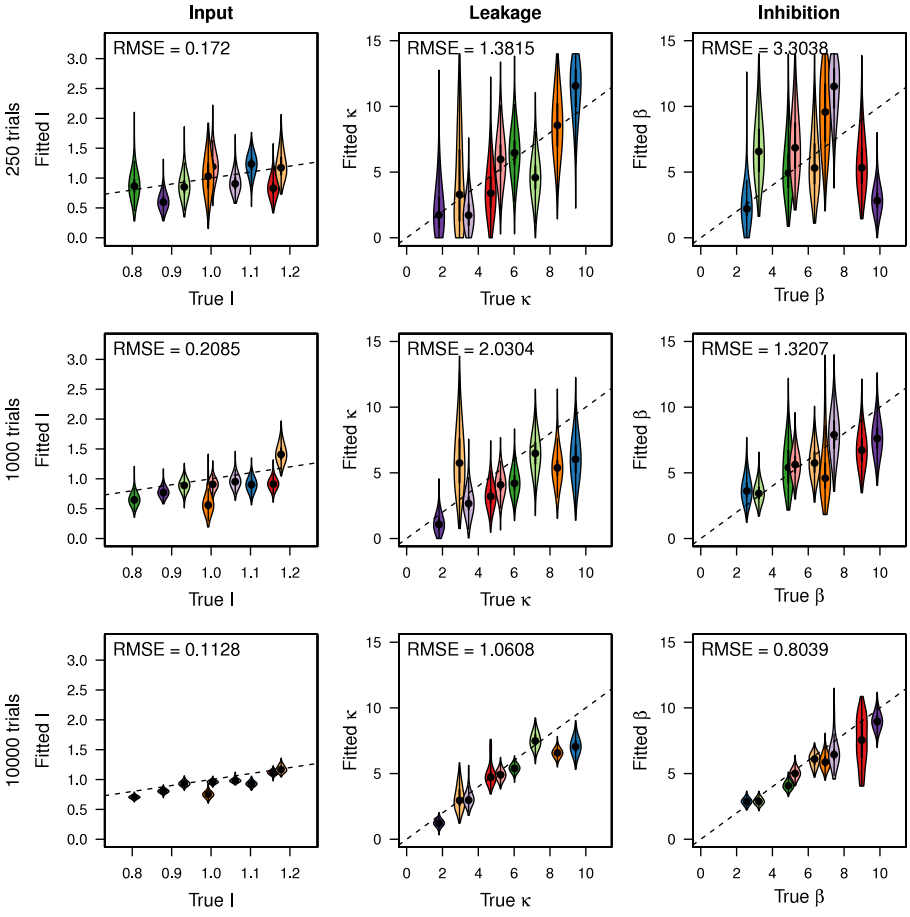


**Fig. 7.** Posterior distributions of parameter estimates obtained by using a DE-MCMC sampler on the nine 250-trial (top), 1000-trial (middle) and 10,000-trial (bottom) data sets. In each panel, the mean of the root mean squared errors is shown. Technical details of these panels are the same as columns 2, 3 and 4 in Fig. 6. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between conditions. For example, in speed–accuracy trade-off experiments, research aims to discover whether emphasizing the importance of being fast versus being accurate influences the response caution (i.e., threshold) or the efficiency of perceptual processing (i.e., drift rate) (Mulder et al., 2013; Rae, Heathcote, Donkin, Averell, & Brown, 2014).

Researchers interested in the LCA might design experiments aimed at finding an effect on inhibition or leakage. Although the absolute values of these parameters are very hard to retrieve, it might be possible to make claims about the difference in a parameter between conditions. To see whether this is possible, a new simulation study was performed to investigate whether a difference in only the inhibition parameter can be detected. The procedure described below can also be used to test for more complex effects, such as interactions.

First, three out of the nine data sets fitted in the previous section were randomly selected. For each of these three 'baseline' data sets, we created four 'variation' data sets by selectively increasing the inhibition parameter by 0.25, 0.5, 1.0, and 2.0. Effectively, this amounts to having three data sets with five conditions that only differ in the underlying inhibition parameter. Then, we estimated the input, leakage, and inhibition parameters, using the DE-MCMC procedure as described above, with the input difference, threshold, and non-decision time parameters fixed to their true values. The width of the priors was slightly increased to decrease ceiling effects (uniform priors with an upper bound of 5, 20, and 20, for the input, leakage, and inhibition parameters). The results of one of the three data sets are shown in Fig. 8, the other two are similar and shown in Supplementary Figs. A.13 and A.14.

The panels show the posterior parameter distributions of the baseline data set and the four variation data sets with a 0.25, 0.5, 1.0 and 2.0 increase in inhibition. Ideally, the posteriors of input and leakage of all data sets would be more or less identical, with only the inhibition posteriors increasing. Unfortunately, this does not seem to be the case. Although there is a general upward trend in the estimates of inhibition in the 10,000 trial case, these posteriors sometimes overestimate the increase in inhibition and underestimate the input and leakage parameters.

To explore what conclusions a researcher might have drawn based on these posteriors, we plotted the differences between posteriors of the variation data sets and the baseline set for the 10,000-trial case (bottom row in Fig. 8). We also calculated the estimated probabilities of each parameter increasing or decreasing as compared to baseline. These results are shown in Fig. 9.

The plot shows the difference distributions between the variation and baseline posteriors for the input (left), leakage (middle), and inhibition (right) parameters, for each of the four real differences in inhibition (0.25, 0.5, 1.0, and 2.0, top to bottom rows). The shaded area is the estimated probability $p$ that a parameter is higher in the variation set than in the baseline set. Thus, ideally, for all the variation sets, the shaded areas of the inhibition parameter should be large, and about 0.50 (probability) for the other parameters.

The results show that the estimated probability that the inhibition parameter is higher in the variation set than in the baseline set is indeed above a probability of 0.50 for all variation sets. However, in case of a true 0.25 increase in inhibition, the probability of detecting a decrease in the input and leakage

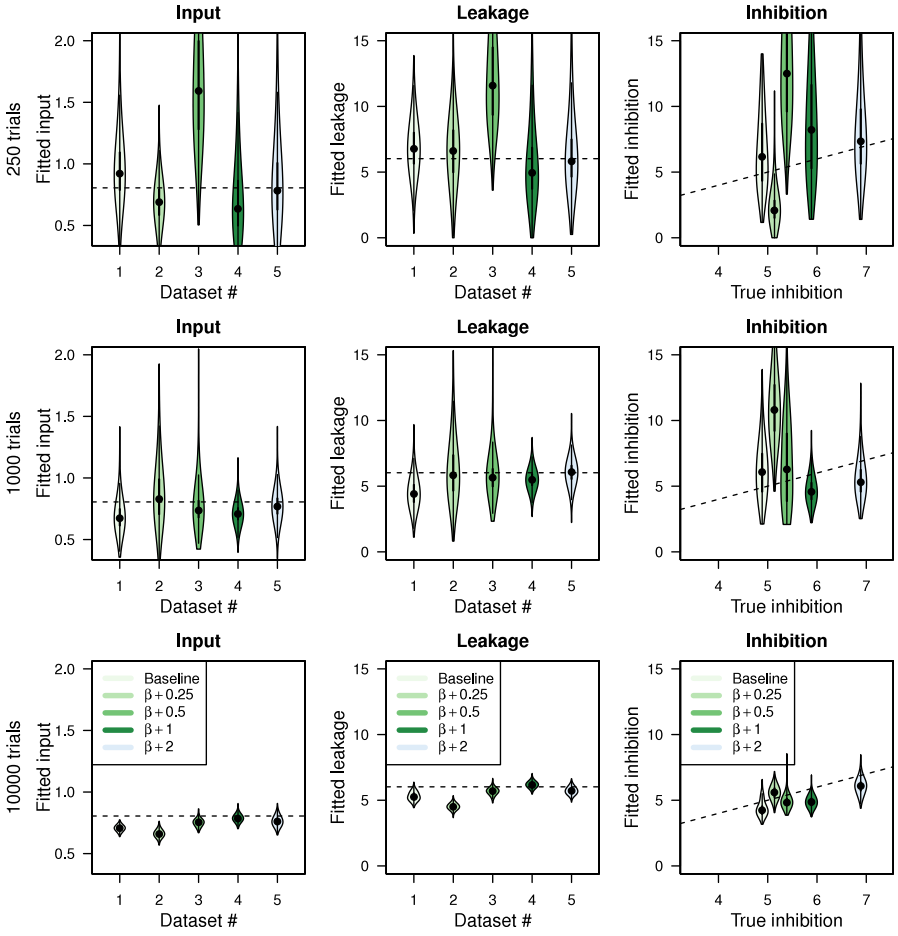


**Fig. 8.** Posterior distributions of the input, leakage, and inhibition parameters of five data sets, of which the data-generating parameter values differ only in the degree of inhibition. The left and middle panels show the recovery of the Input and Leakage parameters for each of the five different inhibition differences; The right panels show the ability to recover the inhibition parameters themselves. Dashed lines indicate the true parameter value.

parameters is very high. This also occurred in all four variation sets of the data set shown in Supplementary Fig. A.13. Reversely, in the variation sets with 0.5, 1.0, and 2.0 increase in inhibition, there is a very high probability of (falsely) detecting an increase in input and leakage (minimum probability of this occurring is 0.82—higher than the probability of detecting an increase in inhibition). This further occurred in two of the four variation sets of the data set shown in Supplementary Fig. A.14.

In summary, in 5 of the 12 contrasts estimated, one would detect an increase in inhibition, but also falsely conclude that there was a decrease in input and leakage. In 5 other cases, one would detect an increase in inhibition, but also falsely detect an increase in input and leakage. The remaining two cases (two variation sets of Supplementary Fig. A.14) are ambiguous; conclusions would highly depend on the threshold in probability that is used to base a conclusion on. In both cases, the correct pattern (an increase in inhibition, no difference in input and leakage) is not clear in the estimates.

On the basis of these results, we conclude that, even in the best-case scenario, in which a researcher is able to obtain 10,000 trials per participant per condition, and is able to find highly accurate estimates of the input difference, threshold, and non-decision time parameters and thereby apply a two-stage fitting procedure, an inhibition contrast that exists in data cannot be reliably detected.

## 5. Correlational structure of the parameters of the LCA

The results so far have indicated that many fitting routines produce inaccurate estimates of the input, leakage, and inhibition parameters (see also the methods discussed in the supplemental materials). The best estimates were obtained using the Bayesian fitting procedure with a data set consisting of 10,000 trials, but even there, there is some uncertainty. In this section we discuss possible reasons for the difficulty to reliably recover parameters for LCA. For this, the posterior distributions from the Bayesian fitting procedure were used to discover the correlations between the three hard-to-estimate parameters. The correlation matrix of the 10,000-trial data set (shown in orange in Figs. 6 and 7) is shown in Fig. 10.

Although there are clear pairwise correlations between the three parameters (including a very strong correlation between input and leakage), this does not necessarily cause fitting problems. For example, both the DDM and LBA also include strong correlations between the boundary separation and drift rate (Ratcliff & Tuerlinckx, 2002; Turner et al., 2013). Yet, it is possible to retrieve parameters in both DDM and LBA. So why is this a problem for the LCA?

Note first that, between data sets, the correlation between input and leakage is consistent and strong; but the correlations between input-inhibition and leakage-inhibition vary (Fig. A.15 displays the correlations of another data set). This, in combination with the fitting results that are suggestive of multivariate trade-offs, suggests that pairwise correlations are perhaps not the best way to visualize the relationships among the three parameters. To get a better impression, the obtained parameter estimates were plotted against each other in Fig. 11, together with the corresponding regression planes.

