THEORETICAL REVIEW

Sequential sampling models without random between-trial variability: the racing diffusion model of speeded decision making

Gabriel Tillman[1,2] · Trish Van Zandt[3] · Gordon D. Logan[2]

Abstract

Most current sequential sampling models have random between-trial variability in their parameters. These sources of variability make the models more complex in order to fit response time data, do not provide any further explanation to how the data were generated, and have recently been criticised for allowing infinite flexibility in the models. To explore and test the need of between-trial variability parameters we develop a simple sequential sampling model of N-choice speeded decision making: the racing diffusion model. The model makes speeded decisions from a race of evidence accumulators that integrate information in a noisy fashion within a trial. The racing diffusion does not assume that any evidence accumulation process varies between trial, and so, the model provides alternative explanations of key response time phenomena, such as fast and slow error response times relative to correct response times. Overall, our paper gives good reason to rethink including between-trial variability parameters in sequential sampling models

Keywords Response time · Sequential sampling models · Decision making

Evidence accumulation is arguably the most dominant theory of how people make speeded decisions (see Donkin & Brown, 2018, for a review), and it is typically instantiated in sequential sampling models (e.g., Ratcliff, 1978; Usher & McClelland, 2001; Brown & Heathcote, 2008). These models provide an accurate account of correct and error response time (RT) distributions as well as the corresponding accuracy rates in speeded decision making tasks. The models also allow researchers to translate the data into the meaningful psychological parameters that generate the data.

Sequential sampling models assume a simple cognitive architecture consisting of stimulus encoding, response selection, and overt response execution. To make a

decision, people begin with an initial amount of evidence for all response options, the *starting point* of evidence accumulation (Fig. 1). From the starting point, more evidence is continually sampled from the stimulus, which accumulates at a rate of *drift rate* towards the corresponding *response threshold*. When the accumulated evidence crosses a response threshold this triggers the corresponding overt response. The quality of evidence sampled from the stimulus governs the drift rate, which can be interpreted as the speed of information processing. Higher response thresholds mean that a person needs more evidence to trigger a response, and so, threshold settings represent how cautious a person is. Starting points and response thresholds can be different for different response options capturing any inherit biases people have. The time necessary for processes outside of evidence accumulation is the non-decision time, which includes the time needed for perceptual encoding and overtly executing a motor response.

Researchers often assume that the core parameters of sequential sampling models, such as drift rates, starting points, and non-decision times vary between trials. There are a number of reasons for these sources of variability. Below we will discuss the evidence for between-trial variability in drift rates, starting points, and non-decision times.

✉ Gabriel Tillman
gabrieltillman.au@gmail.com

[1] School of Health and Life Sciences, Federation University, Ballarat, Australia

[2] Department of Psychology, Vanderbilt University, Nashville, TN, USA

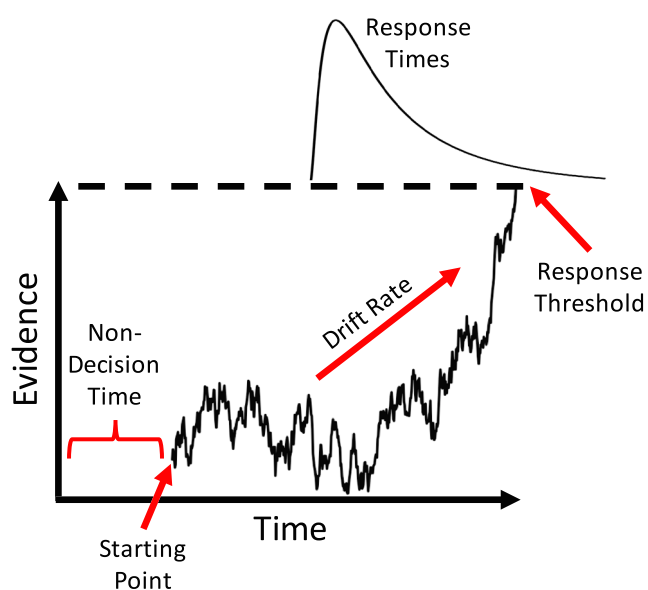[3] Department of Psychology, The Ohio State University, Columbus, OH, USA

Fig. 1 A typical sequential sampling model process. Only one accumulator is illustrated, but multiple alternative decisions are modeled by multiple racing accumulators (Usher & McClelland, 2001; Brown & Heathcote, 2008; Brown & Heathcote, 2005) or single accumulators with two boundaries (Ratcliff, 1978; Link & Heath, 1975)

Evidence for variability in drift rate

Ratcliff (1978) introduced drift-rate variability to model item differences in a recognition-memory task using a single accumulator model with two boundaries, which is known as the drift diffusion model (DDM; Ratcliff and Tuerlinckx, 2002). In this paradigm, between-trial variability in drift rate is thought to be analogous to the memory evidence distributions in signal detection theory (e.g., Egan, 1958). In recognition memory, sequential sampling models produce the same conclusions about evidence variability as signal detection theory (such as conclusions drawn from zROC slopes; Starns & Ratcliff, 2014) and can accurately measure relative differences in evidence variability when it is directly manipulated (Starns, 2014; Osth et al., in press).

Without between-trial variability in drift rate the DDM predicts perfect accuracy with unlimited processing time (Usher & McClelland, 2001; Turner et al., 2017). The variability is also useful because it enables the DDM to predict the common finding of slower errors than correct responses (Ratcliff & Rouder, 1998). In fact, between-trial variability in drift rate is largely determined by the relative speeds of correct and error RT distributions, rather than the overall shape of the RT distributions and accuracy rates (Ratcliff, 2013). Between-trial variability in drift rate is not the only component that can predict relatively slow error RTs. Within the DDM architecture, some researchers

suggest collapsing response boundaries as a function of time (Ditterich, 2006a, b), which means less evidence is needed to make a response over time and results in longer decisions being less accurate. Link and Heath (1975) noted that the random walk, which is a DDM in discrete time, can predict unequal error and correct RT distributions with certain distribution assumptions for the moment-to-moment evidence accumulation rate, such as the Laplace distribution.

Double-pass experiments also provide empirical evidence for between-trial variation in drift rate (Ratcliff et al., 2018). In these experiments, the same perceptual stimuli are presented at two separate trials during an experimental session. Typically there is strong agreement on performance of these 2 trials. But if the drift rates of participants only vary within-trial, then agreement on the two trials is no greater than chance, meaning that between-trial variability in drift rate is needed to capture the empirical data. Although, the experiment does not determine if the variability is random or systematic (Evans et al., in press).

There is also neurophysiological evidence for between-trial variability in drift rates (Ratcliff et al., 2009; Ratcliff et al., 2016). For example, researchers have sorted trials of perceptual discrimination tasks by whether they show high amplitude or low amplitude EEG. They then sorted these two groups of trials by early stimulus related or late decision related components of the EEG. They found that for the early component trials, there were no differences between the drift rates of the low amplitude and high amplitude signals. But for the late component trials there were differences between the drift rates between the low amplitude and high amplitude trials. Because more positive drift rates were obtained from data associated with more positive EEG amplitudes, then it can be concluded that single-trial amplitudes of the late EEG explain some of the trial-to-trial variability in decision-related evidence.

Evidence for variability in starting points

In simple RT tasks, where participants only need to make one response, starting-point is required to account for the possibility of premature sampling (Laming, 1968), where the evidence accumulation process begins before the onset of the stimulus. In simple RT, participants cannot know to start sampling evidence at the presentation of the stimulus because the stimulus detection is itself the evidence being accumulated (i.e., detection cannot be used to trigger the detection process). In choice RT tasks, however, stimulus onset provides the signal to start sampling evidence that discriminates between different types of response options. One may even find that the amount of variability in starting points is smaller in choice paradigms than detection paradigms.

Pre-mature sampling is the theoretical motivation for starting point variability, but in relation to data, starting point variability is often included in sequential sampling models to explain the empirical finding that error responses can be faster than correct responses (Laming, 1968; Smith & Vickers, 1988; Ratcliff & Rouder, 1998; Ratcliff et al., 1999). When considering neurophysiological evidence, we find that differences in starting points across trials may be produced by sequential effects (for example Fecteau & Munoz, 2003).

Variability in non-decision time

Researchers who use the DDM typically assume that non-decision time is variable across trials. For example, Ratcliff et al. (2004) found that including non-decision time variability allowed the DDM to better fit the .1 quantile of RT distributions from a lexical decision task (Ratcliff, 2002; Ratcliff & Smith, 2004). Theoretically, non-decision time variability is plausible because motor response times could have small variability or stimuli may not be perceptually encoded as efficiently on each trial.

Issues with assuming between-trial variability

It is likely that some of the between-trial variability in the core parameters in most sequential sampling models is systematic and some is random (Evans et al., in press). A key job for theory is to sort out the systematic variance and explain it. Despite this, the current strategy in RT modeling is to assume all the between-trial variability is random, characterized by some probability distribution, such as a uniform or a Gaussian. But, there are several issues with the assumptions of random between-trial variability.

First, by adding these between trial sources of noise, sequential sampling models have become more and more complex in order to fit RT data. Models such as the DDM (Ratcliff & Tuerlinckx, 2002) assume between-trial variability in starting point, non-decision time, and drift rate in addition to within-trail variability in drift rate. The problem is that these extra between-trial variability parameters do not always help. Sometimes fixing the values of certain between-trial variability parameters can improve estimation of the more psychologically interesting parameters (Lerche & Voss, 2016). It is relatively difficult to accurately estimate between trial variability parameters (Boehm et al., 2018) and researchers have shown that models that have no between-trial variability are able to detect experimental effects better than the DDM with all between-trial variability parameters (van Ravenzwaaij et al., 2017).

Second, the between-trial variability assumptions are not part of the *process* model that explains how the RT data are generated. The core process of sequential sampling models is the integration of evidence to a threshold, which explains how someone makes a decision, but the between-trial variability assumptions provide no additional explanation of the decision process and are simply added to help the models fit data. That is not to say there is no evidence for decision parameters being different across trials, but that assuming between-trial variability represented as some probability distribution does not do the job of good theory in distinguishing between systematic and random variability.

Finally, given the recent claims of Jones and Dzhafarov (2014) – that certain between-trial drift variability assumptions can allow infinite flexibility of sequential sampling models (but see Heathcote et al., 2014; Smith et al., 2014) – it is important to explore and test models that do not make these assumptions.

Given these issues with the assumptions of between-trial variability, we test the necessity of the assumptions by evaluating the performance of a sequential sampling model that drops the between-trial sources of noise. We use the racing diffusion model (RDM), which relies on within-trial variability in drift rate instead of between-trial variability in drift rate. In the past, versions of racing diffusion models have been fit to choice RT data (Leite & Ratcliff, 2010; Teodorescu & Usher, 2013), but these models were applied using extensive Monte Carlo simulation methods. Moreover, racing Ornstein-Uhlenbeck diffusion processes have been used in more complex models that account for choice RT and confidence judgements (Ratcliff, 2009; Ratcliff & Starns, 2013). Trish Van Zandt developed closed form equations for the likelihood of racing diffusions to account for performance in the stop signal paradigm (Logan et al., 2014). By dropping the stop signal component we can apply a more general model of N-choice decision making, for example see Osth and Farrell (2019) and Turner (2019).

In this paper we fit the RDM to several previously published empirical data sets demonstrating that it can account for the full distribution of correct and error RTs as well as accuracy rates. We also fit the Linear Ballistic Accumulator (LBA; Brown and Heathcote, 2008) to the same data and compare how the RDM model performs relative to the LBA, which assumes drift rates vary between-trials rather than within-trials. We then assess the correspondence between the key RDM and LBA parameters to understand the mimicry between models that assume within-trial variability instead of between-trial variability. Before testing the RDM with model fitting, we first discuss the details of the RDM and how it accounts for

benchmark RT phenomena without between-trial variability parameters.

The racing diffusion model

The RDM represents the process of choosing between N alternatives (1, 2, 3, ..., N) using N racing evidence accumulators. In Fig. 2, we illustrate a two-choice RDM that represents a decision between 'left' and 'right' – perhaps in a task where participants need to determine if dots on a screen are moving to the left or to the right. The accumulators begin with an initial amount of evidence (k) that increases at a noisy rate with mean v toward the response threshold b. The evidence accumulation process stops when the first accumulator reaches its respective threshold. The winning accumulator determines the overt response and the time taken to reach the response threshold determines the decision time. To compute RTs, a constant time T_{er}, which is the parameter that represents non-decision processing time, is added to the decision time. Non-decision processes are often thought to represent both perceptual encoding and preparing motor responding (i.e., [t]ime for [e]ncoding and [r]esponding).

The RDM *can* have two sources of variability, one source of between-trial variability and one source of within-trial variability. The starting points of evidence accumulation are drawn from a uniform distribution between the interval of 0 and A. The second source of noise is the within-trial variation of drift rate, or change in evidence X. This noise

is mathematically described by the following stochastic differential equation (the Wiener diffusion process):

$$dX(t) = v \cdot dt + s \cdot dW(t), \quad (1)$$

where $s \cdot dW(t)$ represents the Wiener noise process with mean 0 and variance $s^2 \cdot dt$.

If the model only has within-trial drift rate variability and no between-trial starting point variability, then the Wald distribution (Wald, 1947) describes the first-passage time distribution for a single accumulator with positive drift rate toward a positive response threshold. The Wald distribution (Wald, 1947) is often used to explain simple RT data, where participants only need to make one type of response (Heathcote, 2004; Ratcliff & Van Dongen, 2011; Ratcliff & Strayer, 2014; Ratcliff, 2015; Schwarz, 2001; Heathcote, 2004; Anders et al., 2016; Tillman et al., 2017). The Wald probability density function is:

$$f_i(t|b, v) = b(2\pi t^3)^{-\frac{1}{2}} \exp\left[-\frac{1}{2t}(vt - b)^2\right], \quad (2)$$

where b and v are the response threshold and drift rate, respectively. When the model includes non-decision time, response times are described by a shifted-Wald distribution, with a shift equal to the non-decision time.

