Parameter recovery, bias and standard errors in the linear ballistic accumulator model

Visser, I.; Poessé, R.

DOI
10.1111/bmsp.12100
Publication date
2017
Document Version
Final published version
Published in
British Journal of Mathematical and Statistical Psychology
License
Article 25fa Dutch Copyright Act

British Journal of Mathematical and Statistical Psychology (2017), 70, 280–296

The British Psychological Society
www.wileyonlinelibrary.com

Special Issue Paper

Parameter recovery, bias and standard errors in the linear ballistic accumulator model

Ingmar Visser* and Rens Poessé
Department of Psychology, University of Amsterdam, The Netherlands

The linear ballistic accumulator (LBA) model (Brown & Heathcote, 2008, *Cogn. Psychol.*, *57*, 153) is increasingly popular in modelling response times from experimental data. An R package, *glba*, has been developed to fit the LBA model using maximum likelihood estimation which is validated by means of a parameter recovery study. At sufficient sample sizes parameter recovery is good, whereas at smaller sample sizes there can be large bias in parameters. In a second simulation study, two methods for computing parameter standard errors are compared. The Hessian-based method is found to be adequate and is (much) faster than the alternative bootstrap method. The use of parameter standard errors in model selection and inference is illustrated in an example using data from an implicit learning experiment (Visser *et al*., 2007, *Mem. Cogn.*, *35*, 1502). It is shown that typical implicit learning effects are captured by different parameters of the LBA model.

1. Introduction

In both experimental and non-experimental settings, response times are of crucial importance in understanding cognitive processes, next to the character of the responses provided by people responding to items or questions. Mathematical models that take both the response and the response time into account have been used in psychology at least since the 1960s and 1970s (Laming, 1968; Ratcliff, 1978). Since that time the use of such models has increased steadily.

In typical applications of response time models, the interest is in testing whether an experimental manipulation has an effect on accuracy and response times, and, in particular, which aspect of cognitive processing is affected by the manipulation. Such aspects are the influence of the difficulty of the task, the caution which participants use to approach the task, a possible trade-off between speed and accuracy, and possible biases they have towards one or another type of response. Most of the models that are used to simultaneously model responses and response times are from the class of so-called evidence accumulation models (Lee & Cummins, 2004), and they are meant to model precisely such aspects of cognitive processes.

Evidence accumulation models assume that evidence for the different response alternatives accumulates over time until a threshold is reached. Such models typically comprise three essential types of parameters related to specific cognitive processes. First,

*Correspondence should be addressed to Ingmar Visser, Amsterdam Brain and Cognition Center, Yield Research Program, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, Amsterdam 1018XA, The Netherlands (email: i.visser@uva.nl).

DOI:10.1111/bmsp.12100

the drift or drift rate parameter(s) are related to the *speed of evidence accumulation* such that higher drift rates lead to faster decisions. Hence, higher drift rates are associated with relatively easier tasks and vice versa. Second, the boundary or threshold parameters are related to the *amount of evidence* that is deemed sufficient to make a decision. A higher boundary thus signals more caution about making decisions. Third, the starting-point parameter is related to the amount of prior evidence or *bias* that is present. A higher starting point for one of the response options than for other options means that less evidence is required to decide for that particular response option.

Although the use of such response time models has increased significantly over recent years, their use remains limited to specialists in this area. For these models to find a larger audience, a number of conditions need to be met. First, naturally, it needs to be clear that their application leads to better understanding of cognitive processes. Meeting this condition requires multiple successful applications of these models, and this condition thus cannot be met by any single paper. We do, however, present an application of an evidence accumulation model to implicit learning. Second, software is required that, by and large, automates the processes of parameter estimation and inference of the model, and provides the necessary statistics to perform model selection. Here we present a software package that facilitates easy model specification and estimation. Third, sound statistical inference methods are required that can be applied and understood by non-specialists. Estimation methods, inference methods, and (easy-to-use) software that implements these, should enable non-specialists to apply these models and test hypotheses based on them. In the current paper we present parameter inference based on maximum likelihood estimation and standard errors and present simulation studies to validate these methods as well as an illustration using real data.

Here we study the linear ballistic accumulator (LBA) model (Brown & Heathcote, 2008), details of which are provided in the next section. The LBA model has a number of attractive properties above other response time models. First, it is relatively easy to understand and interpret while still maintaining the most desirable properties of other common response time models (Brown & Heathcote, 2008; Heathcote & Hayes, 2012), such as the diffusion model (Ratcliff, 1978); see Donkin, Brown, Heathcote, and Wagenmakers (2011) for a comparison of LBA and diffusion models. Those properties are the ability to capture the shape of both correct and incorrect response time distributions and the speed–accuracy trade-off (Brown & Heathcote, 2008). Second, the LBA model can be naturally extended to multiple-choice items, whereas many other response time models are limited to binary-choice tasks (but see van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011; for alternatives). Third, there are closed-form solutions for the densities that follow from the model, making it easy to apply.[1] All of these properties of the LBA model ensure that it has the potential to be used and applied by a wide audience of researchers in psychology and cognitive science.

In the following we first briefly describe the LBA model. Second, we present an R package for fitting the LBA model and discuss the design choices made for the package. Third, we present parameter recovery and bias estimates for a basic model. Fourth, we present a simulation study that compares two methods of computing standard errors of the same basic model and their use in parameter inference. Finally, we illustrate the model

[1] Note that many evidence accumulation models, such as the diffusion model and its variations, require numerical approximation of one or more integral expressions to compute the likelihood; see, for example, Molenaar, Tuerlinckx, and van der Maas (2015) for some computational details of the Q-, D-, and normal drift diffusion models.

by applying it to a data set from an implicit learning experiment (Visser, Raijmakers, & Molenaar, 2007).