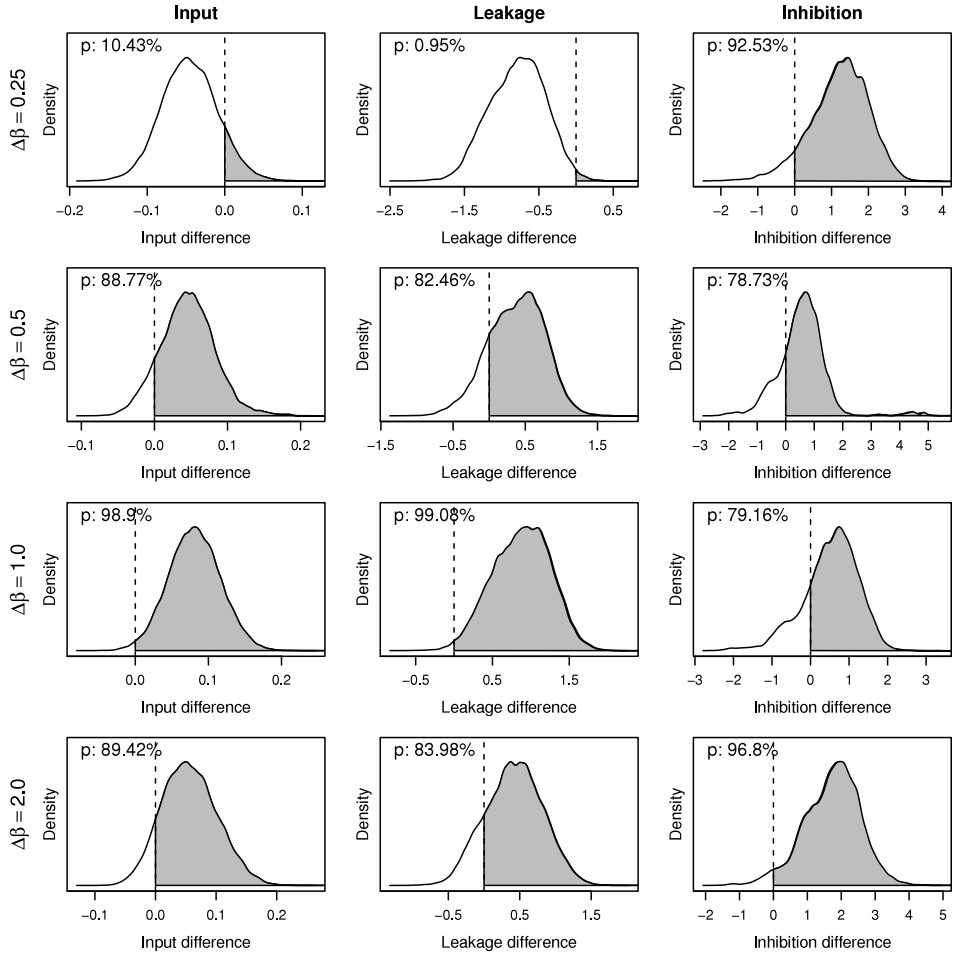


**Fig. 9.** Differences in posterior distributions for the input (left), leakage (middle), and inhibition (right) parameters. Differences are calculated between the variation and baseline data sets. The inhibition parameter in the variation data sets was increased with 0.25, 0.5, 1.0, and 2.0 (top to bottom row) compared to the baseline data set; the other parameters were identical. Shaded area shows the probability $p$ that the parameter is increased in the variation set compared to the baseline set. Ideally, only the probability that inhibition is higher in the variation set than in the baseline set should be high—the other probabilities should lie at around 0.50. The results show that the probability of an increased inhibition is indeed larger than 0.50 for all variation sets; but the probability that the other parameters are increased is also high.

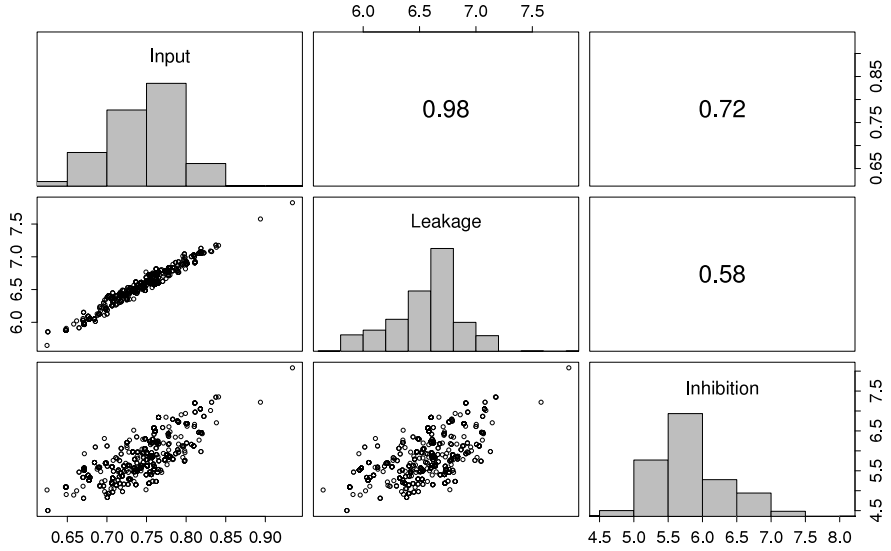


**Fig. 10.** Correlations between the input, leakage, and inhibition parameter estimates of the data set depicted in orange in Fig. 7. Each point is a parameter estimate from the posterior parameter distributions, as obtained using DE-MCMC sampling. Diagonal panels show histograms of the obtained estimates (same data as the orange violin plots in Fig. 7). There is a strong positive correlation between input and leakage, and somewhat weaker positive correlations between input and inhibition, and leakage and inhibition are clear.

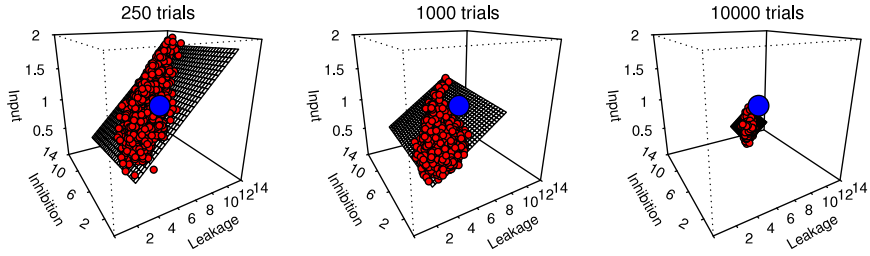


**Fig. 11.** Multivariate correlations between estimates of the input, leakage, and inhibition parameters. Each plot shows 2000 estimates (red dots), randomly sampled from the posterior parameter distributions obtained using DE-MCMC sampling on a 250 trial (left), 1000 trial (middle) and 10,000 trial (right) data set. The parameter estimates used correspond to the orange posteriors in Fig. 7. The blue circle indicates the true underlying set of parameters. Although increasing the amount of trials in a data set decreases the variance of estimates, they do not necessarily converge on the true parameter values.

An important observation is that, in all data set sizes, there is a plane of possible parameter estimates that produce highly similar likelihood estimates. Secondly, data set size influences the size of the surface: with more trials, the surface area decreases; with a decreasing surface area, the estimates' variance (and generally, but not always, accuracy) increases. Thirdly, unless data set size is massive, knowing one true parameter value still leaves a range of possible values of the other two parameters.

In sum, these figures show very large multivariate trade-offs between the input, leakage, and inhibition parameters. There are large parameter spaces that produce practically identical reaction time distributions. Future research should focus on whether it is the strength of the correlations, or the multivariate nature of the correlations, that makes it impossible for the optimization algorithms to disentangle these parameters.

## 6. Discussion

The present study investigated whether the parameters of the leaky, competing accumulator model could be reliably recovered. Using maximum likelihood estimation as well as Bayesian fitting procedures with various data set sizes, it was shown that only a DE-MCMC sampler with a very large data set resulted in reasonably accurate parameter estimates.

Three major causes contributed to the general fitting difficulties. Firstly, there is noise in goodness-of-fit measures, caused by the simulation-based nature of the fitting procedures, most notably the PDA-method. Increasing the number of simulations can decrease the amount of noise, but this comes at a severe practical cost of computation time. One way to overcome this problem (not presently used) would be to use GPU-based computation (e.g., Turner & Sederberg, 2014; Turner et al., 2016). Secondly, the number of trials in most data sets is relatively low. As a result, these data sets contain too little information about the underlying parameters. Thirdly, there are large trade-offs between the parameters of input, leakage, and inhibition. A different way to think about this is that the effects of these three different parameters can be mimicked by other parameters; the unique effects of these parameters are very subtle. Only with massive amounts of data, these unique effects can be detected.

The current study is fully simulation-based, and differs from settings with real data in a number of ways that may be detrimental for estimating parameters in real data. Although the 10,000-trial data set was recovered reasonably accurately using DE-MCMC, the performance will decrease when using more realistic data sets, for two reasons. First, models are always simplifications of the true underlying decision-mechanism. Therefore, the reaction time distributions generated by the true underlying decision-mechanism can only be approximated by the LCA, but are not identical. Second, the current data sets were simulated by the actual LCA and contained no contaminants. Contaminants (e.g., due to mind wandering) will have a significant negative effect on the performance of the likelihood-based methods (Ratcliff & Tuerlinckx, 2002). The obtained posteriors of contaminated data sets will either be wider (more uncertainty) or simply off the true parameter values.

There are ways to impose more constraints on the parameters. For example, one might have strong theoretical motivations for constraining parameters between conditions. A second approach would be to use the actual visual evidence stream to constrain the input parameter (e.g., Purcell et al., 2010; Teodorescu, Moran, & Usher, 2016; Teodorescu & Usher, 2013; Tsetsos et al., 2011). A third example is the use of neural data to constrain parameters, as was recently done using the DDM (Turner et al., 2015).

Fourth, in a variety of contexts, hierarchical Bayesian models have been shown to recover true parameter values more accurately than non-hierarchical models (e.g., Ahn, Krawitz, Kim, Busmeyer, & Brown, 2011; Rouder & Lu, 2005; Shiffrin, Lee, Kim, & Wagenmakers, 2008). Here, the basic approach is to estimate both individual and group-level effects via parameters simultaneously. Hierarchical models constrain the parameters corresponding to individual effects by assuming specific relationships between these parameters across subjects. In this way, parameter inference uses the data from all participants at the same time, instead of focusing on one individual's data at a time. Hence, it is possible that one would not require thousands of trials per subject as we have suggested above, but instead could rely on a few hundred trials per subject, with a handful of subjects. Although hierarchical models can be difficult to fit to data in the context of simulation-based inference as in the LCA model, the Gibbs ABC algorithm (Turner & Van Zandt, 2014) was developed to allow for scalability and has been applied to a few popular models in psychology (Turner et al., 2013; Turner & Sederberg, 2014; Turner & Van Zandt, 2014).

In the current study we particularly focused on a three-alternative choice task, rather than a more common binary choice paradigm. Three alternatives allow for disentangling the leakage and inhibition parameters because the presence of the third alternative provides more modeling constraint on the data. A potential drawback of this approach is that the parameters of a three-choice task may be inherently more difficult to estimate, given the more complex structure of the data. However, based on additional recovery studies with a three-alternative LBA model (see Supplemental Materials Appendix A.6) we believe this is not the case.

With smaller data set sizes, recovery performance of the LCA without such additional constraints is very poor, at least in the parameter region covered in the present study. Indeed, it is plausible that some parameter sets outside the currently used ranges produce reaction time distributions that are realistic, and that these parameter values can be accurately recovered. However, the fact that the currently used parameter ranges do not allow recovery is problematic for two reasons. First, these parameter ranges are capable of producing realistic reaction time distributions, and hence, these should be recoverable. Second, values of $\kappa$ and $\beta$ of 10 can be

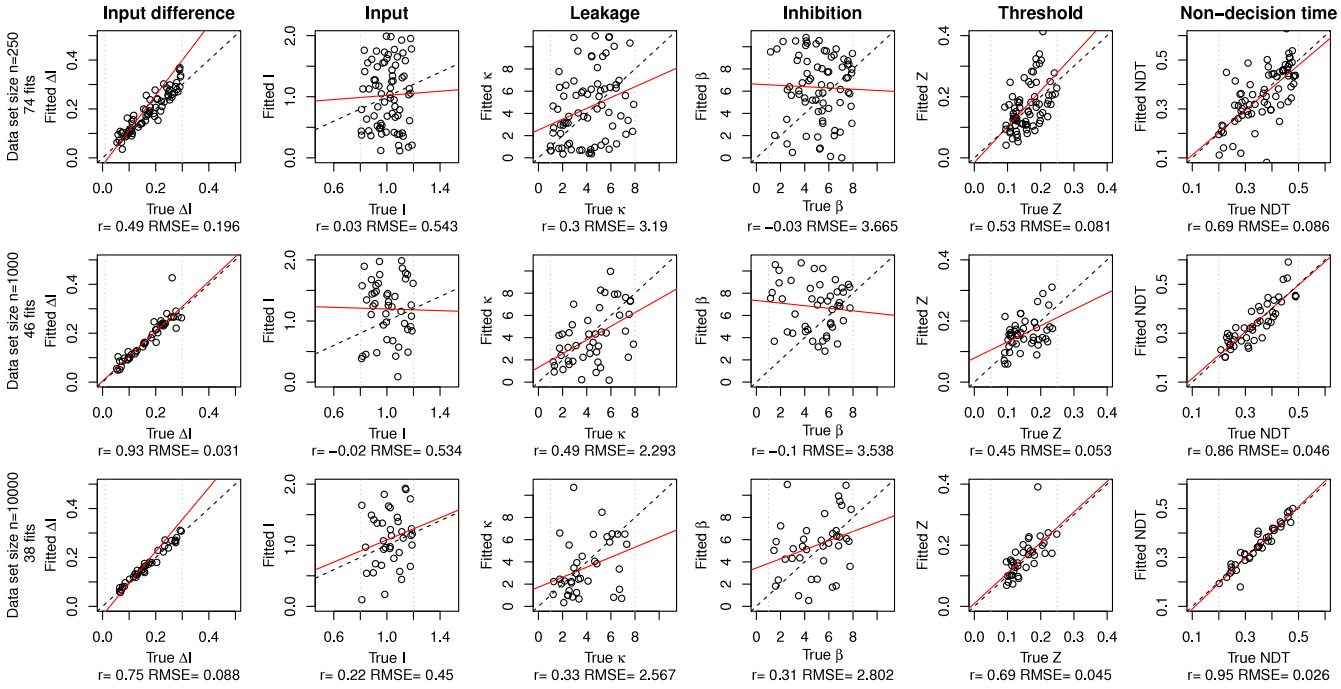


**Fig. A.1.** Parameter recovery results using quantile-based difference minimizing with 5 quantiles per distribution used for binning. The pattern of results is identical to the previously found pattern with maximum likelihood estimation. Input difference, threshold, and non-decision time (with larger data set sizes) can be recovered accurately, but the input, leakage and inhibition parameter estimates are off. Each point's position on the $x$-axis represents the parameter value used to generate the data set, the position on the $y$-axis the parameter estimate obtained by fitting. The vertical dashed lines correspond to the ranges of parameters used to generate data. The red line is the actual correlation between the data-generating parameter values and the fitted estimates. The correlation coefficients (Pearson's $r$) and RMSEs are shown underneath each panel.