The RDM can also include between-trial variability in starting point, but as we show in the subsequent section, it is not required for fast errors. Starting point variability also accounts for when participants are uncertain about when to start sampling evidence (Laming, 1968). On some trials, participants may begin sampling before the trial begins
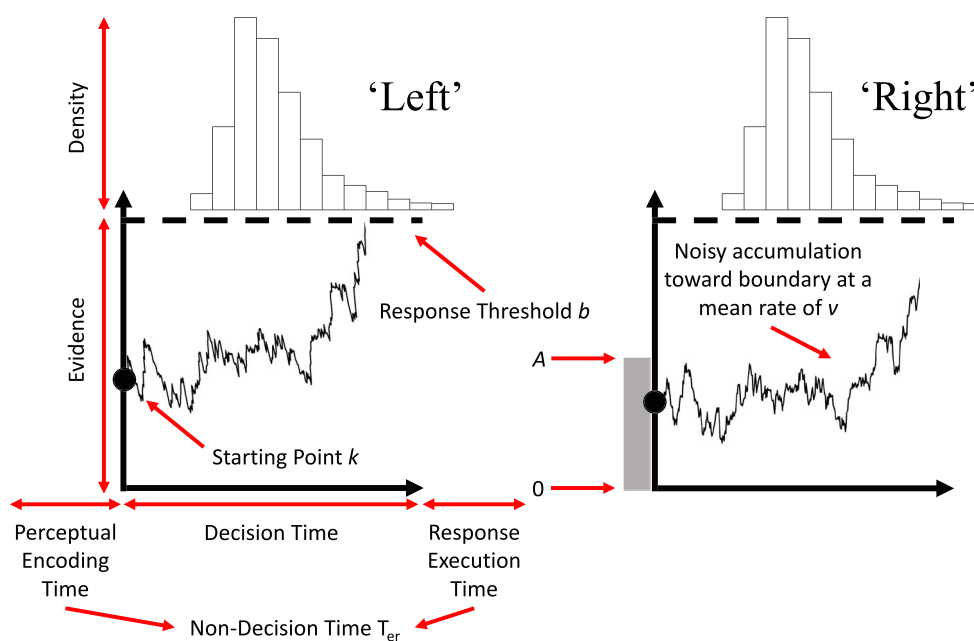

Fig. 2 The Racing Diffusion Model (RDM) and its parameter values: response boundary (b), mean drift rate (v), between trial variability in starting point (A), and non-decision time (T_{er})

resulting in a higher starting point. $\frac{1}{A}$ is the probability density function of the uniform starting point distribution. The probability density function of finishing times for the RDM with within-trial variability in drift rate and between-trial variability in starting point can be computed analytically (Logan et al., 2014). Using $\phi(x)$ and $\Phi(x)$ to represent the density and cumulative distribution functions of the standard normal distribution, respectively, and with

$$\alpha = \frac{(b - A - tv)}{\sqrt{t}} \tag{3}$$

and

$$\beta = \frac{(b - tv)}{\sqrt{t}} \tag{4}$$

then:

$$g_i(t|b, v, A) = \frac{1}{A}\left[-v\Phi(\alpha) + \frac{1}{\sqrt{t}}\phi(\alpha) + v\Phi(\beta) - \frac{1}{\sqrt{t}}\phi(\beta)\right] \tag{5}$$

for $v > 0$ and $A > 0$, and the Wiener process standard deviation is assumed to be 1. If $v = 0$, then

$$g_i(t|b, A) = \frac{1}{A}\left[\frac{1}{\sqrt{t}}\phi(\alpha) - \frac{1}{\sqrt{t}}\phi(\beta)\right], \tag{6}$$

and if $A = 0$ then the associated probability density function is Eq. 2. The cumulative distribution function is presented in the Appendices A, B and C.

In the RDM, there is a race between N accumulators and we want to know the time taken for the first accumulator to reach the threshold. The probability of the first accumulator finishing from all accumulators is given by the following defective probability density function:

$$g_i(t) = f_i(t)\prod_{j \neq i}(1 - F_j(t)), \tag{7}$$

where $f_i(t)$ and $F_j(t)$ are the probability density and cumulative density functions of first-passage time distributions for the ith and j^{th} accumulators.[1] Substituting Eqs. 2, 5, or 6 for the probability density distributions in Eq. 7 allows us to generate likelihood functions to fit the RDM model to data.

Slow errors and fast errors from race models

Multi-accumulator models predict slow errors by virtue of the race architecture, where the error accumulator is slower than the correct accumulator on average (Smith & Vickers, 1988; Townsend & Ashby, 1983). The LBA accounts for fast errors by including both starting point variability and between-trial variability. In a speed emphasized condition,

participants set their response thresholds close to the top of the starting point distribution. If the accumulator corresponding to the error response samples near the top of the starting point distribution and the correct accumulator samples closer to the bottom, then there is little time for the correct accumulator to overtake and win the race. Thus, error responses in this situation occur because the accumulator started close to the threshold, which also results in relatively fast responses.

In Fig. 3, we present a simulation to demonstrate how the RDM predicts fast and slow error RTs relative to correct RT. We generated 10, 000 trials for a single participant for a range of different threshold values. The threshold, b, was systematically varied from 0 to 2. The other parameters were fixed to the following values: $A = 1$ (when included), $V_{Correct} = 4$, $V_{Error} = .5$, $T_{er} = 0.2$. Due to the race architecture, the RDM predicts slow errors when the error drift rate is slower than the correct drift rate and the response thresholds are high enough to negate early terminating errors due to within-trial noise. However, when the response thresholds are close to the starting point, the RDM predicts fast errors due to the within-trial drift rate variability.[2] So the RDM can account for fast and slow errors with only *one* source of noise. Fast errors in the RDM with only within-trial drift rate variability are at most 10 ms faster than the mean correct RT (left panel of Fig. 3). With starting point variability, fast errors from the RDM get up to 40 ms faster than the mean correct RT (right panel of Fig. 3). The lowest thresholds considered produced RTs of 211 ms and the largest thresholds produced RTs of 770 ms, which covers a range that we would expect to see in empirical data. Although, a survey of papers where enough RT data is reported suggests that fast errors are at least 40 ms faster than correct RTs (e.g., Ratcliff & Rouder, 2000; Logan et al., 2014).

Fitting empirical data

In this section, we fit the RDM to empirical data sets from Ratcliff and Rouder (1998), Ratcliff and Smith (2004), Rae et al. (2014), and Van Maanen et al. (2012). These data sets were chosen because the conclusions reported in the original papers included effects on drift rates, thresholds, or both these parameters. In a later section, we also fit a data set with non-decision time effects, but the analysis is used to investigate differences in non-decision time estimates between popular sequential sampling models and the RDM. We also fit the LBA model because it is an

---

[1]The distribution is defective because it is normalized to the probability of its associated response.

[2]The DDM also contains within-trial drift rate variability, but because evidence for one response counts against evidence for the other response, only having within-trial noise leads to the model predicting equally fast correct and error response time distributions.
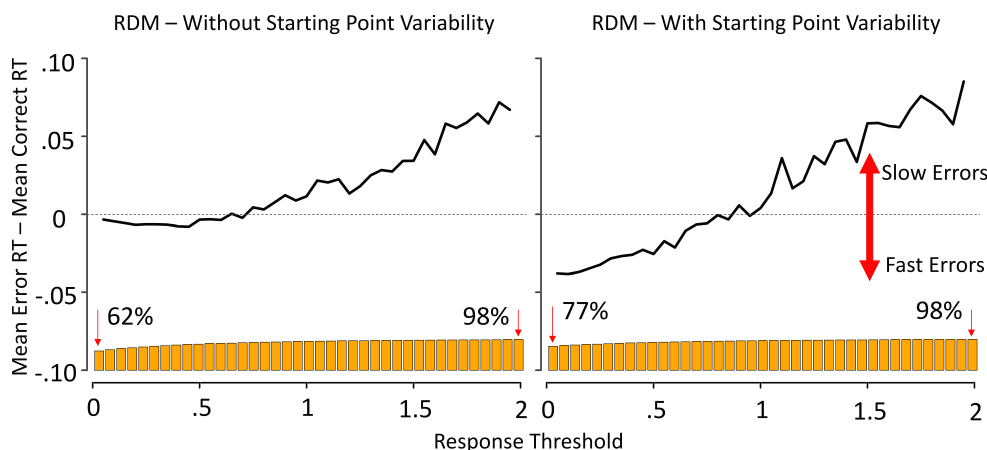
Fig. 3 Results of the RDM simulation. We plot error response times minus correct response times as a function of response threshold. Y-axis values below the dashed line represent fast errors relative to correct and values above the line represent slow errors. Barplots along the x-axis show the accuracy rates for each simulated data set

established model in the field making it a useful benchmark. Furthermore, the RDM and LBA only differ in their assumption about whether to include within-trial variability or between-trial variability in drift rate, and so, comparing the two will highlight differences between within-trial and between-trial drift rate assumptions.

We adapted R code provided with Logan et al. (2014) and hierarchical Bayesian methods to fit the models to data. The parameters of the RDM that we estimate are v, B, A, and T_{er}, where s is fixed to 1 unless mentioned otherwise. For all fits reported, instead of estimating b, we estimate the parameter B, which represents the difference between b and A. This prevented any starting points from being sampled above the response threshold. All other parameter variations will be discussed on a case-by-case basis in the subsequent sections. More details of our fitting procedure are presented in the Appendices B and C.

Rouder et al. 1998

We used experiment 1 in Ratcliff and Rouder (1998), which was originally fit with the DDM. In this experiment, participants made brightness discriminations of pixel arrays displayed on a computer screen. Specifically, they were asked to determine whether the arrays were of high or low brightness. Like the original study, we fixed all parameters across all conditions except for response thresholds and drift rate. There were 33 levels of brightness and we collapsed these into seven, as most brightness levels contained homogenous RT and accuracy rates (cf. Donkin et al., 2009a). We allowed drift rate to vary across the seven brightness conditions. The experiment included blocks with an accuracy emphasis and blocks with a speed emphasis. In the accuracy emphasis condition, participants were told to respond as accurately as possible. In the speed emphasis

condition, participants were told to respond as quickly as possible. Speed or accuracy emphasis are assumed to influence the response thresholds of participants (but see Rae et al., 2014).

In Fig. 4, for each participant, we plot the empirical mean RTs against response probabilities for the 7 collapsed stimulus conditions and speed/accuracy conditions. The response probability axis shows correct responses to the right of the .5 point and errors are to the left of .5. This means that each correct response to the right has a corresponding error response to the left. In both the RDM and LBA panels, the top points show data from the accuracy condition and the bottom points show data from the speed condition.

In the speed condition, RTs were below 400 ms across the entire accuracy range. The blue points in the figure, which are posterior predictives generated from 100 samples from the subject level joint posterior distributions, show that both the RDM (top panel) and the LBA (bottom panel) capture this flat function. In the accuracy condition, RTs are more than double the speed condition and both models predict this trend. However, RTs also vary from 400 ms to 900 ms across response probability and both models appear to significantly miss this trend for participant KR, especially in the middle accuracy range.

The empirical data contain both fast and slow error RTs relative to correct RTs and the RDM could capture these trends. For instance, in brightness condition 6, for participant KR, the empirical mean correct RT is 542 ms in the accuracy condition and 308 ms in the speed condition. Error RT is 568 ms in the accuracy condition and 292 ms in the speed condition. If we inspect the posterior predictive mean RTs generated from the mean of the posterior distribution of the RDM model, we find that in brightness condition 6, for participant KR, correct
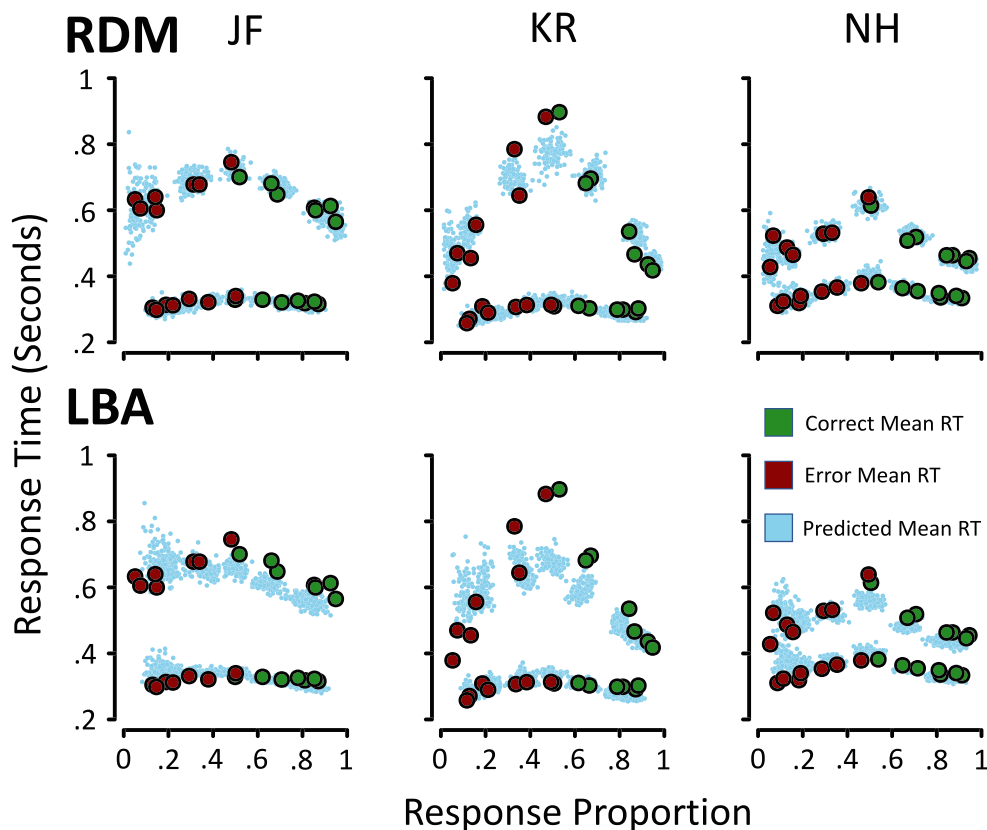
Fig. 4 Mean RT (red and green dots) and posterior predicted mean RT (blue dots) from the RDM model (top panel 'A') and LBA model (bottom panel 'B') for subjects JF, KR, and NH from Ratcliff & Rouder ([1998]). The upper and lower data points are for accuracy and speed emphasis conditions, respectively. The left side of each plot represents error responses and the right side of each plot represents correct responses to the same stimuli

RT is 539 ms in the accuracy condition and 319 ms in the speed condition. Moreover, error RT is 566 ms in the accuracy condition and 289 ms in the speed condition. Overall, the RDM captures many of the complex empirical trends present in the benchmark (Ratcliff & Rouder, [1998]) data set.

The best fitting parameters are shown in Fig. 5. Both the RDM and LBA have higher thresholds in the accuracy condition compared to the speed condition. Moreover, the correct drift rates increased and error drift rates decreased as a function of stimulus difficulty, where 1 and 7 were the easiest conditions and 4 was the hardest condition.

Ratcliff et al. ([2004])

Ratcliff and Smith ([2004]) fit the DDM to data from a lexical decision task in which participants identified whether a string of letters was a word or a non-word. In experiment 1 from the paper, participants were presented with words that occurred often (high frequency), rarely (low frequency), or very rarely (very low frequency) in everyday language. Participants were also presented non-words that were generated from the high frequency, low frequency, or

very low frequency words. Drift rates varied as a function of how difficult the stimuli were to classify, such that, more difficult (i.e., lower frequency words) stimuli produced lower drift rates. We replicate their drift rate finding here with the RDM.