2. The linear ballistic accumulator model

The LBA model is an evidence accumulation model, meaning that over time evidence accumulates until a threshold is reached to make a decision. The LBA model has, as its name implies, one essential property that sets it apart from other evidence accumulation models: evidence accumulation in the LBA is linear and ballistic. This means that evidence accumulation has the same rate throughout a trial, rather than being driven by a random process as is the case in other models (Ratcliff, 1978; Usher & McClelland, 2001).

The working of the LBA model is depicted in Figure 1 for a single accumulator. The LBA model has evidence accumulators for each of the possible response alternatives. After stimulus presentation, evidence accumulation begins at a certain starting point, denoted by k. Evidence accumulation proceeds at rate d, and stops when the threshold is reached. As illustrated in the figure, the time to reach threshold can be simply computed as $(b - k)/d$. As in other accumulator models, whichever accumulator reaches the threshold first generates the response.

More formally, response times in the LBA are modelled as follows:

$$t = t_0 + \frac{b-k}{d}, \quad \text{with } k \sim U(0, A) \text{ and } d \sim N(v_i, s_v), \tag{1}$$

with the following parameters. First, we have the drift rate parameters v_i, for each accumulator i, denoting the (average) rate of evidence accumulation. Second, we have the standard deviation s_v of the drift rates v; in the LBA, the drift rates for the accumulators are normally distributed with means v_i and standard deviation s_v; together with the v_i, this means that the ds are normally distributed according to $N(v_i, s_v)$. Third, we have the starting-point parameter A; the starting point of the evidence accumulation, k, is drawn from a random uniform distribution of which A is the end point, that is, $k \sim U(0, A)$. The starting-point parameter is associated with bias for a particular response (or in a particular condition). Fourth, we have the boundary parameter b, which is the threshold for evidence accumulation. The boundary parameter is associated with response caution as a

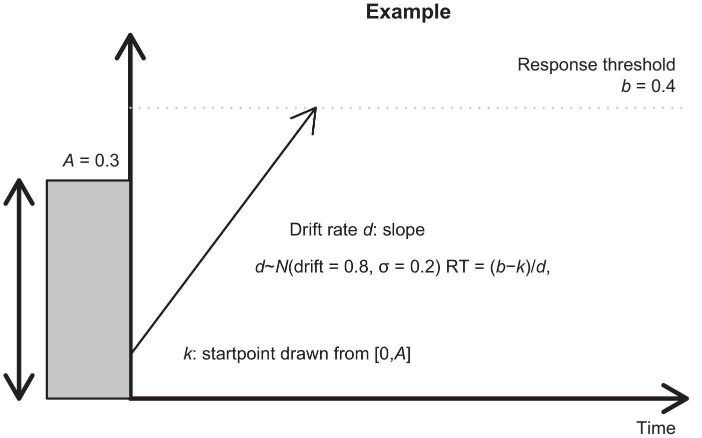

Figure 1. Linear ballistic accumulator model illustrating the parameters; see text for details.

higher boundary sets a higher threshold for evidence to accumulate before a decision is made. Finally, we have the non-decision time parameter t_0; this parameter accommodates processes unrelated to the choice task such as stimulus encoding and the generation of motor responses. Note that d and k necessarily vary with trials, whereas b and t_0 typically do not vary with trials.

The simplest LBA model, for a binary choice task without further conditions, has six parameters: v_1 and v_2 are the (average) drift rates for both choices, A is the starting-point parameter, s_v the variability of the drift rates, b the boundary parameter, and t_0 the non-decision time. Note that this model is not identified without setting a scaling constraint: scaling is necessary because the drift rate, drift rate variability, starting point and boundary parameters can be multiplied with a constant to yield the same response times, as can easily be seen by considering the equation that generates the response time $(b - k)/d$. The most commonly used constraint, which we use throughout this paper, is to set the sum of the drift rates to 1. In the example here, that means that $v_1 + v_2 = 1$, and hence the model has five free parameters left. See Donkin, Brown, and Heathcote (2009) for discussion of the scaling constraint in evidence accumulation models.

The LBA model is typically estimated using the likelihood, although likelihood-free methods have recently been developed as well (Holmes, 2015; Turner & Sederberg, 2014). Assuming response time data Y_t, the likelihood for the data is expressed as

$$L(Y_t|\theta) = \prod_t \left[f_i(Y_t - t_0) \prod_{j \neq i} F_j(Y_t - t_0) \right], \tag{2}$$

where i represents the response category of the response time being modelled, θ represents the parameter vector of the model, $\theta = (t_0, b, A, s_v, v_1, \ldots, v_n)$, for data with n possible categories. Here, $f_i(\cdot)$ and $F_i(\cdot)$ are the probability and cumulative density functions for the ith accumulator reaching threshold, defined in equations 1 and 2 in Brown and Heathcote (2008, p. 159), respectively. Informally, the likelihood for any given response time associated with an accumulator i is the product of the density of that accumulator reaching threshold at time Y_t, and the densities of the other accumulators not yet having reaching their thresholds. We further follow the procedure for estimation of model parameters in Donkin, Brown, and Heathcote (2011, p. 149). This procedure is implemented in an R package, *glba*, which is outlined in the next section.