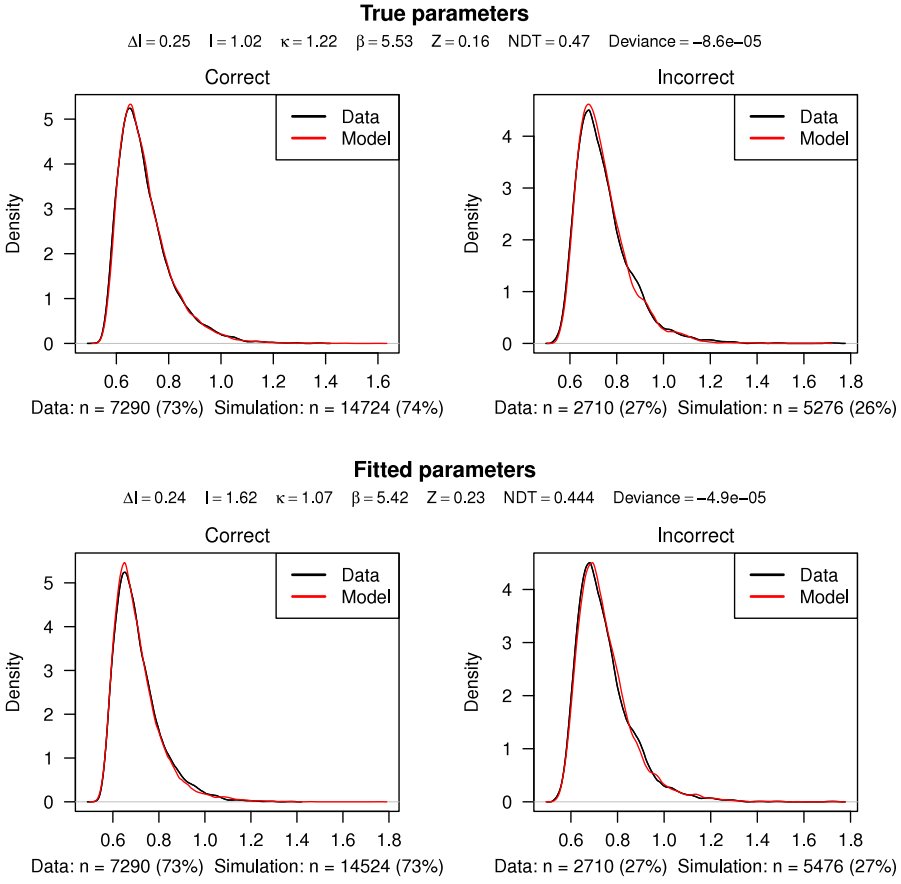


**Fig. A.2.** Comparison of the fit of the true parameter values $\theta$ (top row) and the fit of the best-fitting estimate of the parameters $\theta^*$ on a randomly generated 10,000-trial data set. Fits were obtained using quantile-based difference-minimizing. Density of the data is indicated by the black lines, the red lines indicate the density of a proposal data set generated using $\theta$ (top) or $\theta^*$ (bottom). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

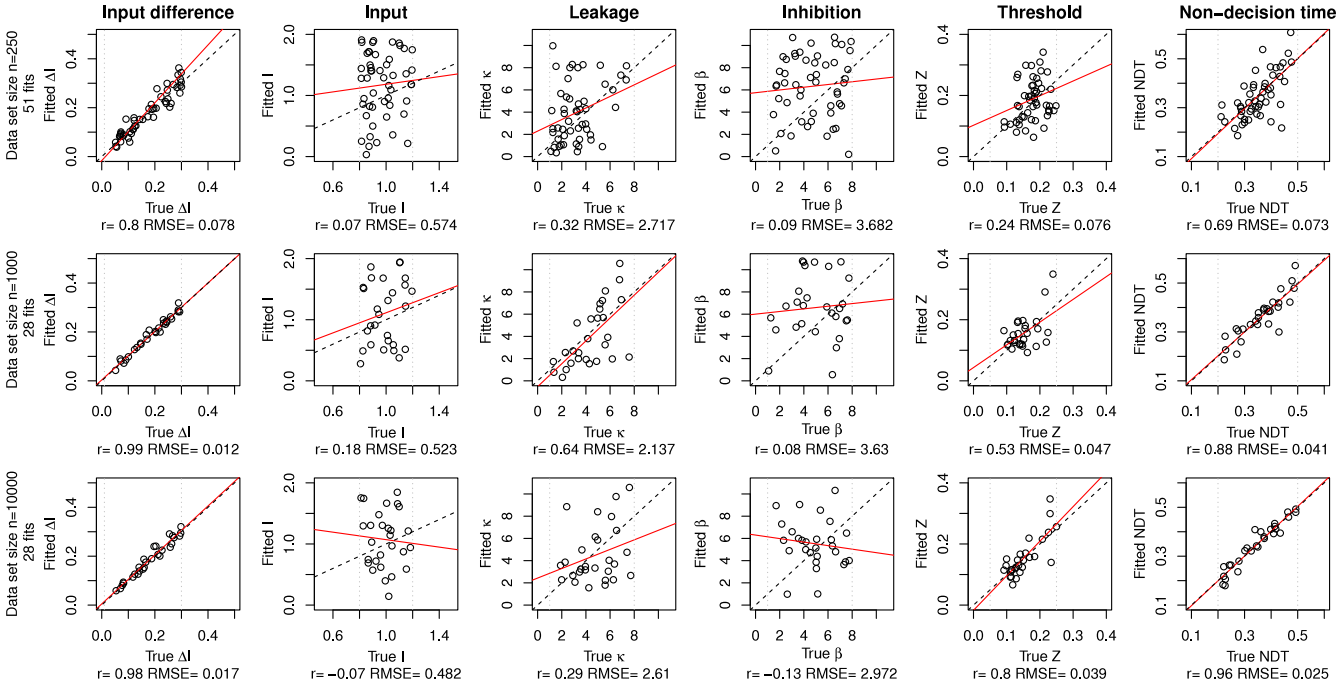


**Fig. A.3.** Parameter recovery using difference minimizing with ten quantiles per distribution. The pattern of results is identical to the previous results. Each point's position on the $x$-axis represents the parameter value used to generate the data set, the position on the $y$-axis the parameter estimate obtained by fitting. The vertical dashed lines correspond to the ranges of parameters used to generate data. The red line is the actual correlation between the data-generating parameter values and the fitted estimates. The correlation coefficients (Pearson's $r$) and RMSEs are shown underneath each panel.

(and have been, Bogacz et al., 2006) argued to be 'high'. Note that a $\kappa$ of 10 corresponds to a half-life of about 70 ms[4]—meaning that an accumulator loses half of its activation in 70 ms only due to leakage. A value of $\kappa$ of 15 corresponds to a half-life of approximately 45 ms. Before increasing $\kappa$ beyond such values, we might want to reconsider the idea behind leakage. On a neural scale, a half-life of 45–70 ms is a long time period, and this amount of leakage can be considered small. However, on a behavioral/cognitive scale, 45–70 ms can be argued to be very short, as this would imply that an accumulator loses 0.90 of its accumulated evidence when no evidence is shown for 210 ms (ignoring additional reduction of activation due to inhibition effects). Such degrees of leakage would be highly energy inefficient.

To answer the question of what amounts to reasonable leakage, we may gain some insight by answering the related question: What is an accumulator? It is improbable that an accumulator, as proposed by SSMs, is neurologically underpinned by a single neuron. Yet, single cell recordings have shown that neurons act like evidence accumulators, and that neurons can account for the reaction time distributions observed (e.g., Purcell et al., 2010). Subsequent simulation studies have shown that ensembles of neurons can act highly similar to individual neurons, under reasonable assumptions about correlations between neurons/accumulators (Zandbelt, Purcell, Palmeri, Logan, & Schall, 2014).

The study by Zandbelt et al. (2014) did not include a leakage-like mechanism in the individual neurons. A follow-up study might provide insights into the importance of leakage in a group of neurons, as compared to individual neurons. In other words, if the total evidence is represented by the activity of an ensemble of neurons, how does the decay of each individual neuron influence the level of activity of the group over time? Although such an approach will not help us solve the fitting problems involved with the LCA, we might gain insight into how problematic the absence of a leakage parameter in other SSMs (like the DDM) actually is.

In conclusion, without additional constraints in the fitting procedures, using DE-MCMC sampling and the PDA method together with very large data sets results in reasonably accurate parameter estimates, that are better than the estimates using any of the other methods used in this study. Although we cannot draw definitive conclusions about the identifiability of the LCA with smaller data set sizes, we hope at least to have shown that fitting the LCA is far from trivial. Researchers interested in fitting the LCA will either have to use massive data set sizes, or constrain parameter estimation in a way not currently performed and then extensively validate the fitting routine used.

## Appendix A. Supplementary model fits

In addition to the maximum likelihood and MCMC approaches reported in the main text, we attempted two additional methods for fitting the LCA model to data. We report these here to help other researchers in making the right choice of fitting routine. Our first additional attempt focusses on minimizing the difference between a set of observed and predicted quantile of the RT data (Smith & Ratcliff, 2015). Our second additional attempt used MLE, but explored the consequences of multiple repeated fitting routines with differences start points.

### A.1. Difference minimizing

A way to circumvent the problem of not having a likelihood function is to abandon likelihoods. One could take a 'leap of faith' and use a summary statistic, and use simulation studies to assess whether the parameter estimates are accurate. A widely used fitting procedure (e.g., Ratcliff & Smith, 2004; Smith & Ratcliff, 2015) without likelihoods is quantile-based difference minimizing. The general idea is that, given two reaction time distributions (i.e., a data set and a proposal data set), we can maximize

[4] $dx = [-\kappa x]\frac{dt}{\tau}$ so $\frac{dx}{dt} = [-\kappa x]\frac{1}{\tau}x$ which has an easy solution: $x_t = x_0 e^{-\frac{\kappa}{\tau}}$ so $t_{\frac{1}{2}} = \frac{\ln 2}{\frac{\kappa}{\tau}}$.