For both the RDM and LBA models, we allowed drift rate and non-decision time to vary over stimulus difficulty, which is a recommended parameterization for lexical-decision tasks (Donkin et al., [2009b]), but see also Tillman et al. ([2017]). To summarize the empirical RT distributions, we present five quantile estimates (.1, .3, .5, .7 and .9) in Fig. 6. The .1, .5, and .9 quantiles represent the leading edge, median, and upper tail of the distribution, respectively. These plots are defective cumulative distribution functions, meaning that RT is shown along the x-axis and the upper asymptote of the function represents the probability of that response. In other words, the proportion of correct (green) and incorrect (red) responses are along the y-axis.

Both models capture the accuracy rates and correct RT distributions well. The models mostly differ in their account of the error RTs. The RDM consistently under predicts the upper tail of the error RT distribution and the LBA consistently over predicts the upper tail. The over prediction
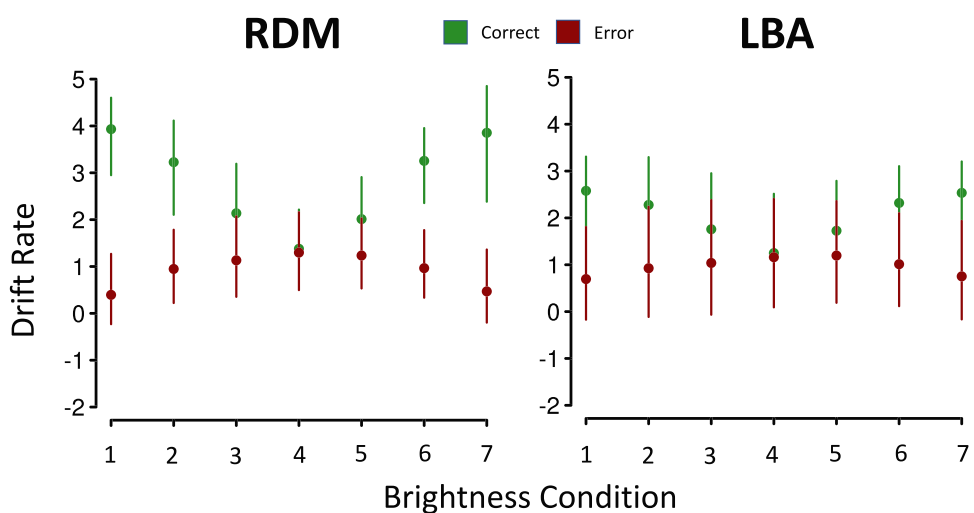
Model	B.Speed	B.Accuracy	A	T_{er}
RDM	.522 (.077, .878)	1.132 (.199, 1.451)	.107 (.006, .499)	.187 (.123, .230)
LBA	.257 (.019 , .724)	.656 (.073 , 1.113)	.313 (.054, .601)	.141 (.015, .554)

Fig. 5 The group-level mean posterior distributions for the RDM and LBA fits to the Rouder et al. data set. The top two panels show correct and error drift rates for the RDM and LBA, with the median of the posterior and 95% credible intervals. In the bottom panel, we table the median *B*, *A*, and *T er* parameters with the 95% credible intervals in parenthesis

is larger and more pervasive than the under prediction, however. The misses are likely due to the difficulty in estimating reliable parameters from relatively few data points, which is the case for the .9 quantile of the error RT

distribution. Importantly, both models capture the key trend of RTs slowing as a function of stimulus difficulty.

The best fitting parameters are shown in Fig. 7. The correct drift rates decreased and the error drift rates
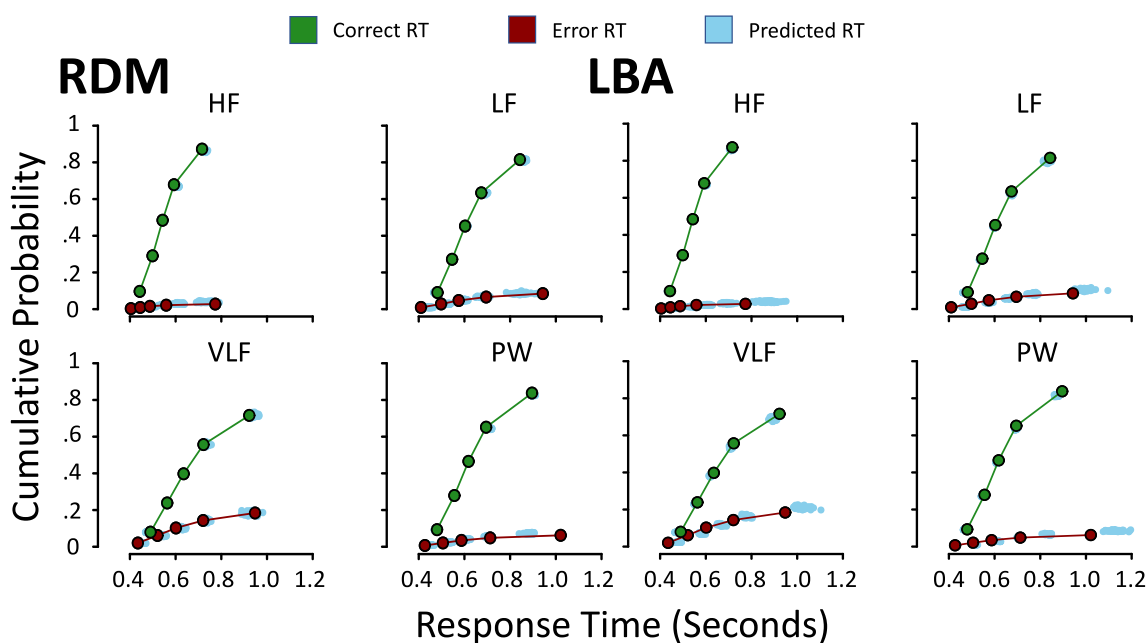

Fig. 6 Empirical and predictive defective cumulative distribution plots of the .1, .3, .5, .7, and .9 RT quantile for each stimulus condition from experiment 1 of Ratcliff et al. (2004) – high frequency (HF), low-frequency (LF), very-low-frequency (VLF), and pseudo-words (PW). Posterior predictives are shown for both the RDM (left) and LBA

(right) model. The green (correct) and red (error) dots represent the empirical data and the blue dots represent the predicted data. Note the predicted data consists of 100 separate data sets generated from the subject-level joint posterior distribution
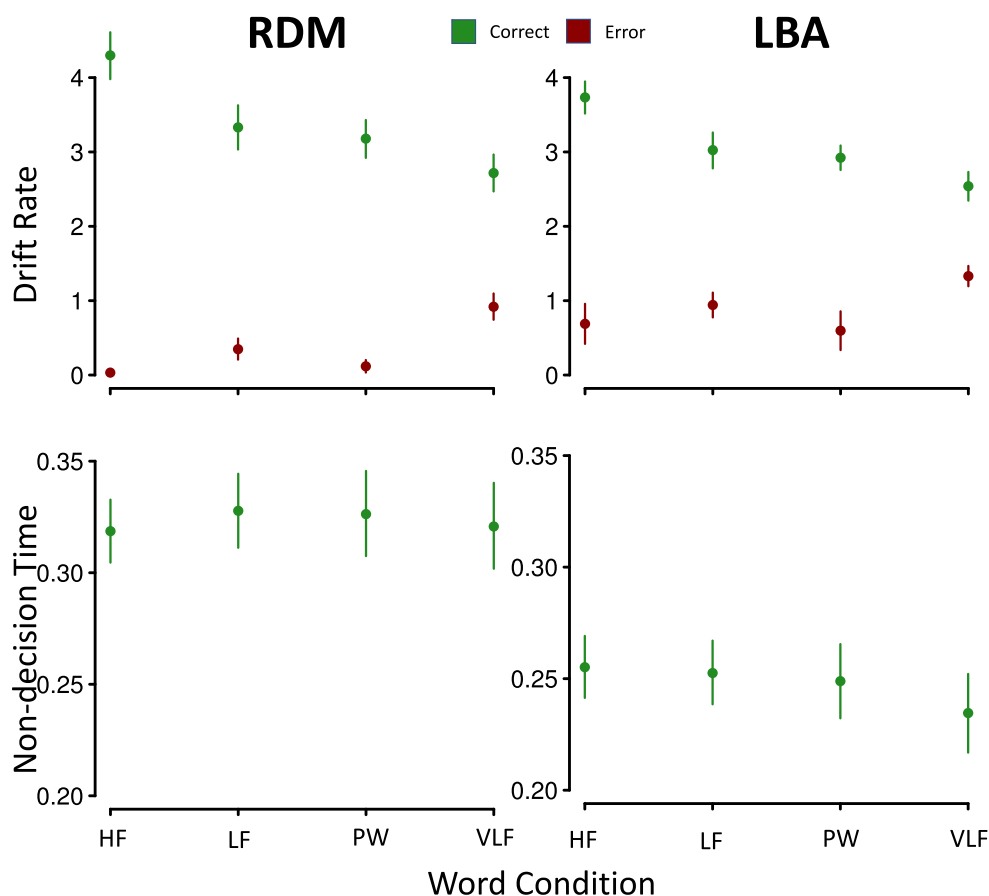
Model	B	A
RDM	.973 (.861, 1.084)	.100 (.005, .283)
LBA	.712 (.535, .846)	.741 (.596, .865)

Fig. 7 The group-level mean posterior distributions for the RDM and LBA fits to the Ratcliff and Rouder (1998) data set. The top two panels show correct and error drift rates for the RDM and LBA, with the median of the posterior and 95% credible intervals. The middle panel shows the non-decision times for the RDM and LBA. In the bottom panel, we table the median *B* and *A* parameters with the 95% credible intervals in parenthesis

increased as a function of stimulus difficulty. The effects of stimulus difficulty on non-decision were quite small (¡5 ms) in the RDM and the RDM also estimated non-decision times more than 70 ms higher than the LBA.

Rae et al. (2014)

The experiment from Rae et al. (2014) was like the experiment from Ratcliff and Rouder (1998). Participants had to choose whether the stimulus was predominately light or dark. The authors manipulated speed and accuracy emphasis on alternating blocks of trials. Despite a similar task design to Ratcliff and Rouder (1998), the authors found that speed instructions, relative to accuracy instructions, not only lead to a decrease in thresholds, but a decrease in the difference between correct and error drift rates. The difference between correct and error drift rates in

accumulator models can be interpreted as discrimination between evidence for correct and error responses, and is comparable to the drift rate in the DDM.

We fit both the RDM and LBA model and allowed response thresholds, drift rates, and non-decision time to vary with speed/accuracy emphasis, which is in line with the original authors' best fitting model. We summarize the data in Fig. 8 with defective cumulative distribution functions. The RDM predicted the tails of the correct RT distribution well, but under predicted the means and leading edge (i.e., the .1 quantile). The LBA under predicted all quantiles for the correct RT distribution and both models over predicted the error RTs. In the empirical data, the accuracy condition did not contain slow errors, which maybe reflective of the small change in accuracy rates across speed and accuracy. Yet, the fast errors that were present in both conditions were captured by the models.
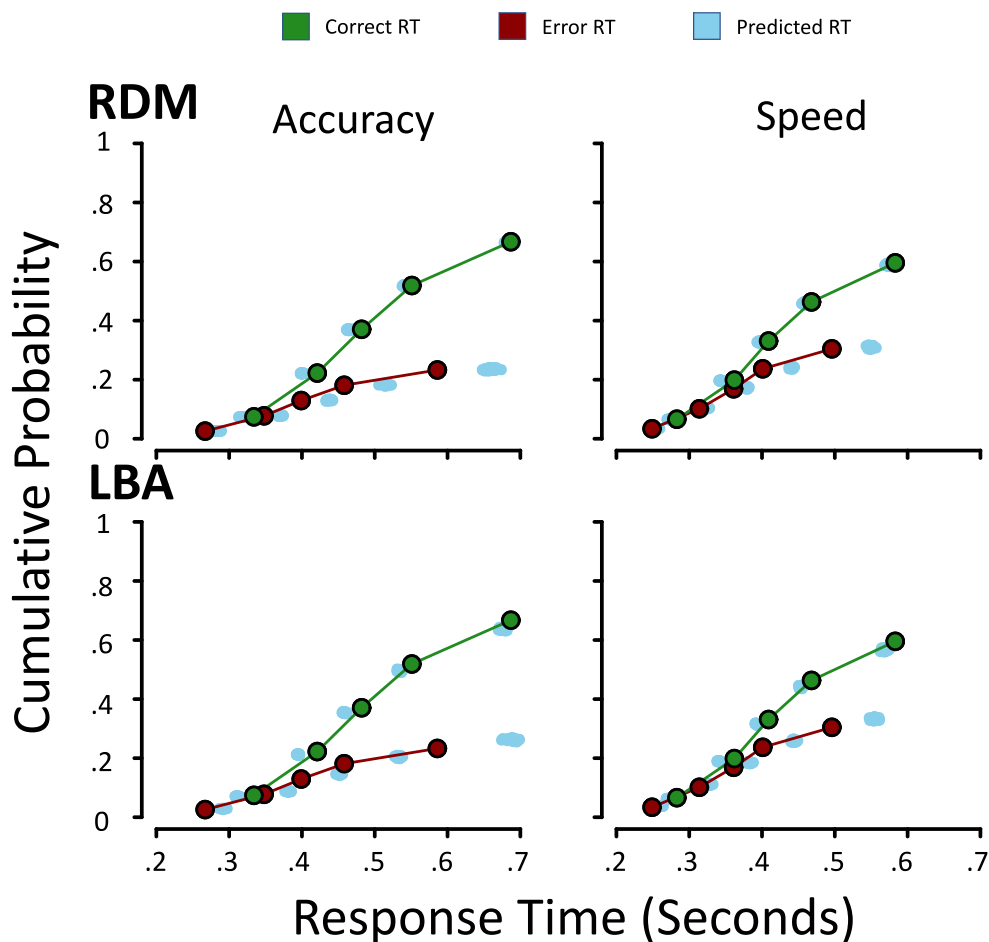
Fig. 8 Empirical and predictive defective cumulative distribution plots of the .1, .3, .5, .7, and .9 RT quantile for speed and accuracy conditions from experiment 1 of Rae et al. (2014). Posterior predictives are shown for both the RDM (top) and LBA (bottom) model. The green (correct) and red (error) dots represent the empirical data and the blue dots represent the predicted data. The predicted data consists of 100 separate data sets generated from the subject-level joint posterior distribution

In terms of best fitting parameters, which we present in Table 1, the difference between correct and error drift rates were larger in the accuracy condition compared to the speed condition for both the RDM and LBA – replicating the original authors novel findings. The response thresholds were also higher in the accuracy condition for both models, which is reflected in the A parameter. Allowing A to vary over speed and accuracy was in line with the original author's parameterization and is analogous to a threshold manipulation because B was fixed across conditions.