3. The *glba* package

To increase the accessibility of the LBA model for applied researchers, here we present an R package that allows easy specification of the model as well as robust starting-value generation, meaning that the chances of finding the global maximum of the likelihood are high.[2] The package further offers parameter standard errors and common model selection statistics to aid in model and parameter inference. Ease of model specification is implemented in *glba* by using the formula interface that is common in R for specifying linear and general linear models. This allows the user to specify the factors or covariates that determine each of the five parameter sets of the LBA model.

In particular, in *glba* all parameters of the LBA model can be made to depend on predictors in the following way:

[2] Without appropriate starting-value generation, optimization may easily get stuck in local maxima of the likelihood, resulting in nonsensical parameter estimates.

$$\theta = \beta X, \tag{3}$$

where θ is any parameter of the LBA model, β is a vector of regression coefficients and X is a design matrix. Note that by having a column of 1's in X, the corresponding element of β can be interpreted as the intercept. Next to typical tests about differences between (discrete) conditions in an experiment, this set-up is equally suitable to test the effects of continuously varied experimental variables. In Section 6 this is illustrated using a real data example.

In *glba*, the logartithm of the likelihood in equation (2) is used to optimize parameter values for given data. More specifically, *glba* uses full-information maximum likelihood (FIML) or continuous maximum likelihood estimation of the unknown parameters (Donkin, Brown, *et al.*, 2011). This means that the likelihood contributions of individual data points are used rather than the likelihoods for summary statistics of the observed distributions such as the quantiles. The latter is done in, for example, Donkin, Averell, Brown, and Heathcote (2009); see Donkin, Brown, *et al.* (2011) for a tutorial on this and other issues in fitting the LBA model.

The reason why *glba* uses FIML instead of quantile maximum probability estimation (QMPE; Heathcote, Brown, & Cousineau, 2004) is that the former allows easier inclusion of predictors on the model parameters. That is, using summary statistics such as quantiles in parameter estimation is relatively straightforward in a factorial design but less so in a designs with continuous predictors.

Quantile maximum probability estimation has been shown to be less biased and more robust against outliers when estimating response time parameters (Heathcote, Brown, & Mewhort, 2002) compared with FIML. However, when the modeller is aware of this, using FIML with robust starting values for the unknown parameters and possibly outlier detection produces good results in estimating the LBA model (Donkin, Brown, *et al.*, 2011). FIML estimation is also usually faster than QMPE[3] and provides the maximum likelihood estimates that can be subsequently used in model selection measures such as the Akaike (AIC) and Bayesian information criteria (BIC).

Other features of the *glba* package that are worth mentioning are the following. First, there is the option to fix parameters to particular values, which is useful for testing hypotheses about them. Second, optionally the Hessian can be estimated at the optimal parameter values, which is then used to provide parameter standard errors; this will be applied in Section 7. Third, *glba* uses the standard R package *optim* to optimize parameter values, which allows for customization of the optimization process, for example, by choosing the optimization method. Another possibility is to include box constraints on parameters when this is deemed necessary or interesting. Fourth, and finally, *glba* optionally employs a function to generate starting values for the parameters, which in our experience is quite robust.

The approach to LBA estimation taken here is different from other recent approaches in that we use maximum likelihood estimation and classical inference methods rather than Bayesian analysis and inference methods. Bayesian analysis has become a popular method in response time modelling in general, and in applying the LBA model in particular. The Bayesian approach has the advantage of being very flexible in the possibilities for model specification. For example, random effects for individual participants and regressor functions on the parameters as in the current paper can

[3] Heathcote *et al.* (2002) mention that the difference is 'an order of magnitude'; the relevance of this, given that computational speeds have increased significantly since 2002, is unknown in the absence of tests to this effect.

also be naturally incorporated in a Bayesian estimation and inference framework (Wiecki, Sofer, & Frank, 2013). An important advantage of the hierarchical Bayesian approach that is frequently mentioned is that parameter uncertainty is automatically part of the analysis, and that the full posterior distributions of the parameters are available and can be inspected for inferential purposes (Wiecki *et al.*, 2013). These advantages, however, may be not as important as they seem, and there are potential downsides as well. First, in many practical applications, rather than using the full posterior distribution of parameters, oftentimes summary statistics of this distribution such as the median or maximum *a posteriori* estimates are used in inference (e.g., Turner, Forstmann, Wagenmakers, Brown, Sederberg, & Steyvers, 2013). Second, Bayesian approaches can lead to large variability in parameter estimates when individual analyses are performed (e.g., Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009), which are then hard to interpret (see, for example, Heathcote & Hayes, 2012; for discussion of the Dutilh *et al.*, 2009, results). Third, the approach adopted here with FIML and linear predictors on the variables provides the flexibility that is common in the hierarchical Bayesian approach, without adding the complexity of Bayesian analyses. Moreover, classical parameter estimation and inference using standard errors provides a measure of parameter uncertainty.

In the current paper, parameter uncertainty is quantified by computing standard errors for parameters that can also be used for parameter inference. To the best of our knowledge, this approach has not been taken in the existing literature on the LBA model. To validate the functionality of the *glba* package, we next study parameter recovery and bias with simulated data.

4. Simulation I: Parameter recovery and bias

To establish the statistical properties of the LBA model, a first requirement is that the parameters are recovered and can be estimated properly and without bias. We present a paradigmatic and simple case in which only the drift rate differs between conditions in an experiment. Parameter recovery and bias were tested by repeatedly simulating data from the model and fitting models to the simulated data. The code for this simulation as well as for the second simulation study and the analysis code for the illustration in Section 6 of this paper is all provided in the online supporting information section.