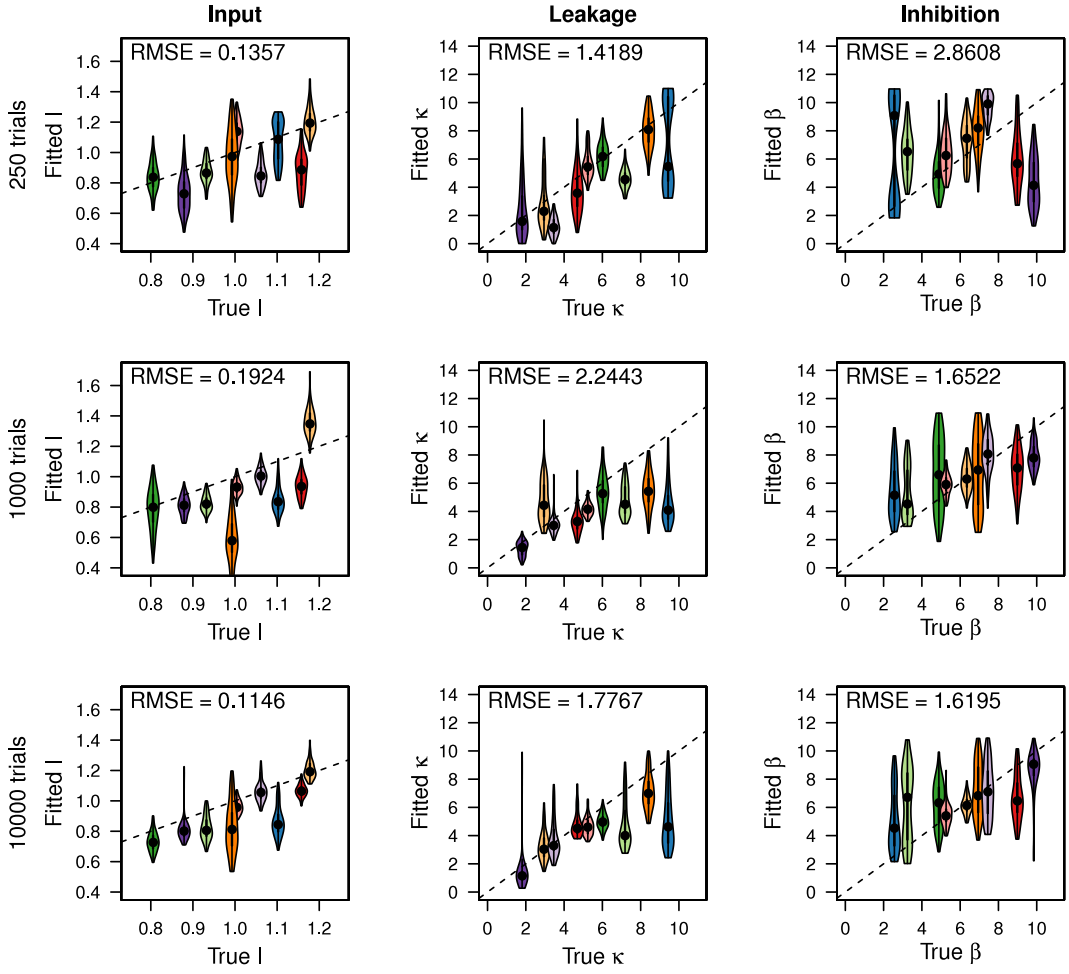


**Fig. A.4.** Violin plots of input, leakage, and inhibition estimates distributions obtained by repeatedly fitting nine data sets using maximum-likelihood estimation for three data set sizes: 250 (top), 1000 (middle), and 10,000 trials (bottom). The leakage and inhibition parameter estimates sometimes show a boundary effect, as the optimizer was constrained to generate values below 11. It is probable that the actual range of obtained estimates could be even larger. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the goodness-of-fit by minimizing the differences between the distributions. Important in the present study, the use of quantiles is, unlike maximum likelihood estimation, robust to outliers and contaminants in the data.

In this fitting procedure, quantiles are used to split the data into a number of bins. The proportion of data (probability mass) in each bin in the data is compared to the proportion of data in the same bin in the proposal data set. A difference function (or loss function) commonly used is a Pearson chi-square function:

$$\chi^2 = \sum_{j=1}^{m} n_j \frac{(p_j - \pi_j)^2}{\pi_j} \tag{15}$$

with $m$ quantiles $j$, $n$ observations, $p$ probability mass in the experimental data, and $\pi$ probability mass in the proposal data.

Although Chi-squared values are easily interpreted, this equation has a practical disadvantage due to the division by $\pi$: some proposal data sets contain no mass in certain bins, causing a division by 0. In the present study, an alternative function was used:

$$\Delta = \sum_{j=1}^{m} n_j (p_j - \pi_j)^2. \tag{16}$$

The absolute values of objective values $\chi^2$ in Eq. (15) and $\Delta$ in Eq. (16) will differ for a given data set and proposal parameter set. However, for each data set, the relative distances between $\chi^2$ values of several proposal parameter sets are similar to the relative differences in $\Delta$ for the same proposal parameter sets. Therefore, both objective functions will generally result in the same parameter values. Key choices in this procedure are the number of quantiles used, and how these quantiles are distributed over the correct and incorrect distributions. A frequently used method is the use of five quantiles (0.1, 0.3, 0.5, 0.7, and 0.9) for both the correct and incorrect distributions, but other choices are possible and often preferred (Anders et al., 2016).

### *A.1.1. Difference minimizing: simulation study*

This simulation study explored whether quantile-based difference minimizing can be used to accurately perform a parameter recovery of the LCA. The procedure is identical to the procedure of the MLE-study, but with a different measure of goodness-of-fit. Instead of estimating likelihoods, we estimate the difference criterion as specified by Eq. (16). Two ways of binning the data were explored.

In the first parameter recovery study binned the data using five quantiles at proportions (0.1, 0.3, 0.5, 0.7, and 0.9), following Ratcliff and Smith (2004), Smith and Ratcliff (2015). For the three data set sizes, the results are shown in Fig. A.1.

The results clearly resemble the results obtained with MLE, except for the estimate of the non-decision time for a 250-trial data set size. Using quantile-based difference minimizing, the difference between inputs $\Delta I$, threshold, and, with larger data

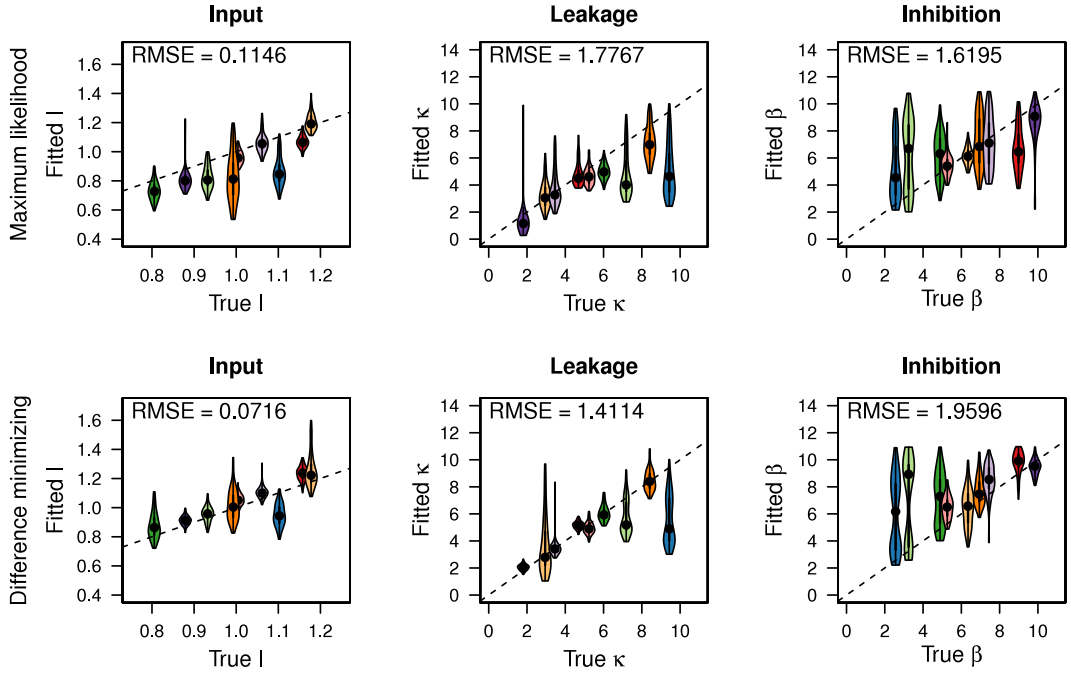


**Fig. A.5.** Violin plots of input, leakage, and inhibition estimates distribution obtained by repeatedly fitting nine 10,000-trial data sets using MLE (top) and difference minimizing (bottom). The fitting method has remarkably little influence on the results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

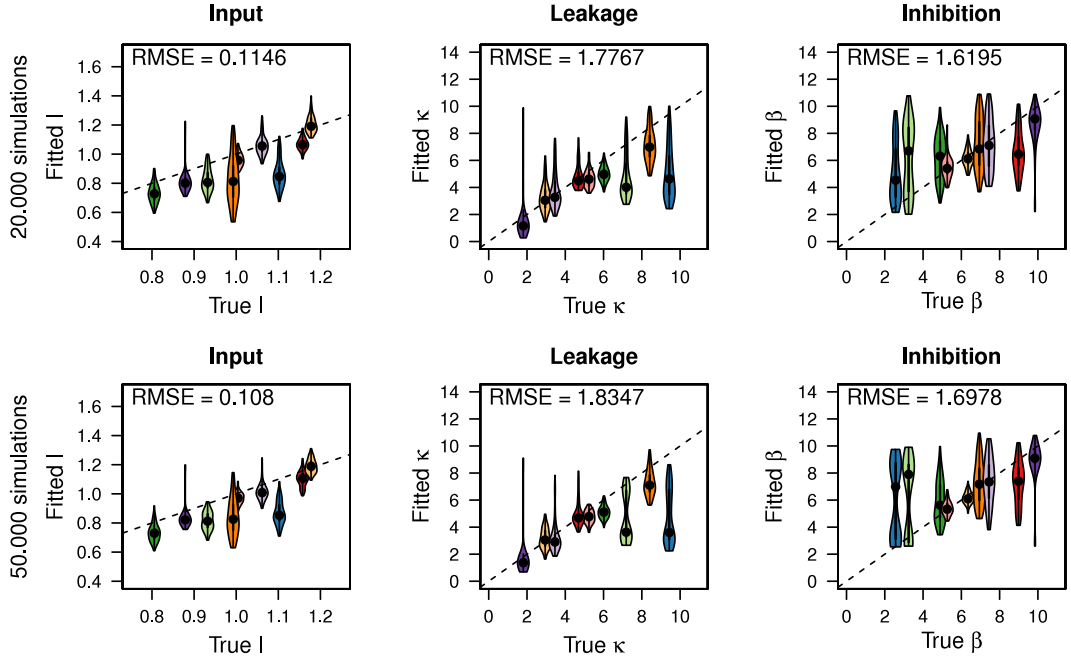


**Fig. A.6.** Violin plots of input, leakage, and inhibition estimates distribution obtained by repeatedly fitting nine 10,000-trial data sets using MLE with 20,000 simulations with $\frac{dt}{\tau} = \frac{3}{1000}$ (top) and 50,000 simulations with $\frac{dt}{\tau} = \frac{2}{1000}$ (bottom) per proposal data set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

set sizes, non-decision time can be estimated rather well, but the input, leakage, and inhibition parameter estimates are very far off the true parameters regardless of data set size.

A typical fit of an individual data set is shown in Fig. A.2. Although the fits are good, as shown by both visual inspection and the difference criterion used, the parameters estimated are not. But as mentioned before, the use of summary statistics is possibly invalid as the summary statistics used might be insufficient. Although it cannot be proven whether this is the cause of the poor results, we can try to reduce the problem by reducing the degree of compression in the summary. Instead of using five quantiles for each distribution, we can use ten: each bin is still a summary of a proportion of the data, but all bins together contain more information about the distribution. With more bins, more subtle details of a distribution can be detected, plausibly increasing the fit.

A new parameter recovery study was performed with difference minimizing using ten quantiles per distribution to bin the data. The procedure is, except for the binning method, identical to

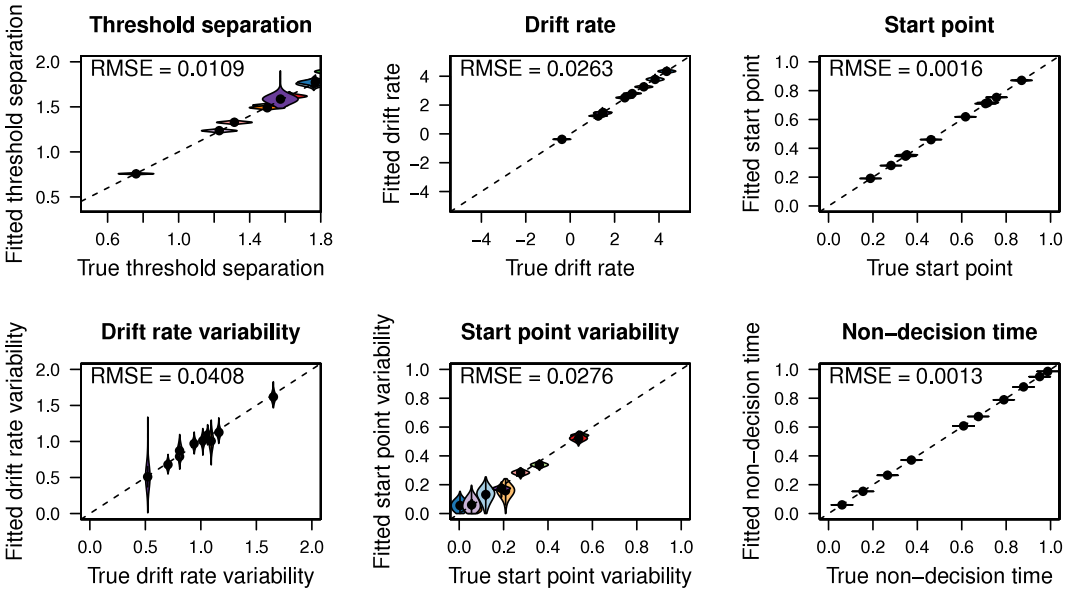


**Fig. A.7.** Posterior distributions of a parameter recovery of the DDM, using a DE-MCMC sampler. The root mean squared errors are shown in each panel. The threshold separation, drift rate, start point, and non-decision time parameters are recovered highly accurately without any uncertainty. The mean of the posteriors of the drift rate variability and start point variability parameters is also highly accurate, although there is some uncertainty surrounding these estimates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