Table 1 Median values from the group-level mean posterior from the RDM and LBA fit to Rae et al. (2014). 95% credible intervals are presented in parenthesis

Model	Condition	v Correct	v Error	A	B	T_{er}
RDM	Speed/Light	4.55 (4.17,4.93)	3.08 (2.77,3.38)	0.39 (0.03,0.71)	1.04 (0.87,1.18)	0.134 (0.091,.160)
	Speed/Dark	4.01 (3.68,4.52)	3.44 (3.12,3.76)	0.28 (0.02,0.55)	1.04 (0.87,1.19)	0.134 (0.091,.160)
	Accuracy/Light	4.27 (3.80,4.73)	2.58 (2.19,2.97)	0.32 (0.02,0.82)	1.04 (0.87,1.18)	0.162 (0.134,.186)
	Accuracy/Dark	4.17 (3.73,4.59)	2.56 (2.15,2.98)	0.49 (0.05,0.89)	1.04 (0.87,1.19)	0.162 (0.134,.186)
LBA	Speed/Light	3.77 (3.49,4.05)	3.29 (3.01,3.57)	0.74 (0.33,0.94)	0.96 (0.76,1.12)	0.010 (0.001,0.038)
	Speed/Dark	3.83 (3.51,4.15)	3.22 (3.00,3.44)	0.86 (0.28,1.14)	0.97 (0.75,1.12)	0.010 (0.001,0.038)
	Accuracy/Light	3.83 (3.61,4.05)	2.72 (2.50,2.96)	1.07 (0.36,1.42)	0.96 (0.76,1.12)	0.025 (0.001,0.068)
	Accuracy/Dark	3.64 (3.38,3.89)	2.82 (2.61,3.03)	0.98 (0.47,1.24)	0.97 (0.75,1.12)	0.025 (0.001,0.068)

Non-decision time is in milliseconds

Van Maanen et al. (2012)

Hick's law is a benchmark phenomenon for speeded N choice experiments (Hick, 1952; Hyman, 1953). In general, it states that the mean RT increases linearly with the logarithm of the number of choice alternatives. If no parameters can vary over set size, then multi-accumulator models predict that the mean RT decreases and the error rate increases with the number of response options (cf. Logan, 1988). In racing accumulator models, the mean RT decreases as a function of N because the observed RT distribution is made up of the minima of finishing time distributions from N accumulators. With more runners in the race there is a higher probability of sampling a faster winner (i.e., statistical facilitation; Raab, 1962).

Multi-accumulator models can account for Hick's law several ways. There could be normalization in evidence input to the model (e.g., Brown et al., 2008; Hawkins et al., 2012). For example, normalization in the RDM occurs when we constrain all the drift rates for all accumulators by having them sum to 1. It is also the case that multi-accumulator models can predict Hick's law by allowing drift rates to vary over set size (Van Maanen et al., 2012; Logan et al., 2014) or by exponentially increasing response thresholds with set size (Usher et al., 2002). The latter explanation may be contested based on neurophysiology. For instance, if we grant that accumulator models are represented by populations of pre-saccadic movement neurons in brain regions such as frontal-eye-field (Teller, 1984; Purcell et al., 2010; Purcell et al., 2012), then we should expect the peak firing rates of frontal-eye-field movement cells to increase with set size. However, evidence from single cell recordings suggest that the movement cells reach a fixed firing rate immediately before initiating saccade responses in visual search tasks (Hanes & Schall, 1996) and the firing rates are invariant across set sizes (Woodman et al., 2008)

Variants of racing diffusion models have accounted for Hick's law in simulation (Leite & Ratcliff, 2010) and in the stop signal paradigm (Logan et al., 2014). To test whether the RDM can be fit to N-choice data and can capture the log-linear relationship between RT and response alternatives, we fit a previously published Hick's law data set from Van Maanen et al. (2012). Rather than commit to any explanation of Hick's law here, we follow the parameterization of Van Maanen et al., and allow A, B, v and T_{er} to vary over set size for both the RDM and LBA.

Specifically, we fit the "spaced" condition from experiment 2, which was a motion dots task that required 5 participants to choose between 3, 5, 7, or 9 decision alternatives. In addition to Hick's law, we are testing whether the RDM can account for both the full distribution of RTs and accuracy rates associated with making N choice decisions.

In Fig. 9, both the RDM and LBA can account for how accuracy, mean RT, and correct and error RT distributions change across 3, 5, 7, and 9 response alternatives. In general, both models capture the correct distributions well and only miss on the .9 quantile for error RTs. In set size 3 for the LBA, the posterior predictions for the last .9 quantile are spread out to 2.2 seconds. However, there were only 37 error RTs in that condition across all participants, and so, accurate parameter estimation is difficult. In the top right of both the set size 3 panels is mean RT plotted as a function of the logarithm of response set size. In green is the empirical data, which shows a linear increase across log set size demonstrating Hick's law. Both the RDM and LBA could predict this trend.

In Fig. 10, the RDM appears to capture changes in the data and Hick's law with a decrease in correct drift rate. Unlike our other fits, the RDM estimated non-decision times lower than the LBA, but the estimates were quite varied as shown by the width of the 95% credible intervals.

Comparing models

To compare the LBA and RDM fits we used a model selection exercise to determine what model has a better account of the empirical data. The aim here is to determine if the RDM does as well as the LBA, which is currently a successful sequential sampling model of choice RT.

One way to choose between models is to assess their "usefulness". One metric of usefulness is how well a model predicts future data. A model's ability to predict future data reflects how well the model captures the important signal (i.e., parts of the data we are interested in) in the data that is common between the current data and future data. Models that do poorly are usually either too simple, meaning that do not capture the signal, or too complex, meaning they capture the signal and noise, where the noise is not the same in future data. We used the widely applicable information criterion (WAIC Watanabe, 2010) to assess the out-of-sample predictive accuracy of the LBA and RDM. The model with the lowest WAIC has the best out-of-sample predictive accuracy and we will select between models with this metric. More details about our model selection exercise are presented in Appendix C.

The main conclusion of our model comparison was that the RDM performs at least as well as the well-known LBA model. In Table 2 we show that the RDM is the preferred model for the Ratcliff and Smith (2004) and Rae et al. (2014) data sets and the LBA is the preferred model for the Ratcliff and Rouder (1998) and Van Maanen et al. (2012) data set. The columns showing the log-pointwise-predictive density (lppd), which is a measure of goodness of fit (without a complexity penalty), show that the winning models are preferred because they better fit
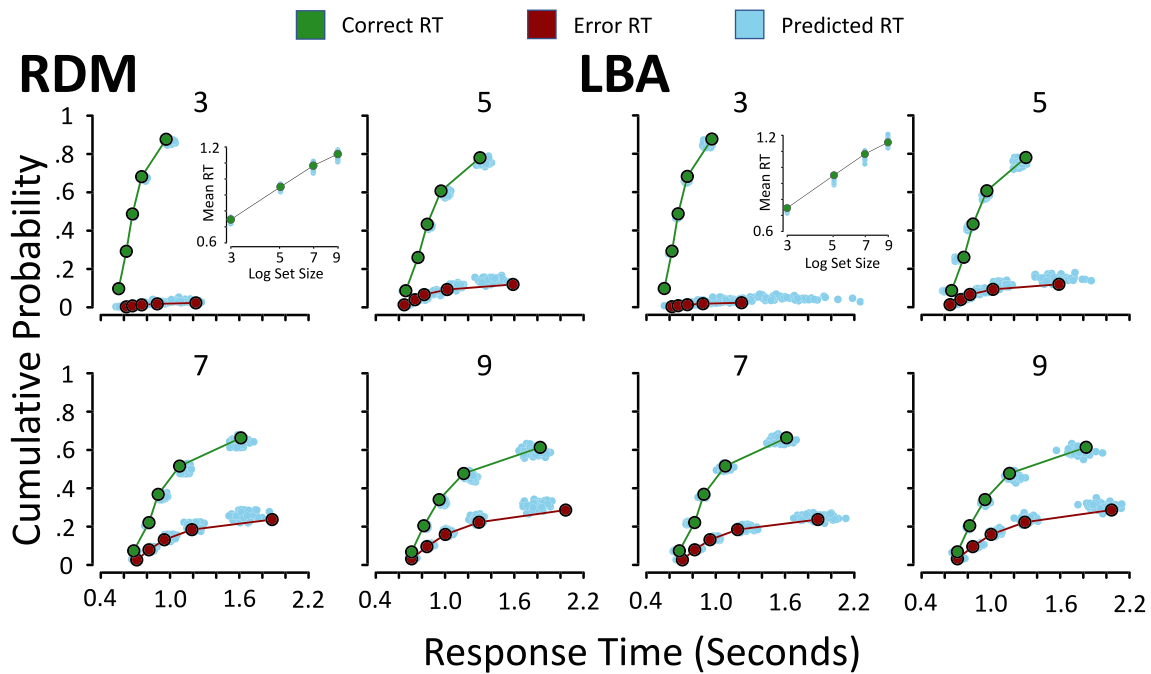
Fig. 9 Empirical and predictive defective cumulative distribution plots of the .1, .3, .5, .7, and .9 RT quantile for each stimulus condition from experiment 2 of Van Maanen et al. (2012) – labels 3, 5, 7, and 9 correspond to conditions that included 3, 5, 7, and 9 response alternatives. Posterior predictives are shown for both the RDM (left) and LBA (right) model. The green (correct) and red (error) dots represent the empirical data and the blue dots represent the predicted data. Note the predicted data consists of 100 separate data sets generated from the subject-level joint posterior distribution. Presented in the set size 3 panel of each model plot is the mean RT as a function of log set size, which demonstrates Hick's law. Both models capture the log-linear relationship between mean RT and set size
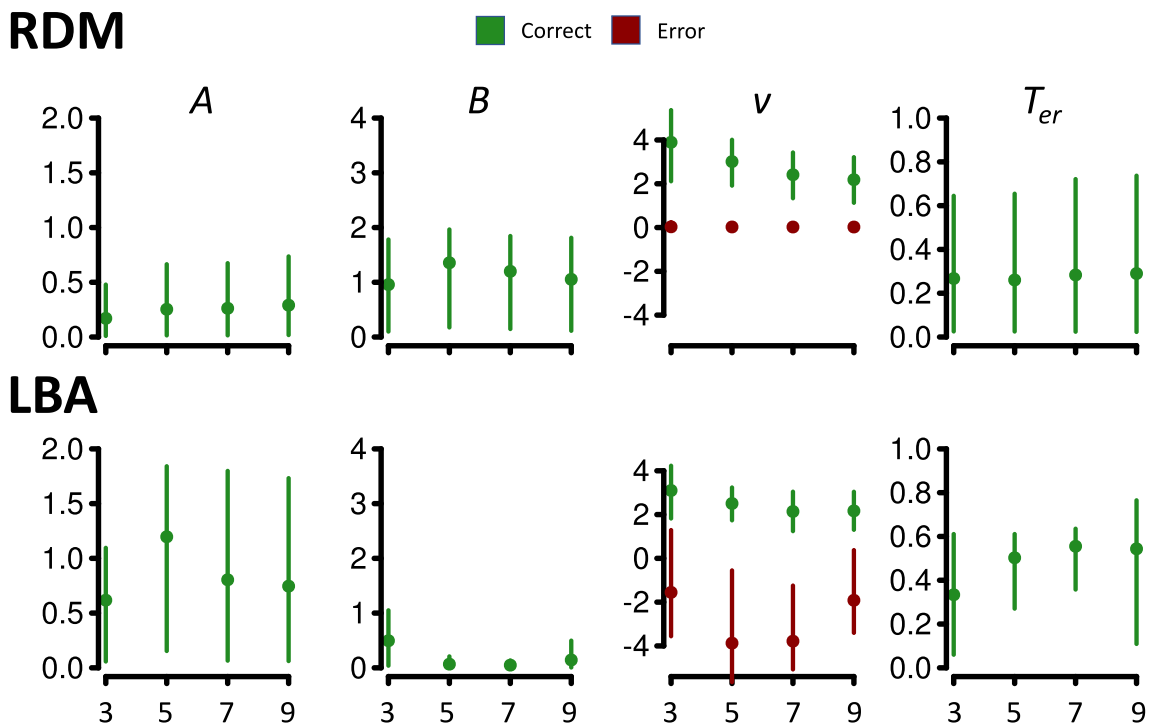

Fig. 10 The group-level mean posterior distributions for the RDM and LBA fits to the Van Maanen et al. (2012) data set. The top panels show median and 95% credible intervals for the A, B, v, and T_{er} parameters for the RDM. The bottom panel shows the median and 95% credible intervals for the A, B, v, and T_{er} parameters for the LBA

Table 2 WAIC model selection results

Data Set	LBA lppd	RDM lppd	LBA pD	RDM pD	WAIC Diff
Ratcliff et al. (1998)	1986	1953	61	71	−56
Ratcliff et al. (2004)	8924	9406	248	213	929
Rae et al. (2014)	16892	17695	538	582	1650
Van Maanen et al. (2012)	−4474	−4536	133	90	−168

WAIC column is the difference in WAIC between LBA and RDM. Positive values are in favor the RDM model

pD = Number of effective parameters

lppd = log pointwise predictive density (i.e., goodness of fit)

aspects of the data in question rather than simply being more parsimonious. In terms of parsimony, we found that the RDM was the simpler model for two of the four data sets. For some data sets, there does not appear to be complete agreement between model selection metrics and the plotted observed and predicted data. For example, in the Rouder et al. (1998) data set, the LBA model qualitatively has larger misses to the mean RT, especially for the subject KR. Despite this, the LBA model was preferred by WAIC. In Table 2 the goodness of fit, the number of effective parameters, and the WAIC values are presented, which can be used to understand this disagreement. From these values we can see that the LBA is selected due to the model having a better fit to the overall data (not just the mean RT) and because it is less complex in terms of the number of effective parameters.

Interim conclusion

So far, we have shown that the RDM model is able to account for benchmark RT phenomena, including full error and correct RT distributions, accuracy rates, N-choice performance, the speed-accuracy trade-off, and fast and slow errors relative to correct. We have also shown that the RDM can fit previously published empirical data at least as well as the LBA model. In what follows, we will assess the extent of correspondence between the RDM's and LBA's key parameters. Following this analysis, we explore

non-decision time estimates of DDM, LBA, as well as the RDM.

Cross-fitting analysis

The cross-fitting analysis determines the correspondence between the drift rate, response threshold and non-decision time parameters of the RDM and LBA models. The analysis involved generating RT and choice data from each model and systematically changing each of the parameters. If the model parameters correspond then systematically changing a parameter of one model, generating data from this model, and then fitting the simulated data with the other model will result in similar changes in the corresponding parameter. Like the method of Donkin et al. (2011), we simulated a single participant in a two-choice experiment with three conditions: easy, medium, and hard. There were 20,000 trials in each condition.