4.1. Design

The model that is tested here is a simple model which generates data in an ‘easy’ versus a ‘difficult’ condition. The two conditions only differ in the drift rate parameters v, which were set at $v_{\text{difficult}} = .6$ and $v_{\text{easy}} = .8$ for the correct responses, respectively. Note that this means that the average drift rates for the incorrect responses were .4 and .2, respectively. Parameter values are otherwise similar to those reported in Donkin, Averell, *et al.* (2009) for ease of comparison. However, they used QMPE fitting of the LBA model rather than FIML. The top row of Table 1 provides the true parameter values used in the simulation. The sample sizes were 80, 150, and 300, respectively, meaning that in each condition 40, 75, and 150 data points were generated. These numbers are both realistic and in the range where potentially bias and/or recovery issues could occur. For each sample size, 1,000 data sets were generated and fitted. No data selection was done prior to

Table 1. Simulation results for the 'difficulty' model, based on 1,000 simulations for each sample size. Results are averages of the parameter estimates, with the root mean squared errors for these estimates in parentheses. 'Fail' indicates the number of models that did not converge or led to inadmissible parameter estimates; see the text for details

	Fail	s_v	A	b	t_0	$v_{\text{difficult}}$	v_{easy}
True		0.200	0.300	0.100	0.200	.600	.800
80	28	0.202 (0.10)	0.313 (0.09)	0.094 (0.07)	0.211 (0.08)	.615 (0.08)	.850 (0.16)
150	8	0.199 (0.06)	0.304 (0.05)	0.098 (0.04)	0.202 (0.05)	.603 (0.04)	.815 (0.10)
300	4	0.198 (0.03)	0.303 (0.03)	0.097 (0.02)	0.203 (0.03)	.601 (0.02)	.807 (0.05)

model fitting, that is, no outlier detection was applied and all the data are used in fitting the models.

4.2. Results

No outlier removal was done prior to model fitting, and hence the convergence results provide information on the robustness of the model. Initially, 302 (out of 3,000) models did not converge or resulted in inadmissible parameter estimates (e.g., negative values for t_0 or b). These models were reinitialized with the starting-value generation function in *glba*.[4] After this second round of optimization, there were 104 models left that did not converge or had inadmissible parameters estimates. These models were reinitialized again and optimized using a different (and slower) optimization algorithm. After this round of optimization, models for 40 data sets, or 1.3% of all data sets, still did not converge or led to inadmissible estimates. The numbers for each sample size are reported in Table 1.

Table 1 shows the recovered parameter averages for the different sample sizes. The average parameter estimates are close to their true values, and, as expected, this is more so for larger sample sizes.

As the table indicates, parameter recovery is good and biases are acceptable at sample sizes 150 and 300. At sample size 80, at least some parameters have large biases that could potentially hinder reliable inference, although it should be noted that the variances are considerably larger as well. Also, at sample size 80, issues with convergence and inadmissible parameter estimates occur more frequently than at larger sample sizes. Hence, for this model, 150 data points seems to be the minimal sample size resulting in unbiased parameter estimates. In the next section we compute the standard errors of this model using this same sample size to study parameter uncertainty.

It should be noted that the parameter bias becomes considerably smaller when only considering the models that converged initially. This is especially so for the parameter with the largest bias, the drift rate parameter in the easy condition. The reason for the large bias in this parameter is that in small data sets as we consider here, combined with a high drift rate, as is the case in the easy condition, it may be that there are few or no error responses in the data. Few error responses can lead to degenerate models where the drift rate estimates are pushed to unity or beyond.

[4] Note that in initial optimization the true parameter values were used as starting values.

5. Simulation II: Parameter covariance and standard errors

Deriving the covariance matrix of parameter estimates is important for several reasons. First, the parameter covariance or correlation matrix can provide information on the identifiability of the model. Correlations between parameters that approach unity are an indicator of a lack of identifiability. Such lack of identifiability hinders parameter estimation and may lead to bias. Second, the parameter covariance matrix can be used to derive standard errors for the parameters. Standard errors, in turn, can be used in inference about parameter estimates and hypothesis testing. For example, testing whether an experimental condition has a significant effect on one of the parameters of the model can be done by using the confidence interval of the estimated parameter or the Wald test.

The latter application can be seen as an alternative to the customary practice in response time modelling of using model selection criteria such as the AIC or BIC. In response time modelling it is customary to specify a large number of models and then use model selection criteria to choose the optimal model. For example, Forstmann *et al.* (2010) used eight models, Ho, Brown, and Serences (2009) selected their model from a set of 28 models, and Ho, Brown, van Maanen, Forstmann, Wagenmakers, and Serences (2012) similarly used the BIC to select among a set of candidate models. Here we use an alternative approach using parameter standard errors in which inference proceeds by selecting models in which effects are significantly different from zero and/or interpreting estimated parameters when that is the case. We present an example of such inference below.

5.1. Method and design

Two methods of computing the parameter covariance matrix (and the resulting standard errors) are compared. The first method to obtain the parameter covariance is a bootstrap procedure. Repeated data samples are constructed by resampling from the data a large number of times, say B. Subsequently, the model is re-fitted to these B samples, resulting in B estimates of the parameters. The parameter covariance matrix can then be approximated by the covariance matrix of the B bootstrap estimates. See Efron and Tibshirani (1986, 1993) for a general introduction to bootstrap-based inference.