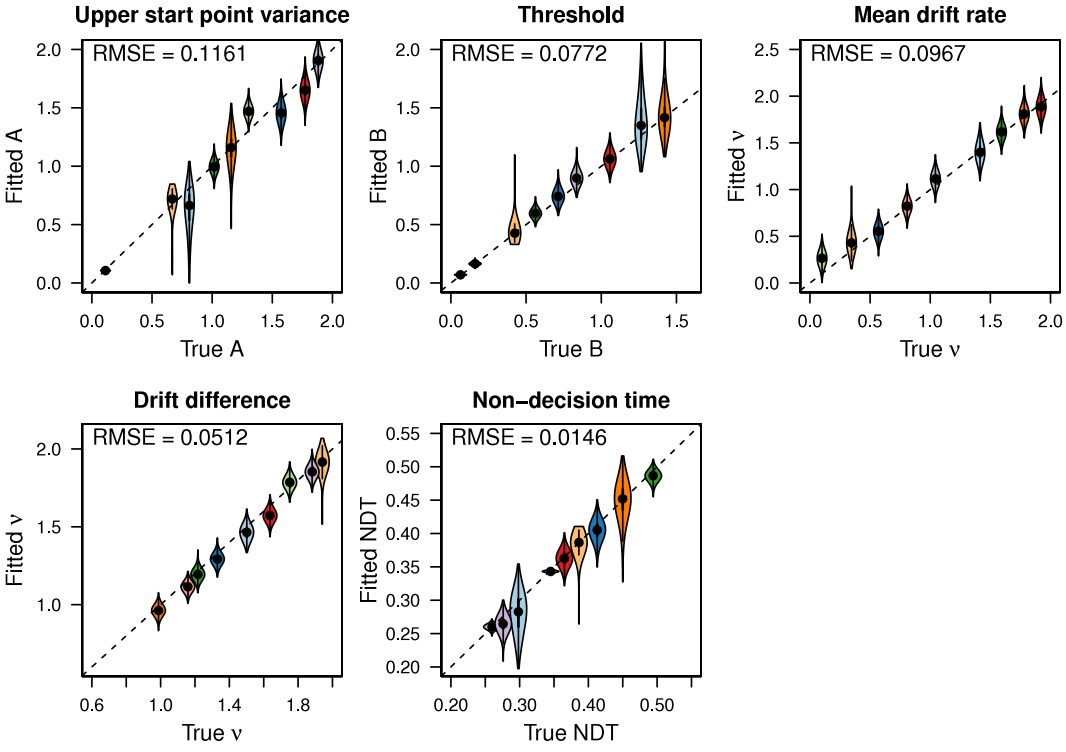


**Fig. A.8.** Posterior distributions of a parameter recovery of the LBA, using a DE-MCMC sampler. The root mean squared errors are shown in each panel. There is some uncertainty surrounding the estimates of start point variance and non-decision time, as well as of the threshold parameter in some data sets. However, the means of all estimates are highly accurate, with very low RMSE values.

the previously used procedure. The results are shown in Fig. A.3. Unfortunately, although the individual fits are good, the parameter estimates of input, leakage, and inhibition are still off. However, it is not easy to see whether they are better.

### *A.2. Repeatedly fitting and quantifying average goodness-of-fits*

Previous results showed that fitting a data set only once results in inaccurate estimates of input, leakage, and inhibition. While single parameter estimates were generally far off the true parameter values, it is possible that different data set sizes or different fitting methods do have an influence on the average estimates' accuracy. This cannot be deduced from previous results, as they do not show a measure of uncertainty in parameter estimates.

To obtain a measure of uncertainty, a single data set needs to be fitted repeatedly in order to obtain a distribution of parameter estimates. The distribution can be used to measure the 'average' deviance (i.e., how far off the is mean of all obtained parameter

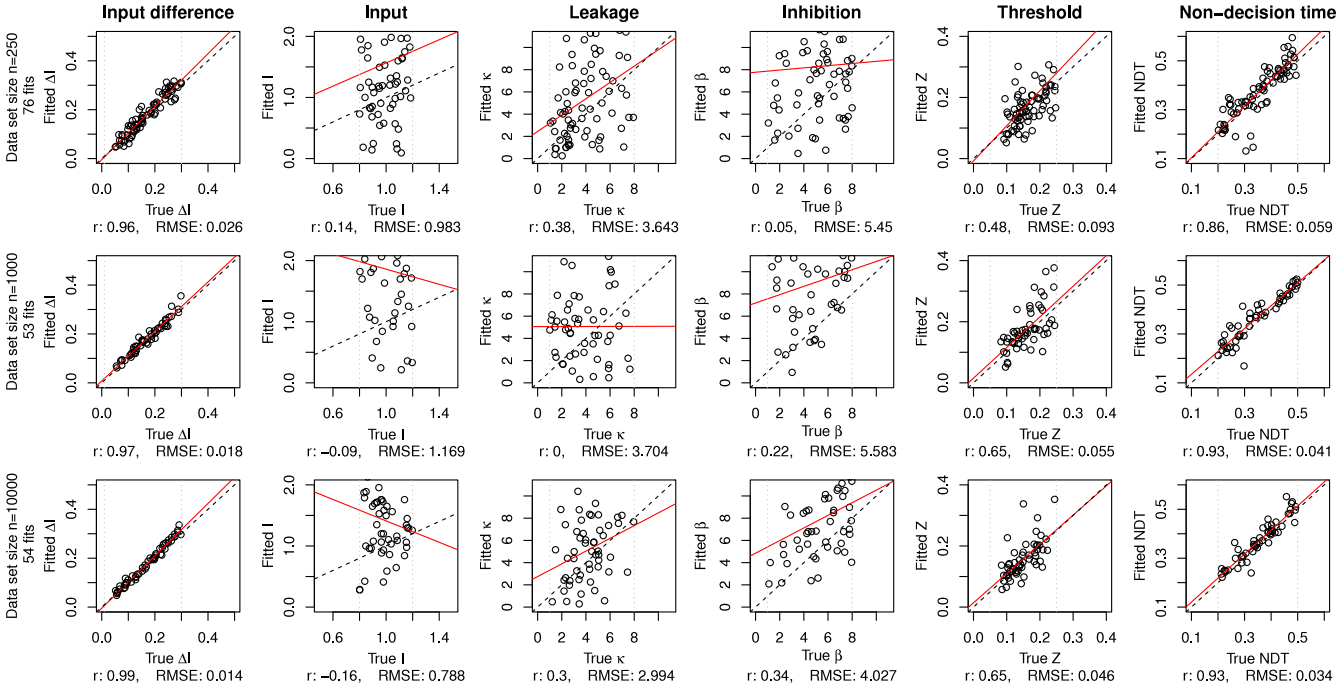


**Fig. A.9.** Results of a parameter recovery study with maximum likelihood estimation, in which the optimization algorithm stopped when it could not improve its best known parameter set for 100 iterations, with a maximum of 10,000 iterations. Each point's position on the $x$-axis represents the parameter value used to generate the data set, the position on the $y$-axis the parameter estimate obtained by fitting. The vertical dashed lines correspond to the ranges of parameters used to generate data. The red line is the actual correlation between the data-generating parameter values and the fitted estimates. The correlation coefficients (Pearson's $r$) and RMSEs are shown underneath each panel.

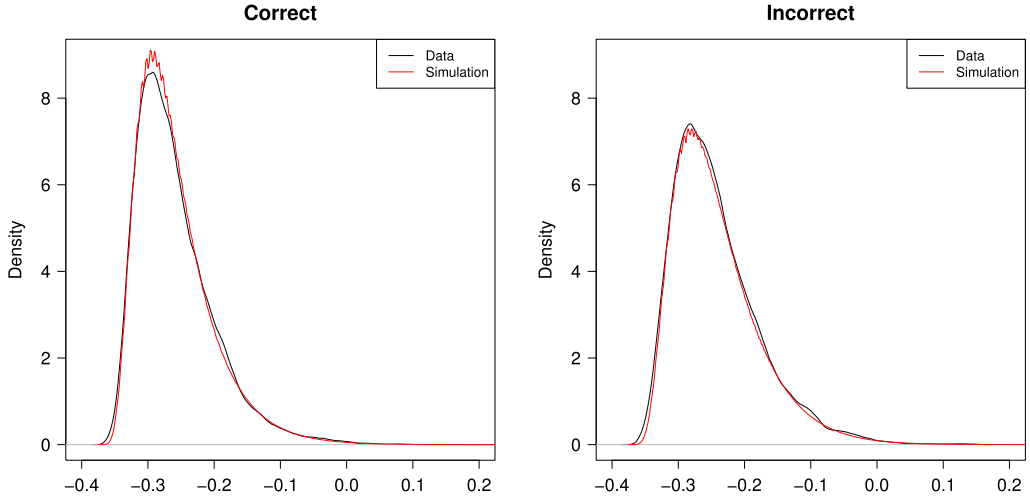


**Fig. A.10.** Density of reaction time distributions of a random data set $D$ (black) and a proposal data set (red), generated with the true parameter values underlying $D$, using $10^8$ trials. In Silverman's rule of thumb, the bandwidth is dependent on the number of observations. With this many simulations, the bandwidth becomes so small that the density is not smoothed anymore. Note that the $x$-axis (reaction time) was log-transformed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

estimates from the true value), and the variance within parameter estimates of a single data set.

The next procedure was designed to explore whether different data set sizes or fitting methods influenced the accuracy of obtained parameter estimates. Nine data sets of each 10,000 trials were created using pseudo-randomly generated parameters. Each parameter set was randomly generated, but it was manually checked whether the corresponding reaction time distribution was reasonable (i.e., not too slow/fast). Each data set was fitted repeatedly for 100 times. The obtained parameter estimate distributions were used to calculate variance and mean deviance from the true parameter value. To keep computation time manageable, only input, leakage, and inhibition were fitted; the other parameters were fixed to their true values. As these parameters can be accurately obtained using any of the previous methods, this should not be a problem when fitting real data.

### *A.2.1. Data size*

Data sets consisting of 250 trials contain less information about the underlying parameters than 1000 or 10,000-trial data sets. Therefore, it is expected that an increase in the amount of trials improves parameter estimation. To test whether this is the case, nine 250-trial and nine 1000-trial data sets were created by randomly sampling 250 and 1000 trials from the 10,000 trial data sets. This resulted in three data set sizes of the same parameter

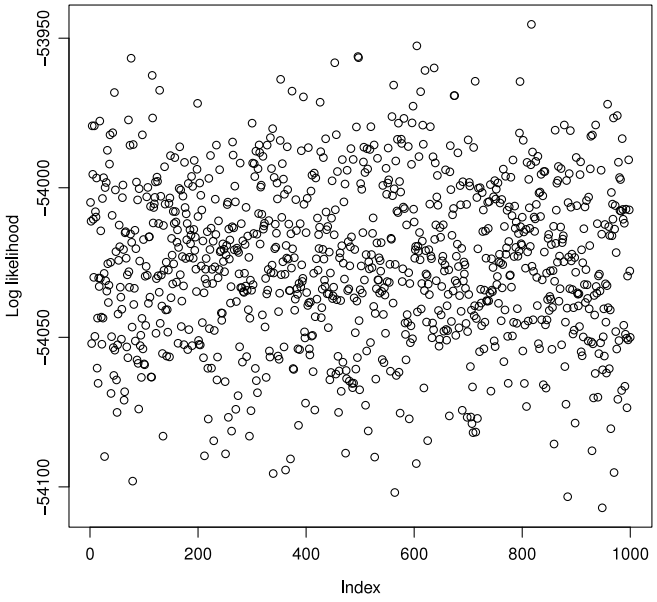


**Fig. A.11.** Scatterplot of 1000 calculated log-pseudo-likelihoods of $\theta$ on a randomly generated data set $D$. Log-pseudo-likelihoods were obtained using Eq. (12); the slowest 5% data points were binned. The figure shows no clusters of estimates, showing that binning the slowest 5% of the data points solves the zero-density problem.

sets. Each data set was repeatedly fitted 100 times using maximum likelihood estimation with PDA. Each proposal data set contained $J = 20{,}000$ simulated trials with a time step size of $\frac{dt}{\tau} = \frac{3}{1000}$.