We generated data for both the RDM and LBA using the parameter values in Table 3. To generate data, we fixed parameter values at their default values (see Table 3) and systematically varied either drift rate, threshold, or non-decision time. Varied parameters took values from 50 equally spaced points with a range like those observed in practical applications. For each parameter set we generated 60,000 observations. For simulation purposes, v was defined as the difference between the correct and error drift rate. We

Table 3 Parameter values used to generate data sets used in the cross-fit simulation. Default values are parameter values used when fixed across simulations

Model		$v.Easy$	$v.Medium$	$v.Hard$	B	T_{er}	s	A
LBA	Min	2.1	2	1.8	.25	.2	.75	–
	Max	3.6	3.5	3.3	.5	.4	1.5	–
	Default	2.3	2.2	2	.5	.3	1	.5
RDM	Min	2.1	2	1.8	.25	.2	–	–
	Max	3.6	3.5	3.3	.5	.4	–	–
	Default	2.3	2.2	2	.5	.3	1	.5

calculated $v_{Correct}$ as $2+(v/2)$ and v_{Error} as $2-(v/2)$. With this parameterization, we just compare the correspondence between the v parameter of the LBA and RDM. Only v could vary over the three difficulty conditions.

For each simulated data set, we estimated posterior distributions via a Bayesian estimation routine. The mean values of the posterior distributions are plotted in Figs. 11 and 12. Both figures show the effect of changing drift rate, threshold, and non-decision time in the data generating model on the parameter values from the recovering model. A monotonically increasing function along the main diagonal suggests the analogous parameters in the RDM and LBA correspond. If the parameters of the generating model only selectively influence the corresponding parameters of the recovering model, then there will be vertical lines with no slope in the off-diagonal panels.

In Figs. 11 and 12, we can see that the main diagonals show the monotonically increasing function, suggesting the corresponding parameters of the RDM and LBA agree. We also observed that the RDM systematically recovered higher non-decision times than the LBA, and the LBA systematically recovered lower non-decision times than the RDM. We replicate this trend in a subsequent analysis and show how this can lead to the two models arriving at different conclusions about non-decision time effects.

In the top center panel of Fig. 11, we can see that increases in drift rate of the LBA can lead to increases in response threshold in the RDM. Furthermore, the center right panel shows that increases in threshold can cause increases in non-decision time. In the center left panel of Fig. 12, we show that increases in the RDM's threshold can lead to increases in the LBA's drift rate. In the center
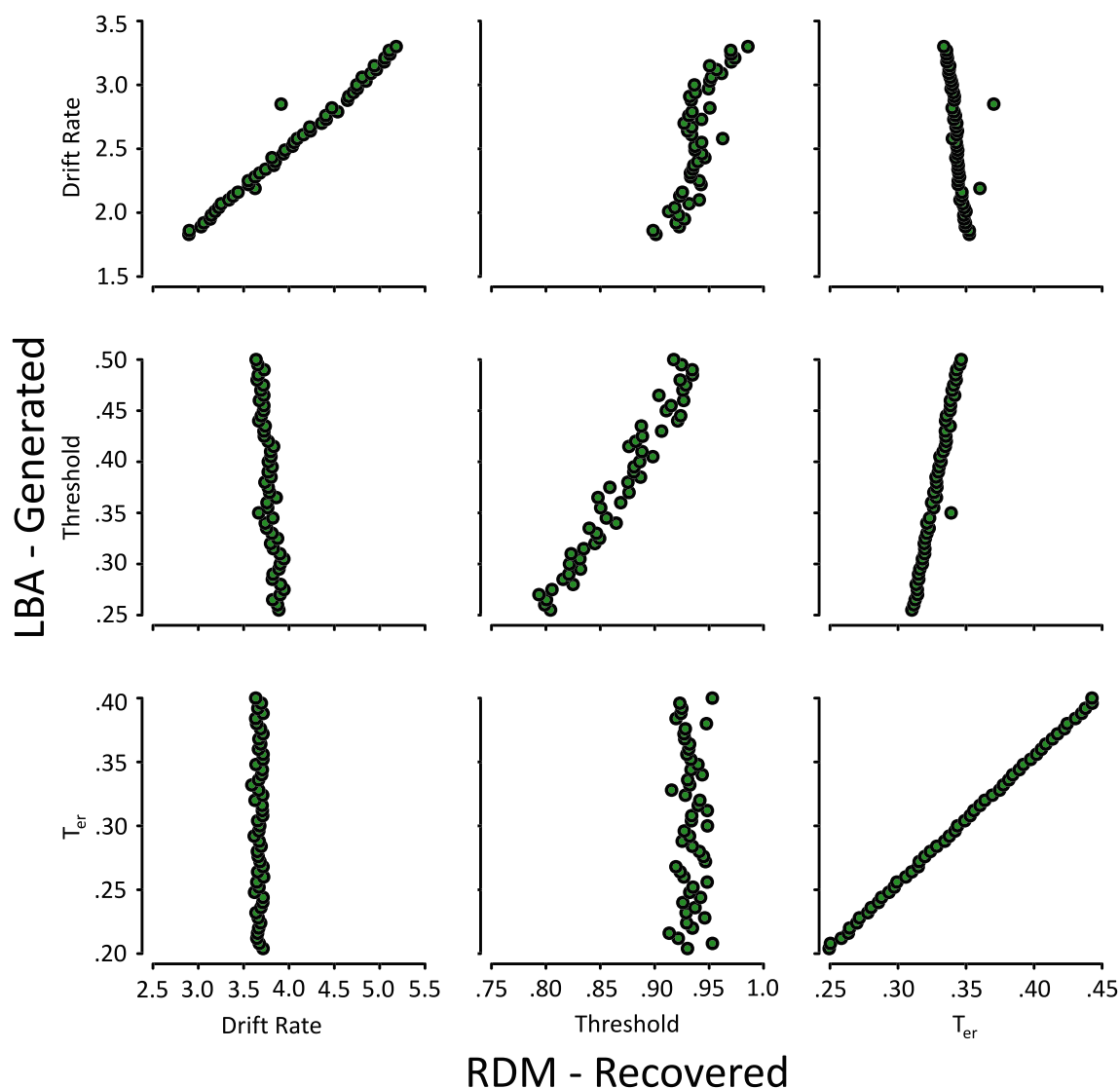

Fig. 11 Changes in parameters corresponding to response threshold, non-decision time, and drift rate in the RDM caused by systematic changing of LBA model parameters
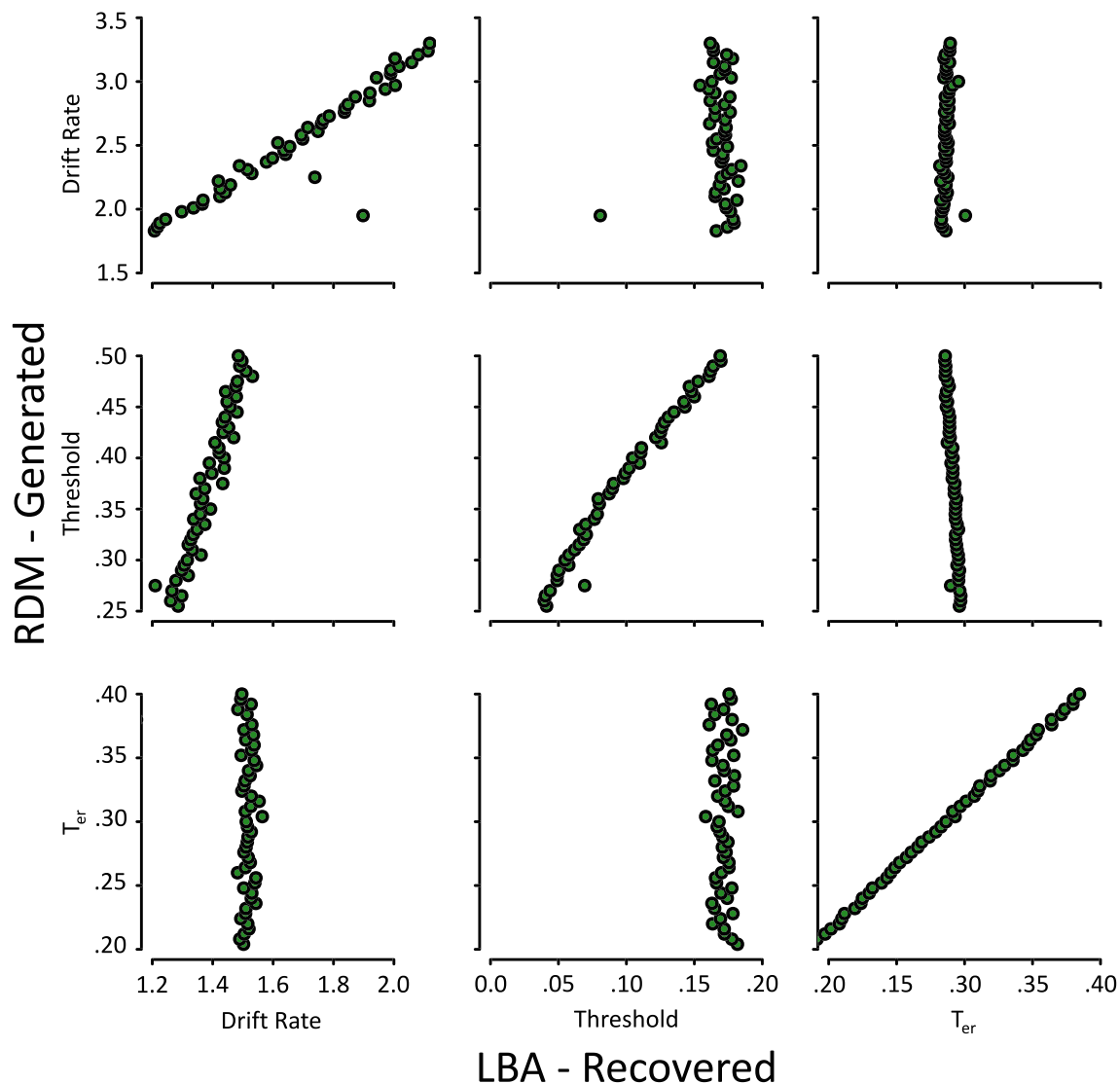
Fig. 12 Changes in parameters corresponding to response threshold, non-decision time, and drift rate in the LBA caused by systematic changing of RDM model parameters

left panel of Fig. 12, we see that increases in the RDM threshold can lead to increases in the LBA drift rate. Overall, and despite some small violations of selective influence we show that the key parameters of the two models agree.

Finally, we also varied between-trial drift rate variability of the LBA over 50 equally spaced values. Unlike the previous analysis, we fit the RDM to these data but only allowed within-trial drift rate variability to vary over the 50 data sets and fixed all other parameters at their default values. In Fig. 13, we can see that increases in between trial variability leads to increases in within-trial drift rate variability. The recovered values of within-trial variability had a smaller range than the generating values of between-trial variability. The RDM likely has a smaller range because noise in the model's decision processes is sampled from a Gaussian distribution at each time step where the LBA is a

single sample from the Gaussian distribution on each trial. Because the RDM samples at each time step rather than each trial, the variability in samples cumulates over time within a trial.

The magnitude of non-decision time

The results from our cross-fitting analysis suggest that the RDM systematically estimates higher non-decision times than the LBA. Because of the sparse literature on the reliability of non-decision time estimates in sequential sampling models, we fit the DDM, LBA, and RDM using hierarchical Bayesian methods to a previously published data set that found effects on non-decision time. van Ravenzwaaij et al. (2012) analyzed data from a motion dots task with the DDM and found that alcohol consumption
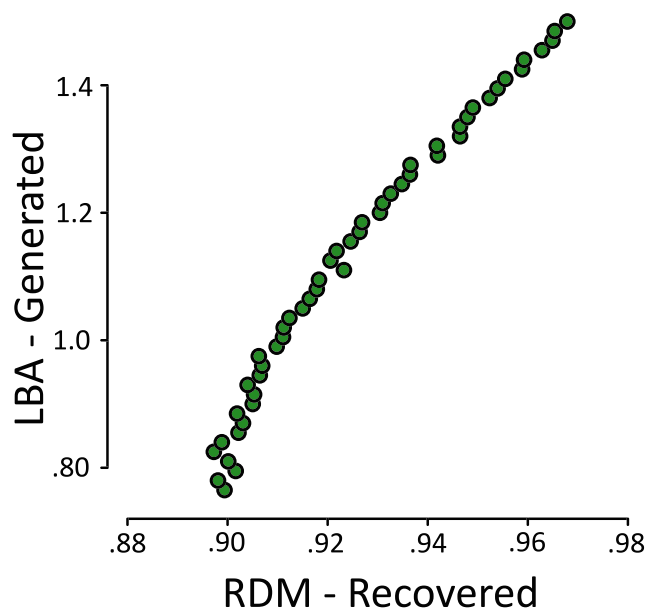
Fig. 13 Changes in the within-trial drift rate variability parameter of the RDM caused by systematic changing of the LBA between-trial drift rate variability parameter

Table 4 Median non-decision time values from the group-level mean posterior. 95% credible intervals are presented in parenthesis

Model	Placebo	Moderate	High
LBA	162 (125,192)	149 (103,182)	140 (117,161)
RDM	251 (245,256)	253 (246,260)	256 (247,264)
DDM	275 (269,281)	275 (270,280)	285 (276,295)

Non-decision time is in milliseconds

decrease non-decision time values. The authors concluded that alcohol impaired cognitive and motor/perceptual encoding capacity.

We calculated the non-decision time values from the median of the group-level posterior mean. As shown in Table 4, non-decision time from the DDM increased from 275 ms in placebo doses to 285 ms for high alcohol doses.[3] With the LBA, non-decision time was 162 ms for placebo doses and this decreased to 140 ms for high alcohol doses. Therefore, if researchers had used the LBA, they would conclude that alcohol consumption does not lead to motor/perceptual encoding deterioration and that alcohol doses reduce motor/perceptual encoding time. Finally, non-decision time from the RDM increased from 251 ms in placebo doses to 256 ms for high alcohol doses, and so, researchers would conclude that alcohol had practically no effect on non-decision processes.

To learn more about the differences in non-decision time estimates from the LBA and RDM we simulated the processes of both models 100 times. In Fig. 14, we present three simulations: the RDM with no start point variability, and the LBA both with and without start point variability. The results show that the RDM without start point variability has its fastest finishing times around 100 ms, but the LBA without starting point variability has its fastest finishing times around 200 ms. If we add a non-decision time value of 100 ms to both these models then the fastest RT for the RDM will be approximately 200 ms

and the LBA will be 300 ms. Thus, the LBA will need a smaller non-decision time value to produce comparable RTs to the RDM. The LBA can produce faster finishing times by increasing the starting-point variability, as shown in the bottom right panel, but this will also affect the accuracy rates of the model.

General discussion

Sequential sampling models account for both response times (RT) and accuracy rates in a range of different speeded decision making paradigms. These models assume decisions result from evidence accumulation to a response threshold. The models allow researchers to draw conclusions about speed of processing, caution, bias, and the time needed for perceptual encoding and motor responding. Yet, the success of these models has been dependent on auxiliary assumptions about the process components of the models varying from trial-to-trial. We argued that these between-trial variability assumptions are not founded on any process explanations and are simply added to help the models fit data. This can be problematic because the predictive ability of sequential sampling models is partly determined by distributional assumptions of between-trial variability.