As the LBA model is here estimated by maximum likelihood, the likelihood function can also be used to derive the parameter covariance matrix. Assuming multivariate normality of the parameter estimates at their optimal values, the inverse of the Hessian matrix is the parameter covariance matrix. Hence, the second method to obtain the parameter covariance matrix is through the Hessian matrix of the log-likelihood function, in particular by computing a finite-difference approximation to the Hessian. The *glba* package provides an optional argument to return the Hessian, which is evaluated at the optimal parameter estimates.

Assuming parameter estimates $\hat{\theta}_i$ and standard errors $\hat{\varepsilon}_i$, 95% approximate confidence intervals are given by (Efron & Tibshirani, 1993)

$$\hat{\theta}_i \pm 1.96 \times \hat{\varepsilon}_i,$$

where $\hat{\varepsilon}_i$ is a standard error estimate obtained from one of the two methods mentioned above. The quality of standard errors is evaluated by studying the coverage probabilities of the confidence intervals constructed by them. The theoretical coverage probability of this interval is 95%.

The model used in this simulation is the same as in the parameter recovery study. The simulation involves generating data from the model, fitting a model to the generated data set, and computing standard errors by the two methods above. The simulation was run 1,000 times, with a sample size of 150. The latter was chosen as the (minimal) sample size from the parameter recovery study that results in adequate parameter estimates. For computing the bootstrap covariance, the number of samples was set to 100.

5.2. Results

Figure 2 shows the average standard errors that were obtained with each method. The bottom panel of Figure 2 shows the coverage rates for each of the parameters and for each of the methods.

As can be seen in Figure 2, the standard errors resulting from both methods are quite similar, although the standard errors from the finite-difference Hessian are the smallest for each parameter. The coverage rates of the parameters are close to their theoretical values, with the Hessian-based method leaning towards conservative estimates and the bootstrap method being slightly too lenient. Note that the bootstrap method is also computationally

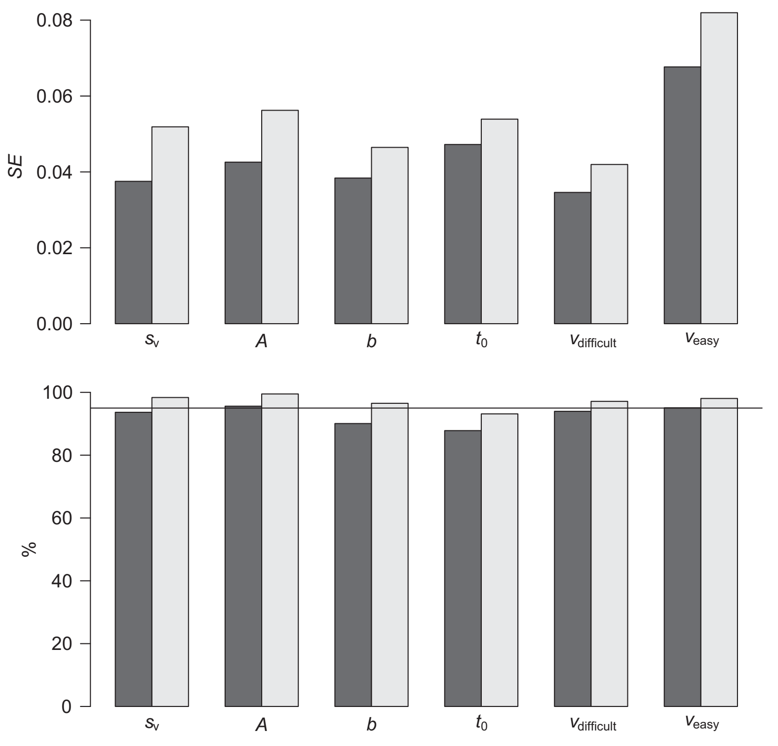

Figure 2. Top: standard errors based on the Hessian (left-hand bars), and the bootstrap methods for each parameter using sample size 150 and 100 bootstrap samples. Results are averages from 977 replications, 23 simulations having failed due to convergence issues or inadmissible parameter estimates. Bottom: coverage probabilities of each parameter of the model; left-hand bars for the Hessian-based method. The horizontal line indicates the theoretical coverage rate of 95%.

Table 2. Example of the 'difficulty' model with parameter estimates and standard errors. All p-values are <.01. The columns after the z-values provide the parameter correlations based on the finite-difference approximation of the Hessian. Parameter estimates and correlations are from a single replication of the model

	Value	*SE*	z	A	b	t_0	$v_{\text{difficult}}$	v_{easy}
s_v	0.200	0.035	5.7	−0.031	−0.065	0.268	.346	.455
A	0.320	0.034	9.5	–	−0.701	0.611	.187	.623
b	0.067	0.023	2.9	–	–	−0.947	−.151	−.559
t_0	0.263	0.027	9.7	–	–	–	.232	.643
$v_{\text{difficult}}$	0.601	0.033	18.2	–	–	–	–	.304
v_{easy}	0.834	0.062	13.4	–	–	–	–	–

much more costly than the finite-difference method; for the latter method, the computation cost in addition to fitting a model is almost negligible, whereas in the bootstrap method fitting as many models as bootstrap samples is required.

By way of example, in Table 2 we provide the parameter estimates for an example of the difficulty model along with the standard errors, and the corresponding p-values based on the z-ratio; that is, assuming that $\theta/SE(\theta)$ follows a standard normal distribution. Similarly, based on the same model, the table also provides the parameter correlation matrix. Note that these parameter estimates as well as the correlation matrix are from a single replication from the simulation study.