The distributions of parameter estimates obtained by this procedure were plotted as violin plots in Fig. A.4. The mean variance of distributions and the mean deviance are shown in each panel. Violin plots illustrate the density of the obtained estimates. Note that, to improve visual interpretation, the width of the violins was adapted. In interpreting these violins, only the relative widths within one violin, and the relative widths between violins within the same panel are meaningful. The use of violins allows us, for example, to see whether or not posteriors are normally distributed around the 'true' mean. This, however, was not the case for any data set size. In fact, it seems that increasing data set size has remarkably little influence on the fitting estimates.

### *A.2.2. Maximum likelihood estimation vs quantile-based difference minimizing*

As shown, the PDA-method is noisy, so perhaps the inaccuracies can be decreased using a different method. To see whether this was the case, maximum-likelihood estimation with PDA was compared to quantile-based difference minimizing. For both methods, there were $J = 20{,}000$ simulations per proposal data set with a time step of $\frac{dt}{\tau} = \frac{3}{1000}$. The difference minimizing method used binned the data in 20 quantiles, as described previously. Only the 10,000-trial data sets were fitted. The results are shown in Fig. A.5. Remarkably, there is little difference between the two methods in the quality of obtained parameter estimates.

### *A.2.3. More simulations*

A different hypothesis is that the cause of the inaccuracies is not the method (MLE or difference minimizing), but what they have in common: they are both simulation based. If noise caused by simulations is the problem, it should be possible to increase the fitting performance by increasing the number of simulations. In the previous fitting procedures, $J = 20{,}000$ simulations with a time step of $\frac{dt}{\tau} = \frac{3}{1000}$ (corresponding to 3 ms) were used. This resolution is similar to or higher than previous research (e.g., Ratcliff & McKoon, 2008; Tsetsos et al., 2011; Turner et al., 2016; Usher & McClelland, 2001). To explore whether the simulation resolution is the underlying problem of the inaccuracies, the same nine data sets were again repeatedly fitted using maximum likelihood estimation. In the current procedure, simulation noise was reduced by increasing the number of simulations to $J = 50{,}000$ per proposal data set and the times step to $\frac{dt}{\tau} = \frac{2}{1000}$. The results are shown in Fig. A.6. Once again, there is hardly any difference.

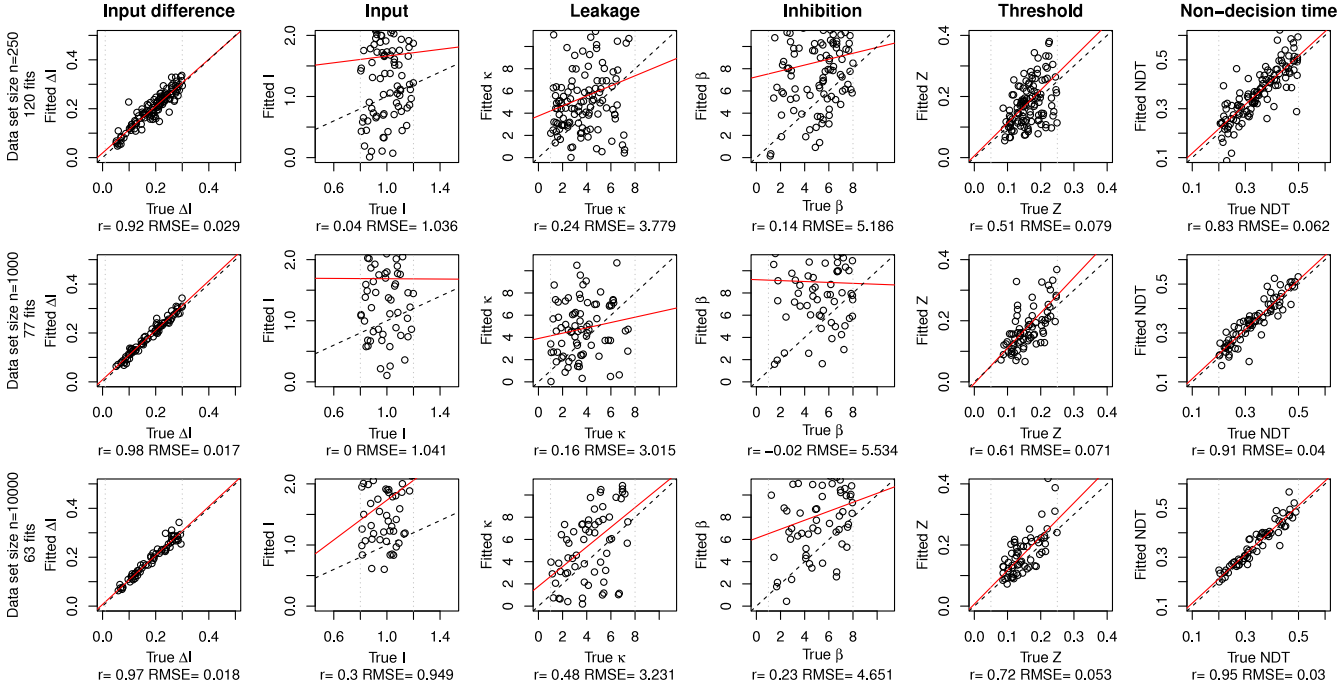


**Fig. A.12.** Results of a parameter recovery study with maximum likelihood estimation using Eq. (12) for the three data set sizes (top row $N = 250$, middle row $N = 1000$, bottom row $N = 10{,}000$). Each point's position on the $x$-axis represents the parameter value used to generate the data set, the position on the $y$-axis the parameter estimate obtained by fitting. The vertical dashed lines correspond to the ranges of parameters used to generate data. The red line is the actual correlation between the data-generating parameter values and the fitted estimates. The correlation coefficients (Pearson's $r$) and RMSEs are shown underneath each panel.

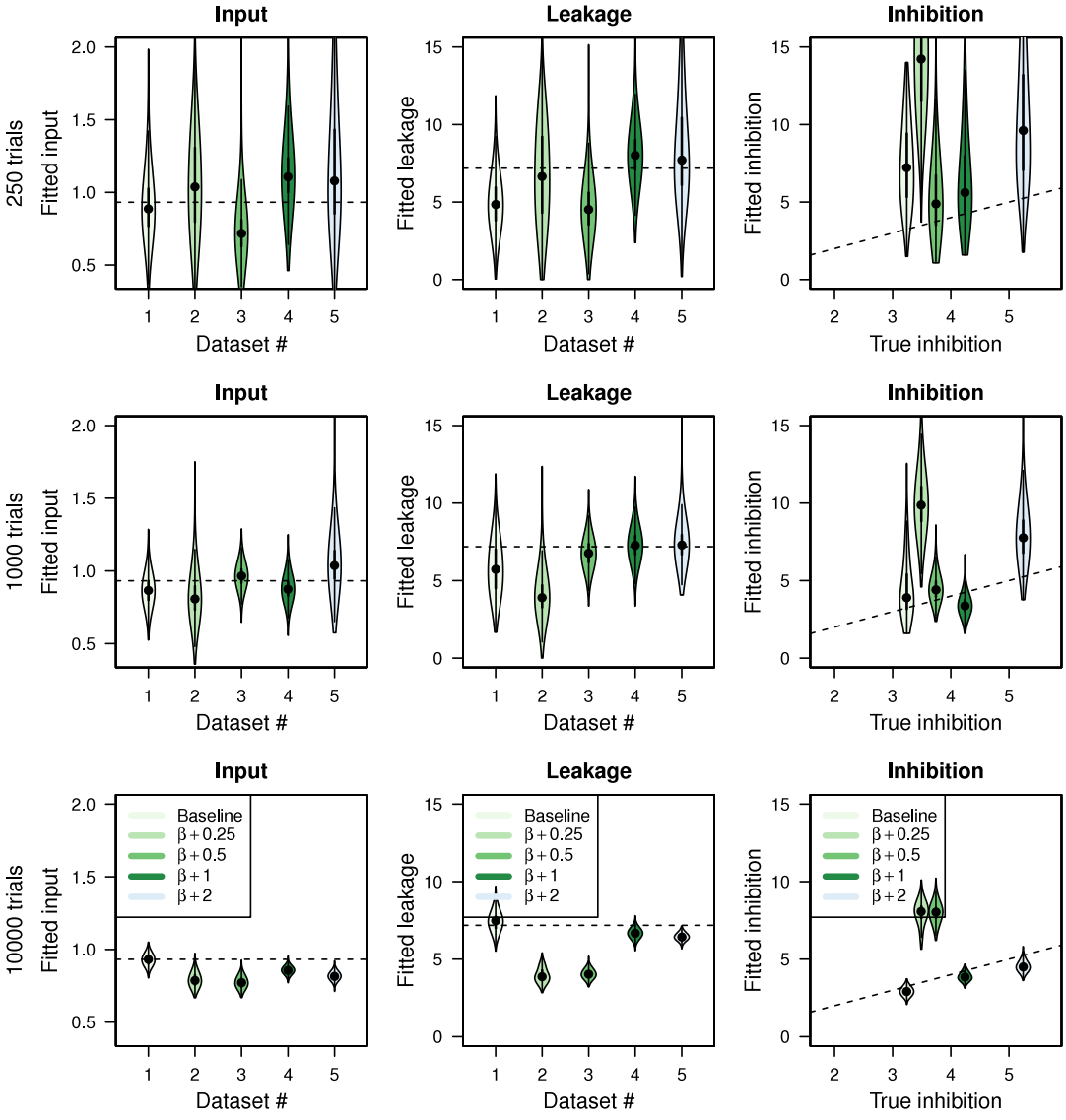


**Fig. A.13.** Same figure as 8 and A.14, with a different data set. Posterior distributions of the input, leakage, and inhibition parameters of five data sets, of which the true parameter values differ only in the degree of leakage. Ideally, a change in the true value of leakage causes only a change in the posterior distribution of the leakage parameter, not the other parameters. The results are similar to the results obtained by changing only the inhibition parameter.

### *A.3. Comparison of the MLE and DE-MCMC recoveries of all parameters*

In Sections 2.2 and 3 we recovered all LCA parameters using MLE and DE-MCMC. In this section, we make a quantitative comparison of the results of these two parameter estimation methods by using a bootstrapping procedure. This procedure is exactly the same procedure as we used in Section 3.1.1. We estimated 20,000 RMSE-values by repeatedly sampling fitted data sets from both the DE-MCMC and MLE procedure, and calculating the RMSE of the sample. The number of samples was equivalent to the number of recoveries for each data set size (i.e., 59, 47, and 37 for data set sizes of 250 trials, 1000 trials, and 10,000 trials in the MLE procedure, 9 for all data set sizes in the DE-MCMC procedure). Then, we calculated the difference between the estimated RMSEs as obtained by the DE-MCMC and MLE procedure for every parameter and data set size. The probability that the RMSE is lower (i.e., more accurate recovery) using DE-MCMC than using MLE is equal to the proportion of differences in estimates that is smaller than 0. The results of this bootstrapping procedure are summarized in Table A.1.

**Table A.1**
Probabilities, for every parameter and data set size, that RMSE obtained using DE-MCMC are lower (i.e., higher recovery accuracy) than RMSE obtained using MLE.

| Parameter | Data set size | Probability of #decrease in RMSE |
|---|---|---|
| Input difference | 250 | 0.15 |
| | 1,000 | 0.74 |
| | 10,000 | 0.18 |
| Input | 250 | 0.00 |
| | 1,000 | 0.00 |
| | 10,000 | 0.00 |
| Leakage | 250 | 0.22 |
| | 1,000 | 0.29 |
| | 10,000 | 0.08 |
| Inhibition | 250 | 0.28 |
| | 1,000 | 0.03 |
| | 10,000 | 0.00 |
| Threshold | 250 | 0.86 |
| | 1,000 | 0.95 |
| | 10,000 | 0.91 |
| Non-decision time | 250 | 0.00 |
| | 1,000 | 0.00 |
| | 10,000 | 0.00 |