To test the necessity of the between-trial variability components, we further developed a multi-accumulator race model – the racing diffusion model (RDM) – that dropped some of these assumptions and showed that it can still provide an accurate account of RT data. We showed that the RDM with only within-trial variation in drift rate can produce key response time phenomena, such as fast and slow errors relative to correct responses. The RDM was compared to the linear ballistic accumulator (LBA), which is a race model with between-trial variability in drift rates and starting points. The RDM replaced the between-trial variability in drift rate with within-trial variability in drift rate and was able to fit data as well as the LBA, accounting for both RT and accuracy rates in tasks involving lexical decisions, brightness discriminations, motion dots, and manipulations of response set size, stimulus difficulty and speed/accuracy emphasis.
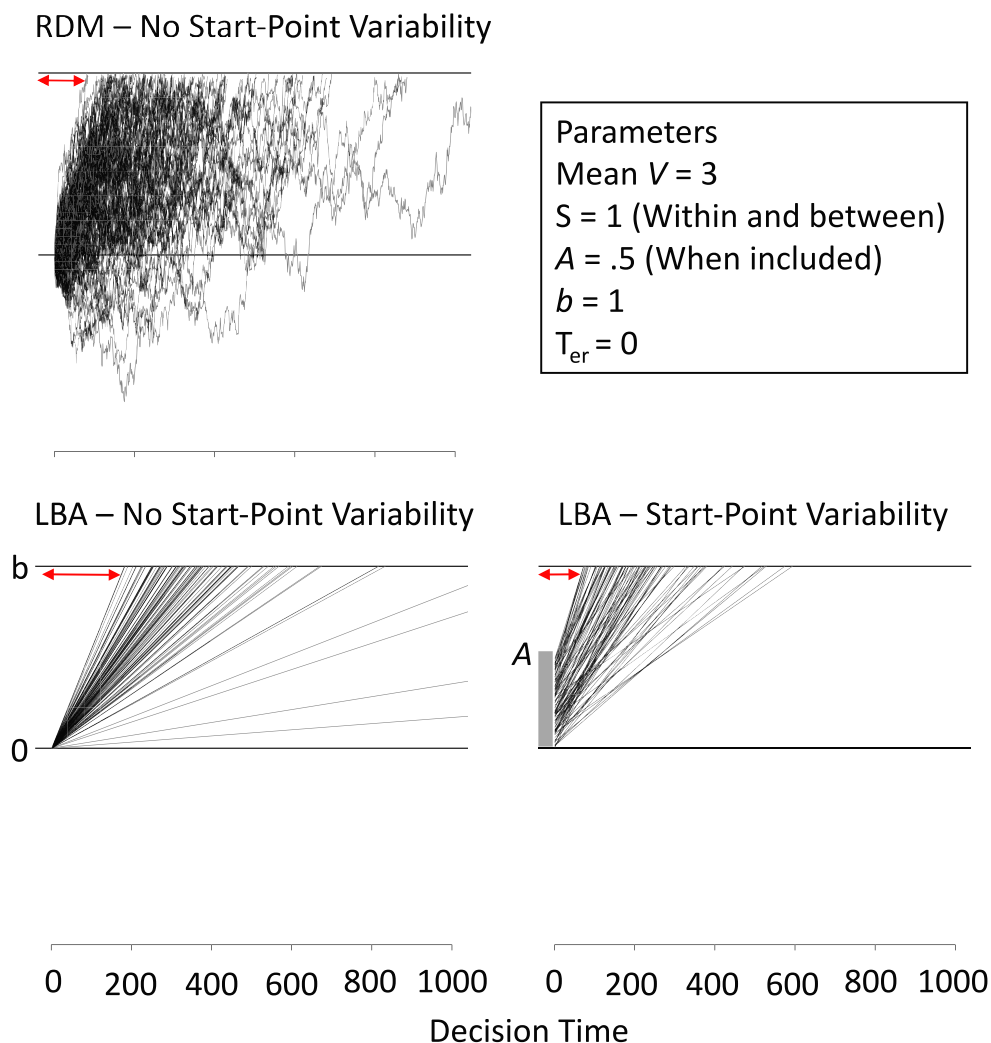
---

[3]Note that the original authors found an increase of 19 ms rather than 10 ms in non-decision time between placebo and high alcohol doses.

RDM – No Start-Point Variability

Parameters
Mean V = 3
S = 1 (Within and between)
A = .5 (When included)
b = 1
T_{er} = 0

LBA – No Start-Point Variability

b

0

LBA – Start-Point Variability

A

0   200   400   600   800   1000   0   200   400   600   800   1000

Decision Time

Fig. 14 Simulations of a single accumulator of the LBA and RDM process. The red arrows in the top left panel and the bottom two panels represents the length of time that goes by without any process finishing. Parameter values are shown in the top right panel. Note that the S parameter is the between-trial variability for the LBA and the within-trial variability for the RDM

Our work questions the necessity of some between-trial variability assumptions that have become routinely assumed by sequential sampling models. Given the evidence that was provided in the introduction for between-trial variability in drift rates, starting points, and non-decision time, we will now discuss the evidence in the context of what we have found testing the RDM.

Evidence for variability in drift rate

The original justification for including between-trial variability in the DDM was to account for item differences in recognition memory experiments (Ratcliff, 1978) and for slower errors than correct responses.

However, the RDM without between-trial variability in drift rate provides a good account of benchmark RT data. The multi-accumulator architecture naturally predicts slow errors as a consequence of the race. Furthermore, eliminating the between-trial variability in drift rate allows for a simpler likelihood equation speeding up the model fitting procedure. And so, when researchers are not interested in questions related to the between-trial variability components, then faster models such as the RDM could be an appropriate alternative. In principle, RDM could have between-trial variability in all its parameters, and it may be fruitful to develop such a model in the future if the data require it. Some of these additions have already been implemented in simple RT experiments (e.g., Tillman et al., 2017).

More recent empirical findings, such as the double pass experiment, also provide evidence for between-trial variability in drift rate (Ratcliff et al., 2018). The problem is that these findings did not distinguish between systematic and random sources of between-trial variability in drift

rate (Evans et al., in press). For example, in the double pass experiment the same perceptual stimuli are presented at two separate trials during an experimental session and so we would assume the two drift rates for the different trials are correlated. However, the between-trial drift rate in most sequential sampling models is just a univariate normal distribution with no information about the correlation between the two drift rates of the same stimulus. The only way to capture the relationship between two instances of the same stimulus is to assign the same (or similar) drift rate to the two trials. And so, although we have evidence that drift rates are not the same across trials, this does not justify capturing between-trial variability using probability distributions that provide no additional explanation.

Infinite model flexibility

Jones and Dzhafarov (2014) claimed that certain between-trial drift rate variability assumptions can allow infinite flexibility of sequential sampling models (cf. Heathcote et al., 2014; Smith et al., 2014). Our cross-fitting analysis can shed light on this claim. The only difference between the RDM with starting point variability and the linear ballistic accumulator (LBA; Brown & Heathcote, 2008), is whether drift rate varies within-trial or between-trial. But what should the distributional form of this variability be?

There is a principled reason to assume that within-trial drift rate variability is Gaussian. If we assume that the random increments in the evidence accumulation process are independent, then the functional central limit theorem states that the distribution of increments converge to a Gaussian (Donsker, 1951).

Jones and Dzhafarov (2014) claim that the between-trial Gaussian distribution assumption of drift rate is simply an implementation detail in the LBA. Yet, we found that the between-trial drift variability in LBA and the within-trial drift variability in the RDM are positively linearly related. In addition to this relationship, Eq. 5 (the RDM density function) is closely related to the LBA's density function, where

$$\alpha = \frac{(b - A - tv)}{ts} \tag{8}$$

and

$$\beta = \frac{(b - tv)}{ts} \tag{9}$$

then:

$$g_i(t|b, v, A) = \frac{1}{A}[-v\Phi(\alpha) + s\phi(\alpha) + v\Phi(\beta) - s\phi(\beta)] \tag{10}$$

where s is the standard deviation of the between-trial drift rate. Equation 5 simply replaces s with $\frac{1}{\sqrt{t}}$, assuming the

RDM has a diffusion coefficient of 1. $\frac{1}{\sqrt{t}}$ in the RDM represents the accumulated noise from taking evidence samples from a normal distribution at each time step. On the other hand, s in the LBA represents the noise from taking a single evidence sample on each trial from the normal distribution. And so, the between-trial drift rate assumptions of the LBA are justified in this regard. The LBA is a tractable proxy for speeded decision models that assume Gaussian within-trial noise, which as we argue, is justified by the functional central limit theorem.

Evidence for variability in starting points

The variability in starting points allow for sequential sampling models to account for error RTs that are faster than correct RTs. The LBA also needs this source of noise to account for the speed-accuracy trade-off. Yet in the current paper, we showed that the RDM with only within-trial drift rate variability can generate fast errors of about 10ms. For errors that are much faster than correct responses, researchers can assume starting point variability in the RDM too.

In future work, researchers could attempt to systematically model some of the starting-point variability using a model such as the RDM that can do without random starting-point variability. For example, the amount of starting-point variability is likely affected by the outcomes of previous decisions – sequential effects (cf. Jones et al., 2013). A key limitation of many sequential sampling models is that they assume each response is independent of the next (but see Brown et al., 2008), and they capture the sequential effects in the starting point variability parameter. In theory, the RDM could be used to estimate the starting point on each trial.

Variability in non-decision time

Non-decision time variability across trials allows the DDM to better fit the .1 quantile of RT distributions from a lexical decision task (Ratcliff, 2002; Ratcliff & Smith, 2004). A major concern is that this source of noise adds another numerical integration to the DDM equation, slowing down the model fitting process. The slowdown may not be justified given the variability in non-decision time is relatively small. The variability in non-decision time will be completely dominated by other sources of noise. For example, the standard deviation in the non-decision time is typically less than one-quarter of the standard deviation in the decision time, meaning it has little influence on the RT distribution (Ratcliff, 2002).

A constant non-decision time provides a substantial computational speed up, which is one reason why a constant value is useful for the more parsimonious LBA

and RDM. Whether a constant value of non-decision time is overly simplistic is yet to be formally investigated – in fact, the distributional properties of non-decision times are not well understood (Ratcliff, 2013; Verdonck & Tuerlinckx, 2015). We know that assuming normal and exponential distributions of non-decision time leads to parameter recovery issues in the DDM (Ratcliff, 2013). Yet, there is some evidence that non-decision times do have a non-negative positive skewed distribution (Verdonck & Tuerlinckx, 2015)

Perhaps the biggest concern is that even without between-trial variability, there are problems with the reliability of non-decision time estimates between different sequential sampling models. In a simulation study, Donkin et al. (2011) found that the DDM reliably estimates larger non-decision times than the LBA. In practice, the DDM also typically produces higher estimates than the LBA. For instance, when the LBA was fit to data from Ratcliff and Rouder (1998), the model produced non-decision time estimates (Brown & Heathcote, 2008, Table 2) between 27-66ms lower than the DDM (Ratcliff & Rouder, 1998, Table 1). In some instances, the DDM estimates non-decision times greater than 400 ms (Gomez et al., 2015), leaving 62-118ms of the mean RTs to be accounted for by the decision process.

It is not clear what the most appropriate magnitude for non-decision time is. Perceptual encoding time for simple visual stimuli can be as fast as 50 ms (Bompas & Sumner, 2011), while key-press responses can be as fast as 66 ms (Smith, 1995, p. 585). More recently, researchers have related measurements of electrophysiology in monkeys (Cook & Maunsell, 2002), and MEG and EEG in humans (Amano et al., 2006; Tandonnet et al., 2005; Vidal et al., 2011) to the RTs to detect simple visual stimuli. By doing so, researchers could partition out the time needed for visual perception, which is approximately 150-200 ms. Given that non-decision time is the sum of perceptual encoding and response production time, for 2 or N choice tasks with a key press response modality and stimuli as complex as motion dots, for example, we should expect non-decision time to be at least 200 ms.

One could argue that the magnitude of non-decision time isn't so important for drawing conclusions from sequential sampling models. Instead, researchers care about reliable differences between conditions of an experiment. However, Heathcote and Hayes (2012) demonstrated that non-decision time decreases with practice in a lexical decision task, but only for the DDM, not the LBA. Our own non-decision time analysis supports the idea that researchers could draw different conclusions from different sequential sampling models about perceptual encoding and motor response times.

Conclusion

In its simplest form, the RDM only has one source of variability: within-trial variability in drift rate. The simplicity of the RDM allowed us to test the necessity of between-trial variability in sequential sampling model parameters. Overall, we found good reasons to rethink including between-trial variability parameters in sequential sampling models. And at the very least, the minimal assumptions of the RDM make it a good candidate to model systematic variability in decision processes.

Appendix A: Cumulative density function

The cumulative density function for the RDM was derived for Logan et al. (2014), but only the R code for the function was available. For completeness, we present the equation here.

The CDF for the RDM is a Wald distribution with variability in start point. The PDF for the Wald distribution is

$$f(t \mid b, v) = \frac{b}{\sqrt{2\pi t^3}} \exp\left\{ -\frac{1}{2} \frac{(vt - b)^2}{t} \right\},$$

where b and v are the threshold and drift rate, respectively, of a diffusion process with a single absorbing boundary (b). This PDF can be integrated over t to give the corresponding CDF

$$F(t \mid b, v) = \Phi\left(\frac{vt - b}{\sqrt{t}} \right) + \exp(2vb)\, \Phi\left(\frac{-vt - b}{\sqrt{t}} \right).$$

The starting point $z < b$ of the process is implicitly equal to 0 in these expressions. Noting that a non-zero start point with threshold b is equivalent to a process with a zero start point and threshold $b - z$, so we can write

$$\begin{aligned}
F(t \mid b, v, z) &= F(t \mid b - z, v) \\
&= \Phi\left(\frac{vt - (b - z)}{\sqrt{t}} \right) + \exp(2v(b - z)) \\
&\quad \times \Phi\left(\frac{-vt - (b - z)}{\sqrt{t}} \right). \quad (11)
\end{aligned}$$

Our goal is to compute the CDF for the mixture the starting point z, where z follows a uniform[0, A] distribution. That is, we desire an expression for

$$F(t \mid b, v, A) = \frac{1}{A} \int_0^A F(t \mid b, v, z)\, dz.$$

We begin by noting first that the CDF Φ of the standard normal distribution can be written as a transformation of the error function:

$$\Phi(x) = \frac{1}{2}\left(\frac{x}{\sqrt{2}}\right) + \frac{1}{2}. \tag{12}$$

Therefore we can rewrite Eq. 11 as

$$F(t \mid b, v, z) = \frac{1}{2}\left[1 + \exp(2v(b-z)) + \left(\frac{vt-(b-z)}{\sqrt{2t}}\right)\right.$$
$$\left. + \exp(2v(b-z))\left(\frac{-vt-(b-z)}{\sqrt{2t}}\right)\right]. \tag{13}$$

Because the integration to be performed has a number of steps, for clarity we write

$$F(t \mid b, v, z) = \frac{1}{2}\left[\alpha(z) + \beta(z) + \gamma(z) + \delta(z)\right],$$

where

$\alpha(z) = 1$,
$\beta(z) = \exp(2v(b-z))$,
$\gamma(z) = \left(\dfrac{vt-(b-z)}{\sqrt{2t}}\right)$, and
$\delta(z) = \exp(2v(b-z))\left(\dfrac{-vt-(b-z)}{\sqrt{2t}}\right)$,

and we will integrate each term over z and then add the results to obtain $F(t \mid b, v, A)$.