This correlation matrix is useful in studying the identifiability of models and data at hand. In the 'difficulty' model, it can be seen that the correlation between t_0 an b is quite high. Even so, at the sample size of 150 data points used in this example, parameter estimation is feasible. In the next section we illustrate the LBA model and parameter inference using the *glba* package with data from an implicit learning experiment.

6. Illustration: Implicit learning

Seger (1994) defines implicit learning as 'non-episodic learning of complex information in an incidental manner, without awareness of what has been learned'. Whether indeed such learning also results in knowledge without awareness and is supported by dual systems is a highly contested issue (for opposing views, see Destrebecqz & Cleeremans, 2001; Shanks & Perruchet, 2002). Regardless of this issue, in implicit learning experiments, participants learn to respond to stimulus material faster or more efficiently without instructions to do so.

A popular task in implicit learning that shows such faster processing is the serial response time task (SRTT; Nissen & Bullemer, 1987). In the SRTT, participants respond to a stimulus that moves from one location to the next between trials. Participants' task is to respond to each different stimulus location with a corresponding, congruently mapped, response key. Unbeknownst to participants, the sequence of locations follows a regular or repeating pattern. After some training with one particular sequence, the sequence of locations switches to a different or a random pattern (see, for example, Pronk & Visser, 2010; Reed & Johnson, 1994, for discussion about these different options).

6.1. Data

Here we model response times and accuracies from an SRTT reported in Visser *et al.* (2007); the data set is included in the *glba* package. The data consist of 12 blocks of trials from a single participant in Experiment 2. Each block consists of 395 forced-choice trials responding to a location sequence with four different positions. The sequence of locations follows a pattern that is generated by a finite-state automaton (also called a regular grammar; see Visser *et al.*, 2007; for details). In blocks 6 and 12 the sequence of locations is (pseudo-)random, with the only restriction that direct repetitions do not occur. The average response times for each block are shown in Figure 3. The effect of changing to the random sequence in blocks 6 and 12 is clearly visible.

There are two important characteristics of the data that we want to capture. First, the response times in sequential blocks decrease rapidly throughout the first few blocks and then stabilize. Second, the switch to the random sequences results in a large increase in response times. The latter effect is the classical effect that is interpreted as evidence for implicit learning. The explanation for the decrease in response times is that, due to the repetitive nature of the sequence, participants develop expectations about which stimulus will appear next. As a result, they can prepare for the upcoming response and hence respond faster. In the random blocks, however, these expectations no longer hold true and hence result in an increase in response times.

6.2. Models

To characterize the implicit learning process in these data, we fitted a number of LBA models with various predictors for the LBA parameters. Before fitting the models, extremely low response times were removed. Although in this task it is common to find

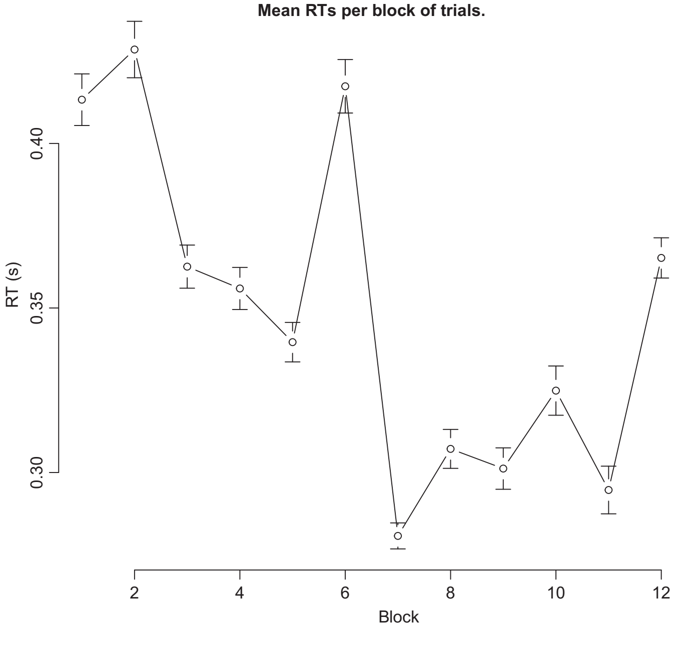

Figure 3. Implicit learning data from a single participant responding to a sequence of locations; data from experiment 2 in Visser *et al.* (2007). Error bars around the means denote standard errors.

Table 3. Parameter estimates of linear ballistic accumulator model of implicit learning, with effect of the sequence factor on A, b and v. *SE* denotes the standard error of the parameter based on the Hessian, z denotes the z-ratio of a parameter and its standard error, and p denotes the corresponding p-value

	Value	*SE*	z	p
s_v	0.27	0.01	25.15	.00
A	0.17	0.02	9.68	.00
A_{sequence}	0.04	0.02	2.41	.01
b	0.18	0.01	18.38	.00
b_{sequence}	−0.05	0.01	9.12	.00
t_0	0.07	0.01	8.60	.00
v	0.89	0.03	33.67	.00
v_{sequence}	0.03	0.02	1.37	.09

very low response times due to the (implicit) learning effects, response times below 150 ms were removed as such a short time hardly provides the opportunity to perceive the stimulus. Using the 150 ms threshold resulted in removing 60 response times from a total of 4,720, amounting to 1.3% of the data.

Expectation in the LBA model is captured by the bias parameter, A. Higher values of A result in faster response times. The difference between sequence and random blocks is hence likely to result in different bias parameters. Although we only hypothesize that the sequence factor affects bias in the model, we also include this factor in the boundary and drift rate parameters as these three parameters together determine the decision speed. Hence, the first model we fitted contains effects of the sequence factor in the experiment on parameters A, b, and v, and the resulting parameter estimates of this model are listed in Table 3. The table also reports the standard errors of the parameters and the associated z- and p-values.