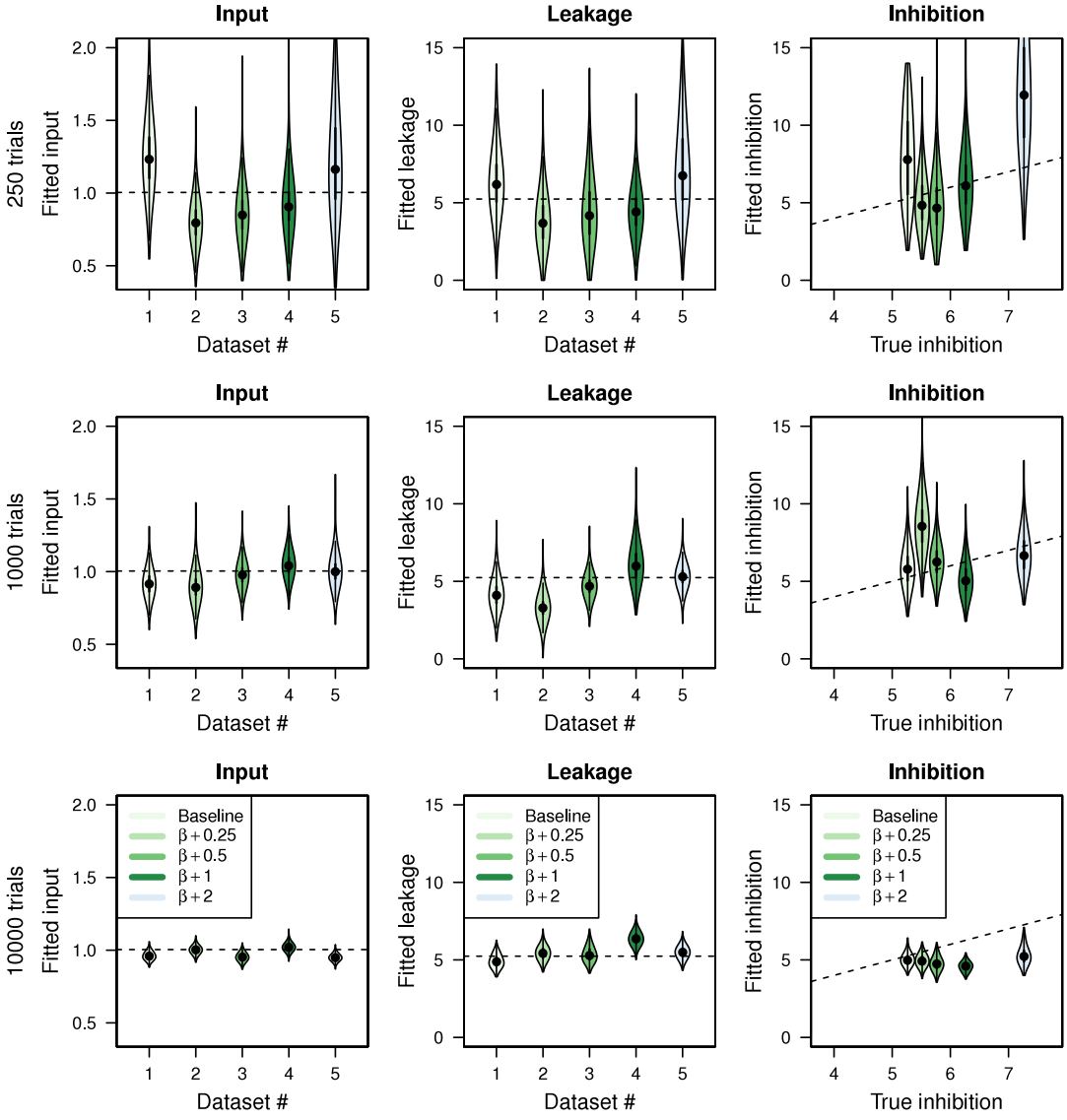


**Fig. A.14.** Same figure as Figs. 8 and A.13, with a different data set. Posterior distributions of the input, leakage, and inhibition parameters of five data sets, of which the true parameter values differ only in the degree of leakage. Ideally, a change in the true value of leakage causes only a change in the posterior distribution of the leakage parameter, not the other parameters. The results are similar to the results obtained by changing only the inhibition parameter.

The results differ per parameter. For the input different parameter, the probability that the recovery accuracy is higher (i.e., the RMSE is lower) with DE-MCMC compared to MLE is between 0.15 and 0.74, depending on data set size. The threshold parameter seems to be recovered better when using DE-MCMC (probabilities of a decreasing RMSE of 0.86–0.95), but the non-decision time parameter is estimated better using MLE.

The input, leakage, and inhibition parameters seem to be estimated better using MLE. However, it must be kept in mind that the recovery of these parameters using both methods is very poor. Therefore, it would be premature to conclude that MLE is a good method to recover these three crucial parameters. Given the very low recovery accuracy of these parameters using both methods, it can only be concluded that the performance of both these methods in this case is too poor to be of practical use.

### *A.4. Comparison of the MLE and DE-MCMC recoveries of only the input, leakage, and inhibition parameters*

In Section 3.1.1, we compared the recovery accuracy of the input, leakage, and inhibition parameters obtained by using MLE the recovery accuracy of these parameters obtained by using DE-MCMC. However, unlike in the MLE procedure, the input difference, threshold, and non-decision time parameters were fixed to their true values in the DE-MCMC procedure. For this reason, one might argue that the comparison used is unfair.

We therefore applied a similar bootstrapping procedure to compare the DE-MCMC results from Section 3 with the MLE results of Appendix A.2.1 (i.e., MLE with input difference, threshold, and non-decision time fixed to their true values). Again, we estimated 20,000 RMSE-values by repeatedly sampling fitted data sets from both the DE-MCMC and MLE procedure, and calculating the RMSE of the sample. The number of samples was equivalent to the number of recoveries for each data set size (i.e., 9 for all data set sizes). Then, we calculated the difference between the estimated RMSEs as obtained by the DE-MCMC and MLE procedure for every parameter and data set size. The probability that the RMSE is lower (i.e., more accurate recovery) using DE-MCMC than using MLE is equal to the proportion of differences in estimates that is smaller than 0. The results of this bootstrapping procedure are summarized in Table A.2.

For data sets with 250 trials, the two recovery methods perform similar. The results seem to indicate a slight preference for the

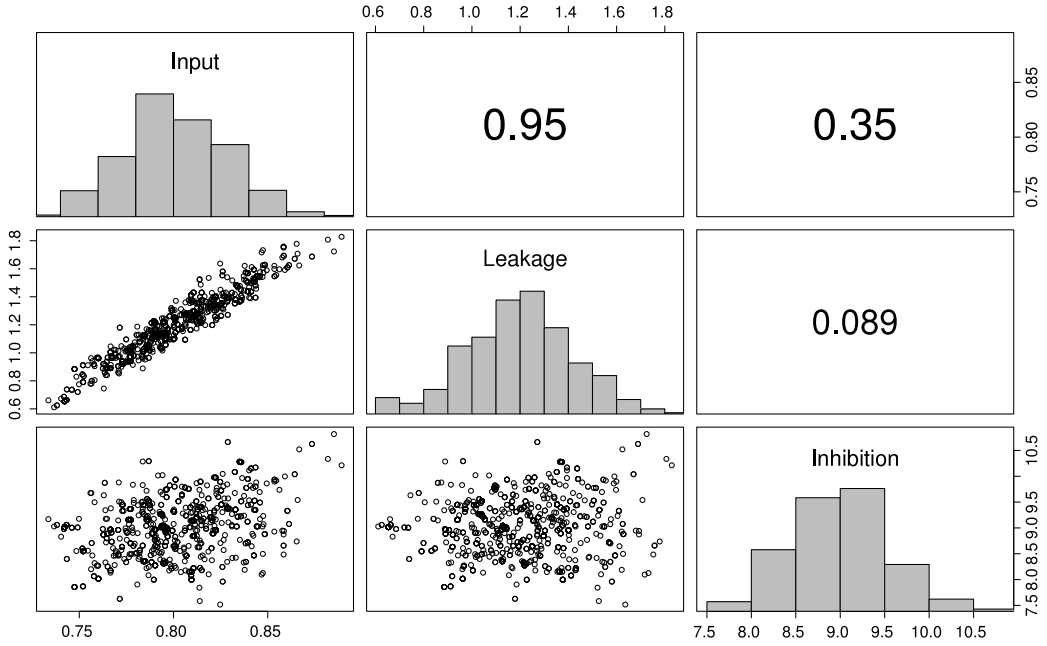


**Fig. A.15.** Bivariate correlations between parameter estimates of the 10,000-trial data set depicted in red in Figs. A.4–A.7. As compared to the bivariate correlations of the orange data set, the correlations between input and leakage, and leakage and inhibition are much less strong.

**Table A.2**
Probabilities, for every parameter and data set size, that RMSE obtained using DE-MCMC are lower (i.e., higher recovery accuracy) than RMSE obtained using MLE.

| Parameter | Data set size | Probability of #decrease in RMSE |
|---|---|---|
| Input | 250 | 0.24 |
| | 1,000 | 0.40 |
| | 10,000 | 0.54 |
| Leakage | 250 | 0.61 |
| | 1,000 | 0.59 |
| | 10,000 | 0.83 |
| Inhibition | 250 | 0.34 |
| | 1,000 | 0.83 |
| | 10,000 | 0.98 |

MLE procedure, but it must be kept in mind that the performance of both methods in estimating parameters using data sets of 250 trials is very poor. However, for data sets of 1000 trials, the DE-MCMC estimates of inhibition are better than the MLE estimates of inhibition. For data sets of 10,000 trials, the estimates of the leakage and inhibition parameters obtained using DE-MCMC are better than those obtained using MLE.

Note that, when interpreting these results, the RMSE metric does not take into account the uncertainty of an individual estimate, only the means. These uncertainties are for DE-MCMC much lower than for the MLE estimates. Furthermore, as noted in Appendix A.2.1, the distributions of MLE estimates show a boundary effect, and are probably more accurate than they would have been under experimental settings with wider optimizer constraints. For these reasons, we believe that the difference in performance between DE-MCMC and MLE is even larger than these results indicate.

### *A.5. Parameter recovery of the DDM*

How well can a parameter recovery be performed using the same method, but with the DDM instead of the LCA? Such a comparison can inform us about the to-be-expected posterior distributions of an identified model. The DE-MCMC procedure was used to perform a parameter recovery on the extended DDM (Eq. (1)). This model has 7 parameters: threshold separation ($a$, cf. threshold in LCA), drift rate ($v$, c.f. input in LCA), start point ($z$), between-trial drift rate variability ($sv$), between-trial start point variability ($sz$), non-decision time ($NDT$), and moment-to-moment noise ($c$). Moment-to-moment noise was held constant with variance $s = 0.1$ to avoid scaling problems; the other parameters were free to vary.

Nine parameter sets were pseudo-randomly generated by sampling from the following distributions:
$a \sim \mathcal{N}(1, 1)$ doubly truncated at $(0, 2)$
$v \sim \mathcal{N}(0, 2)$ doubly truncated at $(-5, 5)$
$z \sim \mathcal{U}(0, 1)$
$sv \sim \mathcal{N}(1, 1)$ doubly truncated at $(0, 2)$
$sz \sim \mathcal{U}(0, 1)$
$NDT \sim \mathcal{U}(0, 1)$.

Randomly generated parameter sets were rejected if they allowed the start point of a single trial to lie beyond the threshold (e.g., a high start point with high variability), or if the percentage of incorrect answers was too high (i.c., above 50%). Each of the accepted parameter sets was used to simulate a data set of 10,000 trials. The nine generated data sets were used to perform a parameter recovery.

The DE-MCMC sampler was implemented with 50 chains. Prior distributions were identical to the distributions used to generate parameters. 1000 burn-in iterations were followed by 6000 sampling iterations, producing posteriors of 300,000 samples. Migration chance was set to 25% during the burn-in. Stuck chains form less of a problem with the DDM due to the use of a (noiseless) likelihood function, and a migration chance after the burn-in was not found necessary. The random perturbation factor was set to 0.001, and the scaling factor $\gamma$ was randomly sampled from $\mathcal{U}(0.5, 1)$ for each proposal. The obtained posterior distributions are shown in Fig. A.7.

The threshold separation, drift rate, start point and non-decision time parameters were recovered with high accuracy and practically no uncertainty. The estimation of threshold and non-decision time is comparable to those estimates obtained when fitting LCA (see also van Ravenzwaaij & Oberauer, 2009). The mean estimates of drift rate variability and start point variability are close to the true values, but there is some uncertainty these estimates.