The sequence factor has significant effects on the bias parameter of the model as well as on the boundary parameter, as evidenced by significant p-values. The effect on the drift rate falls short of significance with p = .09. The effect on the A parameter is in the expected direction, with a higher bias for sequential trials than for random-order trials. The effect on the b parameter can also be readily explained. When the order of trials changes from sequential to random in blocks 6 and 12 of the experiment, it is likely that participants need to adapt their response caution in order to keep performing at the same level of accuracy. Indeed, the accuracy level is high and constant throughout the experiment.

In the second model we fitted, the (non-significant) effect of the sequence factor on drift rate was dropped from the model. The resulting model has parameter estimates that are very similar to those reported in Table 3, and all the parameters now have significant p-values.

The sequence factor in these first two models captures differences between the blocks with sequential trials and those with random-order trials, and thereby captures the essence of the implicit learning effect. As is obvious from Figure 3, there is more variability in these data than only this difference: the response times in the sequence blocks decrease rapidly with training, especially in the first few blocks. To capture this effect in the LBA model, the third model we fitted adds an increasing quadratic trend for the sequential blocks on the drift rate parameter. The parameter estimates for this model are listed in Table 4.

Table 4. Parameter estimates of linear ballistic accumulator model of implicit learning, with effect of the sequence factor on A and b. The drift rate has a quadratic covariate in the sequence blocks, not in the random blocks. *SE* denotes the standard error of the parameter based on the Hessian, z denotes the z-ratio of a parameter and its standard error, and p denotes the corresponding p-value

	Value	*SE*	z	p
s_v	0.24	0.01	22.24	.00
A	0.08	0.03	3.11	.00
A_{sequence}	0.07	0.02	3.86	.00
b	0.23	0.02	13.45	.00
b_{sequence}	−0.04	0.01	5.19	.00
t_0	0.02	0.01	1.52	.06
v	0.77	0.02	44.70	.00
v_{quad}	0.20	0.01	13.20	.00

As can be seen in Table 4, all the relevant parameter estimates remain significant. The effect of the covariate labelled v_{quad} in the table, is large and highly significant, whereby the drift increases from .77 to .95 over the course of learning. In other words, the task becomes significantly easier in the sequence blocks of the experiment, and not in the random blocks of the experiment. Note that the non-decision parameter t_0 is no longer significant in this model.

6.3. Conclusion

The LBA models fitted here nicely capture the two most important aspects of implicit learning data. The difference between sequential and random trial blocks is modelled by a difference in the starting-point parameter A. This is consistent with the idea that in the SRTT, participants come to expect the locations of the stimuli based on the stimuli they have processed prior to the current stimulus (Cleeremans & McClelland, 1991). Second, learning in the task is also captured by an increase in the drift rate in the model for sequence blocks, but not so for random blocks. That is, processing of the sequential stimuli becomes easier with practice. This effect is independent of and additional to the effect of the bias that participants have for sequential stimuli. This increase in drift rate may be interpreted as an increase in the fluency of decision-making and perceptual processing of the stimulus that has been theorized to underly implicit learning and implicit memory effects (Buchner, Steffens, & Rothkegel, 1997; Kinder, Shanks, Cock, & Tunney, 2003).

7. General discussion

The LBA model is arguably the simplest evidence accumulator model that can be used to simultaneously model response times and accuracy while still being flexible enough to capture typical phenomena (see Donkin, Heathcote, & Brown, 2009; for discussion on the ‘simplicity’ of the model). As such, the model stands a good chance of being adopted by a larger audience of applied researchers. In order to facilitate such adoption by a larger audience, easy-to-access methods for model fitting, parameter inference and model selection are a necessary condition. The R package *glba* provides such easy access, intuitive model specification, and, robust starting-value generation.

In the current paper, the *glba* package is validated by successful parameter recovery and by showing that parameter estimates are unbiased given sufficient data. In addition, and more importantly, we proposed and validated two methods for computing standard errors of the LBA model parameters. To the best of our knowledge, standard errors have not been reported before in the extant literature on the LBA model. Standard errors can in turn be used in model selection procedures, alongside the typical model selection measures such as AIC and BIC, and we have shown examples of this.

The quality of standard errors by both methods is good, with the Hessian-based method being slightly more conservative. The computational cost of the bootstrap methods is much higher than that of the finite-difference approximation and hence the latter method is preferred. Parameter standard errors are a valuable tool in assessing the strengths of experimental factors. Their availability is thus important in opening up new areas of application for the LBA model. Next to standard errors, the Hessian was also used to compute the parameter correlation matrix. When encountering very high correlations between parameters it may make sense to either gather more data for individual participants or pool data for different participants together.

In applications of the LBA model it is common to estimate model parameters based on data from single individuals. Subsequently, these parameter estimates are aggregated to draw conclusions about the effects of experimental conditions. A second approach is to allow the parameters of the model to vary randomly over individuals, that is, including multilevel structure in the parameters. Wiecki *et al.* (2013) have taken such an approach with Bayesian estimation, and Molenaar *et al.* (2015) implement a variant of the diffusion model with random effects; both papers also present software that can be used to carry out the analyses. A third possible approach is to have some parameters identical across individuals while leaving others free to vary. The *glba* set-up allows for such parameter designs to be specified without needing additional programming.

The LBA model and the use of *glba* were illustrated using real data from an implicit learning experiment. The response time models for implicit learning indicate that in such a learning environment a number of different cognitive processes are at work. Learning results in changing expectations of stimuli as well as faster processing of these same stimuli. Moreover, switching between blocks of sequence and random trials results in a change in the setting of the speed–accuracy trade-off. Work in related areas also shows that experimental manipulations seldom affect only a single cognitive process. Rae, Heathcote, Donkin, Averell, and Brown (2014) discuss a manipulation of the speed–accuracy trade-off that is shown to affect not only response caution, as expected, but also the drift rate. Similarly, Dutilh *et al.* (2009) show that as a result of practice with a task, almost all parameters of the drift diffusion model are affected. Response models such as the LBA are essential in capturing such, sometimes subtle, effects in the data. The conclusion for implicit learning at the very least is that both strategic and more implicit processes are at work, as evidenced by changes in three parameters of the LBA model. The ability of the LBA model to capture such experimental phenomena makes it a promising model in many areas of research, and the *glba* package can make such applications more feasible for non-specialist researchers.

References

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

Buchner, A., Steffens, M. C., & Rothkegel, E. E. R. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *Quarterly Journal of Experimental Psychology: Section A*, *50*, 631–663. https://doi.org/10.1080/713755723

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235–253. https://doi.org/10.1037/0096-3445.120.3.235

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? new evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*, 343–350. https://doi.org/10.3758/BF03196171

Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods*, *41*, 1095–1110. https://doi.org/10.3758/BRM.41.4.1095

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, *16*, 1129–1135. https://doi.org/10.3758/PBR.16.6.1129

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, *55*(2), 140–151. https://doi.org/10.1016/j.jmp.2010.10.001

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, *18*(1), 61–69. https://doi.org/10.3758/s13423-010-0022-4

Donkin, C., Heathcote, A., & Brown, S. (2009). Is the linear ballistic accumulator model really the simplest model of choice response times: A Bayesian model complexity analysis. In *9th International Conference on Cognitive Modeling – ICCM2009*, Manchester, UK.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E. J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*, 1026–1036. https://doi.org/10.3758/16.6.1026

Efron, B., & Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, *1*, 54–77. https://doi.org/10.1214/ss/1177013817

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E. J., . . . Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences, USA*, *107*, 15916–15920. https://doi.org/10.1073/pnas.1004932107

Heathcote, A., Brown, S., & Cousineau, D. (2004). QMPE: Estimating lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, & Computers*, *36*, 277–290. https://doi.org/10.3758/BF03196299

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

Heathcote, A., & Hayes, B. (2012). Diffusion versus linear ballistic accumulation: Different models for response time with different conclusions about psychological mechanisms? *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 125–136. https://doi.org/10.1037/a0028189

Ho, T. C., Brown, S., & Serences, J. T. (2009). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience*, *29*, 8675–8687. https://doi.org/10.1523/JNEUROSCI.5984-08.2009

Ho, T. C., Brown, S., van Maanen, L., Forstmann, B. U., Wagenmakers, E. J., & Serences, J. T. (2012). The optimality of sensory processing during the speed-accuracy tradeoff. *Journal of Neuroscience*, *32*, 7992–8003. https://doi.org/10.1523/JNEUROSCI.0340-12.2012

Holmes, W. R. (2015). A practical guide to the probability density approximation (pda) with improved implementation and error characterization. *Journal of Mathematical Psychology*, *68*, 13–24. https://doi.org/10.1016/j.jmp.2015.08.006

Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General*, *132*, 551–565. https://doi.org/10.1037/0096-3445.132.4.551

Laming, D. R. J. (1968). *Information theory of choice reaction times*. New York, NY: Academic Press.

Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the 'take the best' and the 'rational' models. *Psychonomic Bulletin & Review*, *11*, 343–352. https://doi.org/10.3758/BF03196581

Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. J. (2015). Fitting diffusion item response theory models for responses and response times using the R package diffIRT. *Journal of Statistical Software*, *66*(4), 1–34. https://doi.org/10.18637/jss.v066.i04

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Pronk, T., & Visser, I. (2010). The role of reversal frequency in learning noisy second order conditional sequences. *Consciousness and Cognition*, *19*, 627–635. https://doi.org/10.1016/j.concog.2009.12.008

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1226–1243. https://doi.org/10.1037/a0036801

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 585–594. https://doi.org/10.1037/0278-7393.20.3.585

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*(2), 163–196. https://doi.org/10.1037/0033-2909.115.2.163

Shanks, D. R., & Perruchet, P. (2002). Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomic Bulletin & Review*, *9*, 362–367. https://doi.org/10.3758/BF03196294

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. https://doi.org/10.1016/j.neuroimage.2013.01.048

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, *21*, 227–250. https://doi.org/10.3758/s13423-013-0530-0

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, *118*, 339–356. https://doi.org/10.1037/a0022749

Visser, I., Raijmakers, M. E., & Molenaar, P. C. (2007). Characterizing sequence knowledge using online measures and hidden Markov models. *Memory & Cognition*, *35*, 1502–1517. https://doi.org/10.3758/BF03193619

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics* 7, article 14. https://doi.org/10.3389/fninf.2013.00014

Received 3 April 2016; revised version received 20 March 2017

Supporting Information

The following supporting information may be found in the online edition of the article:

Data S1. Simulation I: Parameter recovery and bias.
Data S2. Simulation II: Parameter covariance and standard errors.
Data S3. Illustration: Implicit learning.