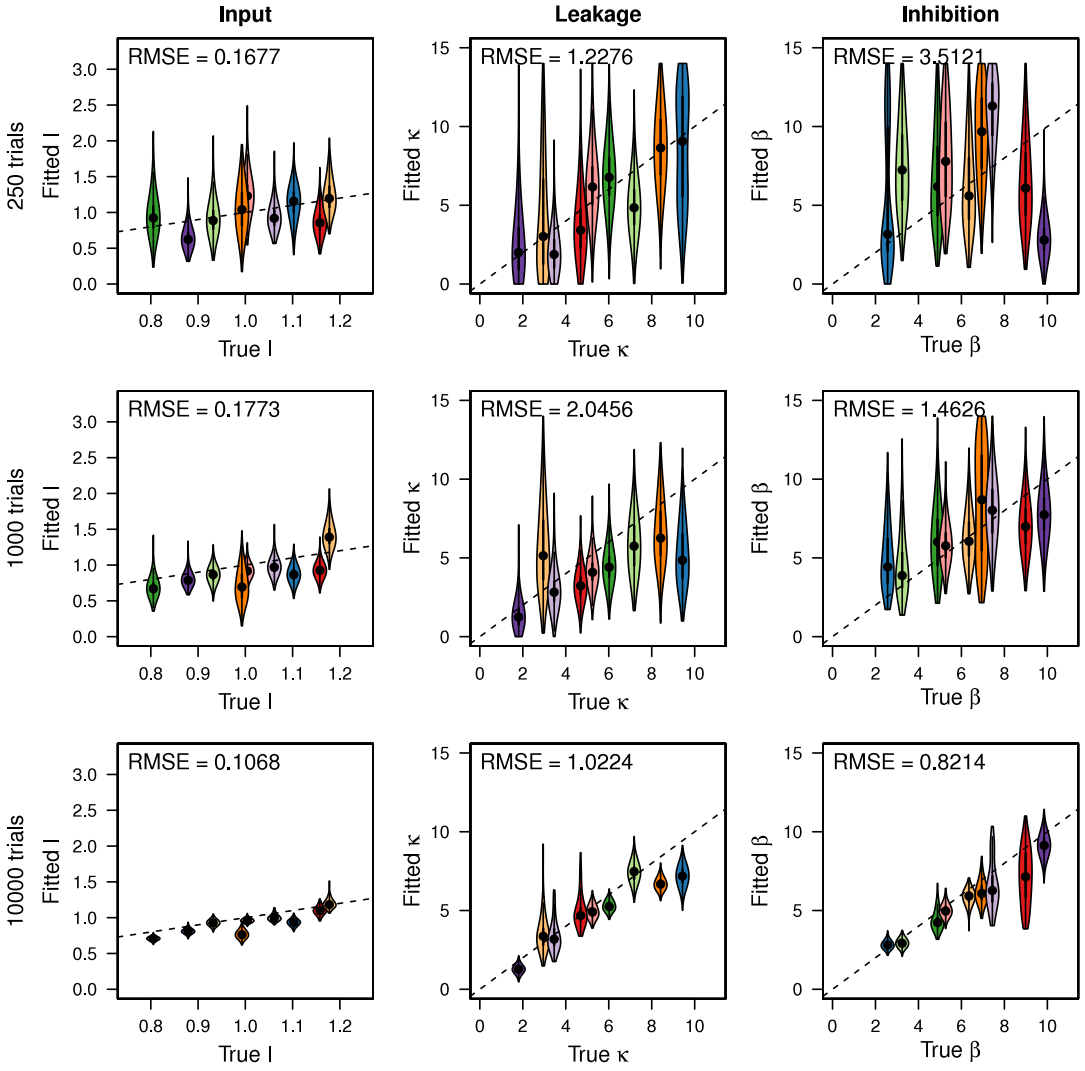


**Fig. A.16.** Posterior distributions of parameter estimates obtained by using a DE-MCMC sampler on the nine 250-trial (top), 1000-trial (middle) and 10,000-trial (bottom) datasets. These results were obtained by implementing the DE-MCMC sampler with a migration chance of 5% throughout instead of the 25% used to create Fig. 7. The results are highly similar.

### *A.6. Parameter recovery of a 3-choice LBA*

To exclude the possibility that recoverability is a general problem with 3-choice models, we used a DE-MCMC sampler to perform a parameter recovery for a 3-choice linear ballistic accumulator model (Brown, Steyvers, & Wagenmakers, 2009). The LBA we implemented has 6 parameters: the upper bound of the start point variance distribution $A$, threshold $B$, mean drift rate $v$, difference in drift rate between the correct and incorrect answers $\Delta v$, drift rate variance $sv$, and non-decision time $NDT$. Nine parameter sets were pseudo-randomly generated by sampling from the following distributions:

$A \sim \mathcal{U}(0, 2)$
$B \sim \mathcal{U}(0, 2)$
$v \sim \mathcal{U}(0.01, 2)$
$\delta v \sim \mathcal{U}(0, 2)$
$NDT \sim \mathcal{U}(0.2, 0.5)$.

Parameter $sv$ was fixed to 1 to prevent scaling problems. Randomly generated parameter sets were rejected if the amount of incorrect answers was above 50%, or if the range of reaction times was below 400 ms. Each of the nine parameter sets was used to simulate a data set of 10,000 trials. These nine data sets were used to perform a parameter recovery.

The DE-MCMC sampler was implemented with 15 chains. Prior distributions were identical to the distributions used to generate parameters. 400 burn-in iterations were followed by 5000 sampling iterations, producing posteriors of 75,000 samples. Migration chance was set to 5% during the burn-in. Stuck chains form less of a problem with the LBA due to the use of a (noiseless) likelihood function, and a migration chance after the burn-in was not found necessary. The random perturbation factor was set to 0.001, and the scaling factor $\gamma$ was randomly sampled from $\mathcal{U}(0.5, 1)$ for each proposal. The obtained posterior distributions are shown in Fig. A.8.

Although there is some uncertainty surrounding estimates of the start point variance and non-decision time parameters, as well as of the threshold parameter in some data sets, the mean estimates of all parameters for all data sets are highly accurate. It is possible that recovery might improve further (i.e., less uncertainty) with more chains or more iterations.

## Appendix B. Supplementary figures

This section contains figures that are mentioned in the main text but are considered supplementary to the main message of the paper.

## References


Abbott, L. (1991). Firing-rate models for neural populations. In O. Benhar, C. Bosio, P. Del Guidice, & E. Tabet (Eds.), *Neural networks: From biology to high-energy physics* (pp. 179–196).

Ahn, W.-Y., Krawitz, A., Kim, W., Busmeyer, J. R., & Brown, J. W. (2011). A model-based FMRI analysis with hierarchical Bayesian parameter estimation. *J Neurosci Psychol Econ*, *4*(2), 95–110.

Amit, D. J., Brunel, N., & Tsodyks, M. V. (1994). Correlations of cortical Hebbian reverberations: theory versus experiment. *Journal of Neuroscience*, *14*(11 Pt 1), 6435–6445.

Anders, R., Alario, F. X., & Van Maanen, L. (2016). The shifted Wald distribution for response time data analysis. *Psychological Methods*, *21*(3), 309–327.

Ardia, D., Arango, J. O., & Gomez, N. G. (2011a). Jump-diffusion calibration using differential evolution. *Wilmott Magazine*, *55*, 76–79. URL http://www.wilmott.com/.

Ardia, D., Boudt, K., Carl, P., Mullen, K. M., & Peterson, B. G. (2011b). Differential evolution with DEoptim: An application to non-convex portfolio optimization. *The R Journal*, *3*(1), 27–34.

Ardia, D., Mullen, K.M., Peterson, B.G., & Ulrich, J. 2015. DEoptim: Differential Evolution in R. Version 2.2-3. URL http://CRAN.R-project.org/package=DEoptim.

Boehm, U., Van Maanen, L., Forstmann, B. U., & Van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *Neuroimage*, *96*, 95–105.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–765.

Brooks, S. P., & Gelman, A. (1998). Generative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

Brown, E., & Holmes, P. (2001). Modeling a simple choice task: Stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics*, *1*, 159–191.

Brown, S., Steyvers, M., & Wagenmakers, E. J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, *53*(6), 453–462. URL http://dx.doi.org/10.1016/j.jmp.2009.09.002.

Buch-larsen, T., Nielsen, J., Guillen, M., & Bolance, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, *39*, 503–516.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459.

Churchland, A. K., & Ditterich, J. (2012). New advances in understanding decisions among multiple alternatives. *Current Opinion in Neurobiology*, *22*(6), 920–926.

Conn, A. R., Gould, N. I., & Toint, P. (1991). A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, *28*(2), 545–572.

Conn, A. R., Gould, N. I., & Toint, P. (1997). A globally convergent lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Mathematics of Computation*, *66*(217), 261–288.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 17538–17542.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2015). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Physiology*.,.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

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

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning* (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc..

Keuken, M. C., Van Maanen, L., Bogacz, R., Schäfer, A., Neumann, J., Turner, R., et al. (2015). The subthalamic nucleus during decision-making with multiple alternatives. *Human Brain Mapping*, *36*(10), 4041–4052.

Latimer, K., Yates, J., Meister, M., Huk, A. C., & Pillow, J. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, *349*(6244), 184–187.

Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception & Psychophysics*, *72*, 246–273.

Lo, C. C., & Wang, X. J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience*, *9*(7).

McMillen, T., & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, *50*(1).

Moran, R. (2016). Thou shalt identify! the identifiability of two high-threshold models in confidence-rating recognition (and super-recognition) paradigms. *Journal of Mathematical Psychology*, *73*, 1–11.

Mulder, M., Keuken, M., Van Maanen, L., Boekel, W., Forstmann, B. U., & Wagenmakers, E.-J. (2013). The speed and accuracy of perceptual decisions in a random-tone pitch task. *Attention, Perception & Psychophysics*, *75*, 1048–1058.

Mulder, M., Van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences—A model-based review. *Neuroscience*, 872–884.

Mullen, K., Ardia, D., Gil, D., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, *40*(6), 1–26. URL http://www.jstatsoft.org/v40/i06/.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

Noorbaloochi, S., Sharon, D., & McClelland, J. L. (2015). Payoff information biases a fast guess process in perceptual decision making under deadline pressure: Evidence from behavior, evoked potentials, and quantitative model comparison. *Journal of Neuroscience*, *35*(31), 10989–11011.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735.

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., & Donner, T. H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Current Biology*, *23*(11), 981–986.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Science*, *6*(10), 421–425.

Price, K. V., Storn, R. M., & Lampinen, J. A. (2006). Differential evolution—A practical approach to global optimization. In *Natural computing*, Springer-Verlag, ISBN 540209506.

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*, 1113–1143.

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.

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

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*(3), 195–237.

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

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

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychological Review*, *120*(3), 697–719. URL http://www.ncbi.nlm.nih.gov/pubmed/23915088.

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

van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E.-J. (2012). Optimal decision making in neural inhibition models. *Psychological Review*, *119*(1), 201–215.

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.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604.

Shadlen, M. N., Kiani, R., Newsome, W. T., Gold, J. I., Wolpert, D., Zylberberg, A., Ditterich, J., de Lafuente, V., Yang, T., & Roitman, J. D. (2016). Comment on "Single-trial spike trains in parietal cortex reveal discrete steps during decision-making". *Science*, *351*(6280), 1406.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.

Sheather, S. J. (2004). Density estimation. *Statistical Science*, *19*, 588–597.

Sheather, S. J., & Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, *53*, 683–690.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, *123*(4), 425–451.

Smith, P. L., & Ratcliff, R. (2015). An introduction to the diffusion model of decision making. In B. U. Forstmann, & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 49–70). New York: Springer.

Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: violations of value invariance in human decision making. *Psychonomic Bulletin and Review*, *23*(1), 22–38.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: from independence to competition. *Psychological Review*, *120*(1), 1–38.

Tsetsos, K., Usher, M., & McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in Neuroscience*, *5*, 63.

Turner, B. M., Dennis, S., & Van Zandt, T. (2013). Likelihood-free Bayesian analysis of memory models. *Psychological Review*, *120*(3), 667–678.

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *Neuroimage*, *72*, 193–206.

Turner, B. M., & Sederberg, P. B. (2012). Approximate Bayesian computation with differential evolution. *Journal of Mathematical Psychology*, *56*(5), 375–385. URL http://www.sciencedirect.com/science/article/pii/S0022249612000752.

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychological Bulletin Review*, *21*(2), 227–250.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368–384.

Turner, B. M., Sederberg, P. B., & McClelland, J. L. (2016). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, *72*, 191–199.
Turner, B. M., Van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: The neural drift diffusion model. *Psychological Review*, *122*, 312–336.
Turner, B. M., & Van Zandt, T. (2014). Hierarchical approximate Bayesian computation. *Psychometrika*, *79*(2), 185–209.
Turner, B. M., & Zandt, T. V. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, *56*(2), 69–85. URL http://www.sciencedirect.com/science/article/pii/S0022249612000272.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.
Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769.
Usher, M., Olami, Z., & McClelland, J. (2002). Hick's law in a stochastic race model with speed-accuracy tradeoff. *Journal of Mathematical Psychology*, *46*, 704–715.
Van Maanen, L., Grasman, R. P. P. P., Forstmann, B. U., Keuken, M., Brown, S. D., & Wagenmakers, E. J. (2012). Similarity and number of alternatives in the random-dot motion paradigm. *Attention, Perception & Psychophysics*, *74*, 739–753.
Van Maanen, L., Grasman, R. P. P. P., Forstmann, B. U., & Wagenmakers, E. J. (2012). Piéron's law and optimal behavior in perceptual decision-making. *Frontiers in Decision Neuroscience*, *5*, article 143.
Wang, X. J. (2008). Decision making in recurrent neural circuits. *Neuron*, *60*(2), 215–234.
Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, *7*, 14.
Winkel, J., Keuken, M. C., van Maanen, L., Wagenmakers, E.-J., & Forstmann, B. U. (2014). Early evidence affects later decisions: why evidence accumulation is required to explain response time data. *Psychonomic Bulletin and Review*, *21*(3), 777–784.
Zandbelt, B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2848–2853.