$\alpha(z)$ and $\beta(z)$

The first two terms are trivial:

$$\frac{1}{A}\int_0^A \alpha(z)dz = 1,$$

and

$$\frac{1}{A}\int_0^A \beta(z)dz = \frac{\exp 2vb}{2vA}\left(1 - \exp(-2vA)\right).$$

$\gamma(z)$

We solve the integral of $\gamma(z)$ by noting first that

$$\int_0^x (u)du = x(x) + \frac{1}{\sqrt{\pi}}\left(\exp(-x^2) - 1\right).$$

Through the change of variable

$$u = \frac{vt-(b-z)}{\sqrt{2t}},$$

we see that

$$\frac{1}{A}\int_0^A \gamma(z)dz = \frac{1}{A}\int_0^A \left(\frac{vt-(b-z)}{\sqrt{2t}}\right)dz$$
$$= \frac{\sqrt{2t}}{A}\int_{a_1}^{a_2} (u)du, \tag{14}$$

where

$$a_1 = \frac{vt-b}{\sqrt{2t}} \quad\text{and}\quad a_2 = \frac{vt-(b-A)}{\sqrt{2t}}.$$

Equation 14 is equal to

$$\frac{\sqrt{2t}}{A}\left[\int_0^{a_2} (u)du - \int_0^{a_1} (u)du\right], \tag{15}$$

and the signs of a_1 and a_2 are irrelevant given the symmetry of the function (x). Because

$$\int_0^{a_i} (u)du = a_i(a_i) + \frac{1}{\sqrt{\pi}}\left(\exp(-a_i^2) - 1\right),$$

substitution into Eq. 15 yields

$$\frac{1}{A}\int_0^A \gamma(z)dz = \frac{\sqrt{2t}}{A}\left[a_2(a_2) + \frac{1}{\sqrt{\pi}}\left(\exp(-a_2^2) - 1\right)\right.$$
$$\left. - a_1(a_1) - \frac{1}{\sqrt{\pi}}\left(\exp(-a_1^2) - 1\right)\right].$$

Letting $\alpha_i = \sqrt{2}a_i$ and substituting back the transformation of the error function in Eq. 12 gives

$$\frac{1}{A}\int_0^A \gamma(z)dz = -1 + \frac{2\sqrt{t}}{A}\left\{[\alpha_2\Phi(\alpha_2) - \alpha_1\Phi(\alpha_1)]\right.$$
$$\left. + [\phi(\alpha_2) - \phi(\alpha_1)]\right\}, \tag{16}$$

where $\phi(x)$ is the standard normal PDF.

$\delta(z)$

The function $\delta(z)$ must be integrated by parts. We first apply a change of variable as was used to integrate $\gamma(z)$, that is,

$$x = \frac{-vt-(b-z)}{\sqrt{2t}}.$$

Then

$$\frac{1}{A}\int_0^A \delta(z)dz = \frac{1}{A}\int_0^A \exp(2v(b-z))$$
$$\times \left(\frac{-vt-(b-z)}{\sqrt{2t}}\right)dz$$
$$= \frac{\sqrt{2t}}{A}\int_{b_1}^{b_2} \exp\left(-2v(x\sqrt{2t} + vt)\right)(x)dx$$
$$= \frac{\sqrt{2t}}{A}\exp\left(-2tv^2\right)$$
$$\times \int_{b_1}^{b_2} \exp\left(-2v\sqrt{2t}x\right)(x)dx,$$

where

$$b_1 = \frac{-vt-b}{\sqrt{2t}} \quad\text{and}\quad b_2 = \frac{-vt-(b-A)}{\sqrt{2t}}.$$

Setting

$$u = (x) \quad\text{and}\quad v' = \exp\left(-2v\sqrt{2t}x\right)$$

and noting that

$$u' = \frac{2}{\sqrt{\pi}} \exp\left(-x^2\right) \text{ and } v = -\frac{1}{2v\sqrt{2t}} \exp(-2v\sqrt{2t}x),$$

integrating by parts gives

$$\frac{1}{A} \int_0^A \delta(z) dz$$
$$= \frac{\sqrt{2t}}{A} \exp\left(-2tv^2\right) \int_{b_1}^{b_2} \exp\left(-2v\sqrt{2t}x\right)(x) dx$$
$$= \frac{\sqrt{2t}}{A} \exp\left(-2tv^2\right) \left[\left(-\frac{1}{2v\sqrt{2t}} \exp(-2v\sqrt{2t}b_2)(b_2)\right.\right.$$
$$\left. + \frac{1}{2v\sqrt{2t}} \exp(-2v\sqrt{2t}b_1)(b_1)\right)$$
$$\left. + \int_{b_1}^{b_2} \frac{1}{v\sqrt{2t\pi}} \exp\left(-x^2 - 2v\sqrt{2t}x\right) dx\right]$$
$$= \frac{\sqrt{2t}}{A} \exp\left(-2tv^2\right) \left[\left(-\frac{1}{2v\sqrt{2t}} \exp(-2v\sqrt{2t}b_2)(b_2)\right.\right.$$
$$\left. + \frac{1}{2v\sqrt{2t}} \exp(-2v\sqrt{2t}b_1)(b_1)\right)$$
$$\left. + \frac{1}{v\sqrt{t}} \exp(2tv^2) \int_{b_1}^{b_2} \frac{1}{\sqrt{2\pi}} \exp\left(-(x + v\sqrt{2t})^2\right) dx\right]$$
$$= \frac{\exp\left(-2tv^2\right)}{2vA} \left[\left(-\exp(-2v\sqrt{2t}b_2)(b_2)\right.\right.$$
$$\left. + \exp(-2v\sqrt{2t}b_1)(b_1)\right)$$
$$\left. + \exp(2tv^2)\left((b_2 + v\sqrt{2t}) - (b_1 + v\sqrt{2t})\right)\right].$$

Transforming (x) back to $\Phi(x)$ and substituting $\beta_i = \sqrt{2}b_i$, and recalling that $\alpha_i = \sqrt{2}a_i$, we obtain

$$\frac{1}{A} \int_0^A \delta(z) dz$$
$$= \frac{1}{vA} \left\{[\Phi(\alpha_2) - \Phi(\alpha_1)] - [\exp(2v(b-A))\Phi(\beta_2)\right.$$
$$\left. - \exp(2vb)\Phi(\beta_1)]\right.$$
$$\left. + \frac{1}{2}\left[\exp(2v(b-A)) - \exp(2vb)\right]\right\}. \quad (17)$$

The solution

The CDF of the Wald distribution with uniform variability in start point is given by

$$F(t \mid b, v, A) = \frac{1}{2A} \int_0^A (\alpha(z) + \beta(z) + \gamma(z) + \delta(z)) \, dz.$$

Adding the four integrals computed in Sections "$\alpha(z)$ and $\beta(z)$" through "$\delta(z)$" gives

$$\frac{1}{A} \int_0^A (\alpha(z) + \beta(z) + \gamma(z) + \delta(z)) \, dz$$
$$= \frac{1}{vA} (\Phi(\alpha_2) - \Phi(\alpha_1)) + \frac{2\sqrt{t}}{A} (\alpha_2 \Phi(\alpha_2) - \alpha_1 \Phi(\alpha_1))$$
$$- \frac{1}{vA} \left[\exp(2v(b-A))\Phi(\beta_2) - \exp(2vb)\Phi(\beta_1)\right]$$
$$+ \frac{2\sqrt{t}}{A} (\phi(\alpha_2) - \phi(\alpha_1)).$$

Therefore,

$$F(t \mid b, v, A)$$
$$= \frac{1}{2vA} (\Phi(\alpha_2) - \Phi(\alpha_1)) + \frac{\sqrt{t}}{A} (\alpha_2 \Phi(\alpha_2) - \alpha_1 \Phi(\alpha_1))$$
$$- \frac{1}{2vA} \left[\exp(2v(b-A))\Phi(\beta_2) - \exp(2vb)\Phi(\beta_1)\right]$$
$$+ \frac{\sqrt{t}}{A} (\phi(\alpha_2) - \phi(\alpha_1)).$$

Appendix B: Model structure and fitting method

Each parameter for each subject was stochastically dependent on a group level distribution, ϕ_θ, where the subscript θ denotes the subject level parameter. We assumed that each group level distribution ϕ_θ was a truncated normal distribution, where $\phi_\theta \sim N(\mu, \sigma) \mid (lower, upper)$. For each group level distribution, the distribution range was the same as the hyper-priors on the group level mean parameter. We set hyper-priors on the group level parameters where the mean of ϕ_θ had the following priors: $A \sim N(.5, .5)|(0, \infty)$, $B \sim N(.5, .5)|(0, \infty)$, $v \sim N(2, 2)|(-\infty, \infty)$, $T_0 \sim N(.3, .3)|(0, 1)$. The standard deviation of ϕ_θ had a Gamma prior with shape parameter α and rate parameter β, where standard deviation $\sim \Gamma(\alpha = 1, \beta = 1)$. Priors supported a range of values identified in a literature review reported by Matzke and Wagenmakers (2009).

We estimated posterior distributions of parameter values using the differential evolution Markov Chain Monte Carlo method (DE-MCMC; Ter Braak, 2006; Turner et al., 2013). DE-MCMC has been shown to efficiently estimate parameters of hierarchical versions of models similar to the RDM (e.g., Turner et al., 2015; Turner et al., 2013). For all model fits in the paper we ran the DE-MCMC algorithm with 40 chains. The starting points of these chains were drawn from the following distributions: $A \sim N(.5, .5)|(0, \infty)$, $B \sim N(.5, .5)|(0, \infty)$, $v \sim N(2, 2)|(-\infty, \infty)$, $T_{er} \sim N(.3, .3)|(0, 1)$, where $N(m, sd)$ indicates a normal distribution with mean m and standard

distribution sd and the numbers after the | indicate the distribution range.

Part of approximating posterior distributions via sampling is deciding when convergence has been obtained, at which we are confident that samples represent the posterior distribution. All samples prior to convergence are discarded. To decide the point of convergence we both visually inspected the chains and discarded all samples prior to the $\hat{R}$ statistic being less than 1.01 (Gelman & Rubin, 1992). $\hat{R}$ represents the stability of parameter estimates within and between chains. The calculation involves comparing the within-chain and between-chain variances, as differences between the two sources of variance indicates a lack of stability in estimates, and potentially non-convergence. $\hat{R}$ has a value upwards of 1 with values closer to 1 indicating that the variance between chains is similar to the variance within chains, and thus, indicating better convergence.

Upon reaching the $\hat{R}$ criterion, we drew 5000 additional samples for each chain. To save memory during computing, and given the high auto-correlation within-chains, we thinned the posterior by only keeping every 10th iteration. These 20000 (i.e., 40 chains $\times$ 500 iterations) samples constituted our posterior distribution estimates. For all model fits the Wiener process standard deviation is fixed to $s = 1$. We also provide R code to implement the RDM model, which we host on the related Open Science Foundation web page: https://osf.io/m4btq/.

Appendix C: Model selection method

The gold standard for estimating the out-of-sample predictive accuracy of a model is cross-validation (Geisser & Eddy, 1979), but this method is computationally expensive and therefore we used a computationally cheap approximation to cross-validation. We used the widely applicable information criterion (WAIC; Watanabe, 2010) to assess the out-of-sample predictive accuracy of the LBA and RDM. WAIC requires calculating a goodness-of-fit value of a model and subtracting a value from this that represents the complexity of the model. In this regard, WAIC is like the well-known Akaike's information criterion (AIC; Akaike, 1974) or the Bayesian information criterion (BIC; Schwarz, 1978), but is applicable to hierarchical Bayesian models.

The first step to calculating WAIC is to compute for each posterior sample of $post_1, ..., post_S$ the likelihood of each data point y_i from data $y_1, ..., y_N$. For each data point, we calculate the average likelihood $Pr(y_i)$ over the entire posterior distribution as follows:

$$Pr(y_i) = \frac{1}{S} \sum_{s=1}^{S} Pr(y_i | \theta^s) \tag{18}$$

We then sum the log-likelihood over all data points to get the log point-wise predictive density, lppd, where:

$$lppd = \sum_{n=1}^{N} \log Pr(y_i). \tag{19}$$

The lppd is a biased estimate of how well the model predicts future data. It is biased because the data we use to evaluate the model is the same data we use to build the model. Essentially, in addition to fitting the signal in the data, the model has been optimized to fit noise in the data that will not be present in future data. The lppd overestimates the model's predictive accuracy and to approximate an unbiased estimate we subtract a measure of the model's complexity from lppd. One measure of complexity is the effective number of parameters p_{WAIC}. The effective number of parameters is a count of the total number of model parameters, but the metric accounts for the fact that all parameters in the model do not contribute to model's fit equally, and so, a parameter's contribution to the count can be values between 0 and 1. To compute p_{WAIC}, we first calculate the variance in log-likelihood across data points for each posterior sample, where:

$$V(y_i) = Var_{s=1}^{S}(\log Pr(y_i | \theta^s)). \tag{20}$$

We then sum the variance in log-likelihood over data points to get an estimate of the effective number of parameters, where:

$$p_{WAIC} = \sum_{n=1}^{N} V(y_i) \tag{21}$$

Using Eqs. 19 and 21 we can compute WAIC as follows:

$$WAIC = -2(lppd - p_{WAIC}) \tag{22}$$

Appendix D: Model recovery

We generated a synthetic data set by simulating the RDM model for an experiment with easy, medium, and hard difficulty, where difficulty effects were generated from systematic changes in the drift rate parameter. Thirty subjects were generated from the group-level distribution and we fit all of the simulated data sets with the RDM model using same parameterization as the generating distributions. Presented in Fig. 15 we display the generating group-level parameter values superimposed on the recovered group-level posterior distributions. The generating values of the group level distribution fall within the group-level posterior estimates of each parameter. In Fig. 16, we show that the mean of the subject-level posterior distribution correlates well with the subject-level generating parameters. Both our group-level and subject level simulation show a good recovery of parameters for the RDM.
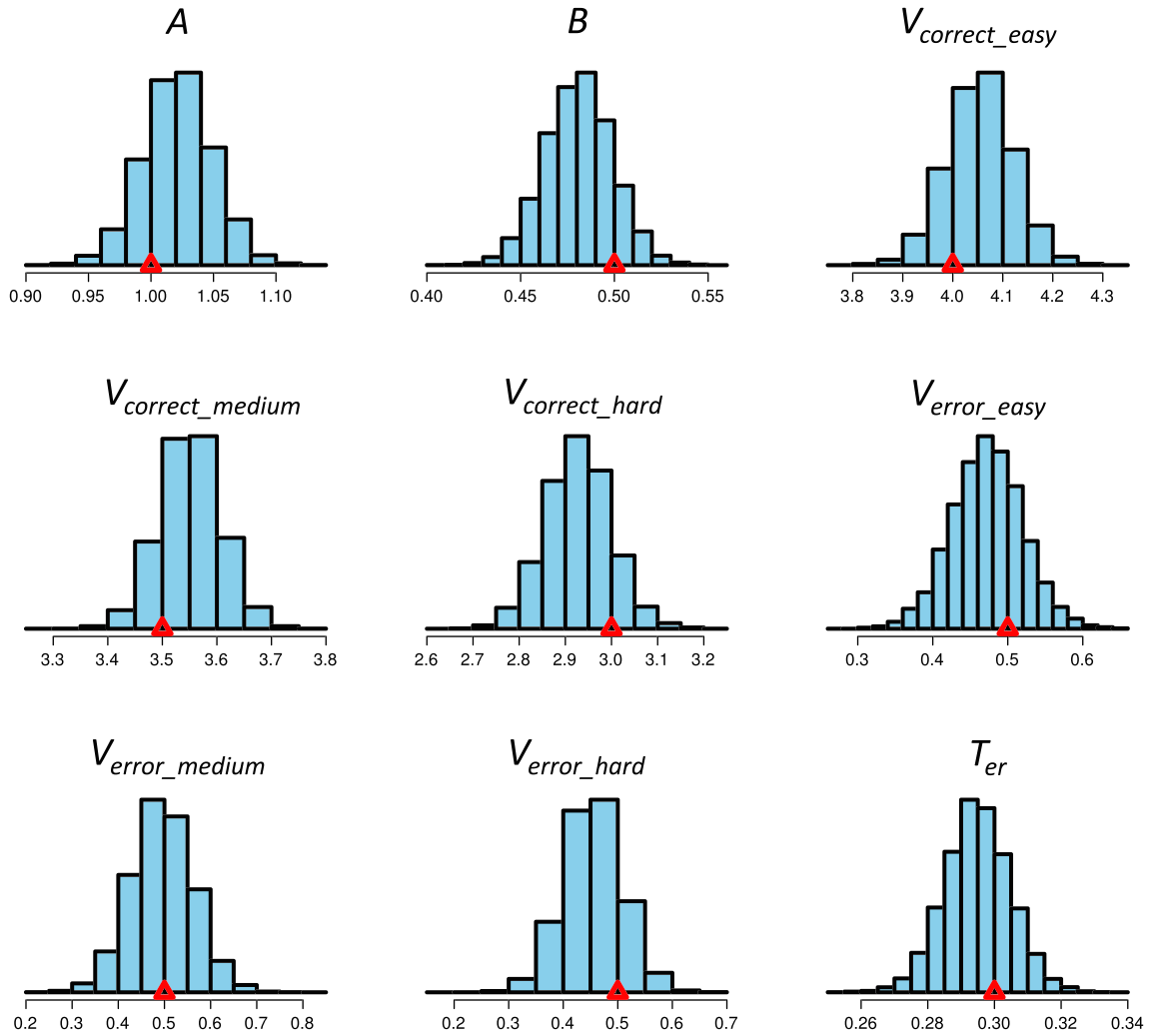
Group-Level Recovery

Fig. 15 Mean of group-level posteriors of generating parameters values are represented by the red triangle and the group-level posteriors of recovered parameter values are denoted by the blue histograms. Red triangles are all contained within the posterior distributions suggesting good recovery of parameters
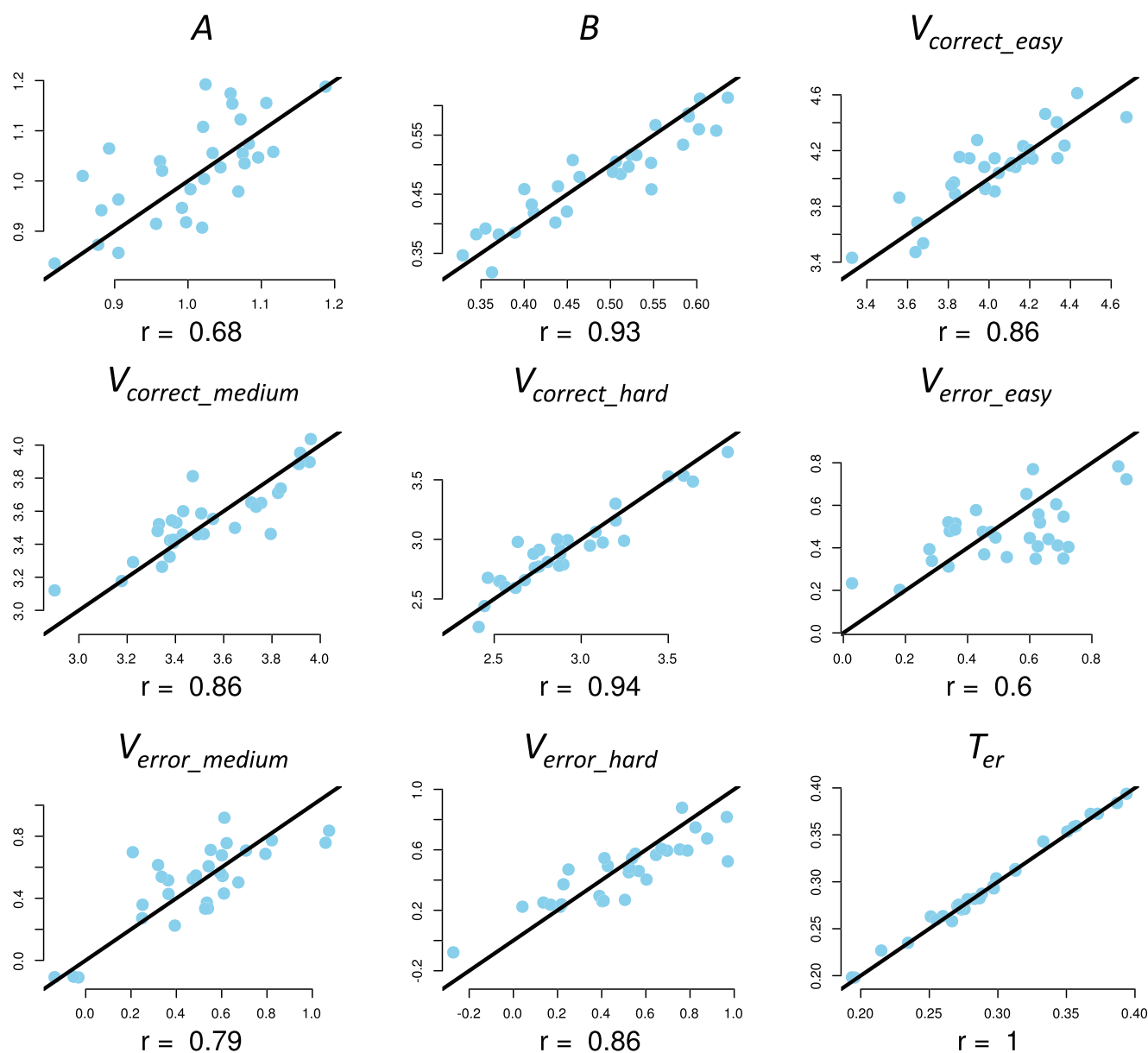
Subject-Level Recovery

Fig. 16 Scatterplots showing the mean of subject-level posteriors of generating parameters as a function of the mean subject-level recovered parameter values. Dots along the diagonal represent good recovery of parameters

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Amano, K., Goda, N., Nishida, S., Ejima, Y., Takeda, T., & Ohtani, Y. (2006). Estimation of the timing of human visual perception from magnetoencephalography. *The Journal of Neuroscience*, *26*(15), 3981–3991.

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

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., & et al. (2018). Estimating across-trial variability parameters of the diffusion decision model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46–75.

Bompas, A., & Sumner, P. (2011). Saccadic inhibition reveals the timing of automatic and voluntary signals in the human brain. *The Journal of Neuroscience*, *31*(35), 12501–12512.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*(1), 117.

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

Brown, S. D., Marley, A., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, *115*(2), 396.

Cook, E. P., & Maunsell, J. H. (2002). Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nature Neuroscience*, *5*(10), 985–994.

Ditterich, J. (2006a). Evidence for time-variant decision making. *European Journal of Neuroscience*, *24*(12), 3628–3641.

Ditterich, J. (2006b). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks*, *19*(8), 981–1012.

Donkin, C., & Brown, S. D. (2018). Response times and decision making. In Wixted, T., & Wagenmakers, E. J. (Eds.) *The Stevens' handbook of experimental psychology and cognitive neuroscience*. (4th ed., Vol. 5. New York: Wiley.

Donkin, C., Brown, S. D., 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.

Donkin, C., Heathcote, A., & Brown, S. (2009a). 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.

Donkin, C., Heathcote, A., Brown, S., & Andrews, S. (2009b). Non-decision time effects in the lexical decision task. In *Proceedings of the 31st annual conference of the cognitive science society*. Austin: Cognitive Science Society.

Donsker, M. D. (1951). An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*.

Egan, J. P. (1958). Recognition memory and the operating characteristic. *USAF Operational Applications Laboratory Technical Note*.

Evans, N. J., Tillman, G., & Wagenmakers, E.-J. (in press). Systematic and random sources of variability in perceptual decision-making: Comment on Ratcliff, Voskuilen, and Mckoon (2018). Psychological Review.

Fecteau, J. H., & Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nature Reviews Neuroscience*, *4*(6), 435.

Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, *74*(365), 153–160.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511.

Gomez, P., Ratcliff, R., & Childers, R. (2015). Pointing, looking at, and pressing keys: A diffusion model account of response modality. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1515–1523.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427.

Hawkins, G. E., Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2012). An optimal adjustment procedure to minimize experiment time in decisions with multiple alternatives. *Psychonomic Bulletin & Review*, *19*(2), 339–348.

Heathcote, A. (2004). Fitting Wald and ex-Wald distributions to response time data: An example using functions for the S-PLUS package. *Behavior Research Methods*, *36*, 678–694.

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), 125–36.

Heathcote, A., Wagenmakers, E. J., & Brown, S. D. (2014). The falsifiability of actual decision-making models.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188.

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, *120*(3), 628.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*(1), 1.

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.

Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics*, *72*(1), 246–273.

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology, 7*.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*, 77–105.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492.

Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review*, *121*(1), 66–95.

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798–817.

Osth, A. F., Dennis, S., & Heathcote, A. (in press). Likelihood ratio sequential sampling models of recognition memory. *Cognitive Psychology*.

Osth, A. F., & Farrell, S. (2019). Using response time distributions and race models to characterize primacy and recency effects in free recall initiation. *Psychological Review*, *126*(4), 578.

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*(4), 1113–1143.

Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From salience to saccades: Multiple-alternative gated stochastic accumulator model of visual search. *The Journal of Neuroscience*, *32*(10), 3433–3446.

Raab, D. H. (1962). Division of psychology: Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*(5 Series II), 574–590.

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*(5), 1226–43.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, *9*(2), 278–291.

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

Ratcliff, R. (2015). Modeling one-choice and two-choice driving tasks. *Attention, Perception, & Psychophysics*, *77*(6), 2134–2144.

Ratcliff, R., Gomez, P., & McKoon, G. M. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182.

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences*, *106*(16), 6539–6544.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127.

Ratcliff, R., Sederberg, P. B., Smith, T. A., & Childers, R. (2016). A single trial analysis of EEG in recognition memory: Tracking the neural correlates of memory strength. *Neuropsychologia*, *93*, 128–141.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–67.

Ratcliff, R. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*(1), 59–83.

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.

Ratcliff, R., & Strayer, D. (2014). Modeling simple driving tasks with a one-boundary diffusion model. *Psychonomic Bulletin & Review*, *21*(3), 577–589.

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

Ratcliff, R., & Van Dongen, H. P. (2011). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*(27), 11285–11290.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*(2), 261.

Ratcliff, R., Voskuilen, C., & McKoon, G. (2018). Internal and external sources of variability in perceptual decision-making. *Psychological Review*.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods*, *33*(4), 457–469.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*(3), 567–593.

Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dzhafarov (2014). *Psychological Review*, *121*(4), 679–688.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*(2), 135–168.

Starns, J. J. (2014). Using response time modeling to distinguish memory and decision processes in recognition and source tasks. *Memory & Cognition*, *42*(8), 1357–1372.

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language*, *70*, 36–52.

Tandonnet, C., Burle, B., Hasbroucq, T., & Vidal, F. (2005). Spatial enhancement of EEG traces by surface Laplacian estimation: Comparison between local and global methods. *Clinical Neurophysiology*, *116*(1), 18–24.

Teller, D. Y. (1984). Linking propositions. *Vision Research*, *24*(10), 1233–1246.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, *120*(1), 1–38. https://doi.org/10.1037/a0030776

Ter Braak, C. J. (2006). A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, *16*(3), 239–249.

Tillman, G., Osth, A., van Ravenzwaaij, D., & Heathcote, A. (2017). A diffusion decision model analysis of evidence variability in the lexical decision task. *Psychonomic Bulletin & Review*, *24*(6), 1949–1956. Retrieved from https://doi.org/10.3758/s13423-017-1259-y

Tillman, G., Strayer, D., Eidels, A., & Heathcote, A. (2017). Modeling cognitive load effects of conversation between a passenger and driver. *Attention, Perception, & Psychophysics*, *79*(6), 1795–1803.

Townsend, J. T., & Ashby, F. G. (1983). Stochastic modeling of elementary psychological processes. CUP Archive.

Turner, B. M. (2019). Toward a common representational framework for adaptation. *Psychological Review*, *126*(5), 660.

Turner, B. M., Gao, J., Koenig, S., Palfy, D., & McClelland, J. L. (2017). The dynamics of multimodal integration: The averaging diffusion model. *Psychonomic Bulletin & Review*, *24*(6), 1819–1843.

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), 368–84.

Turner, B. M., van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: The neural drift diffusion model. *Psychological Review*, *122*(2), 312–336.

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

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed–accuracy tradeoff. *Journal of Mathematical Psychology*, *46*(6), 704–715.

Van Maanen, L., Grasman, R. P., Forstmann, B. U., Keuken, M. C., Brown, S. D., & Wagenmakers, E.-J. (2012). Similarity and 1399 number of alternatives in the random-dot motion paradigm. *Attention, Perception, & Psychophysics*, *74*(4), 739–753.

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review*, *24*(2), 547–556.

van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2012). A diffusion model decomposition of the effects of alcohol on perceptual decision making. *Psychopharmacology*, *219*(4), 1017–1025.

Verdonck, S., & Tuerlinckx, F. (2015). Factoring out non-decision time in choice RT data: Theory and implications. *Psychological Review*, *123*(2), 208–218.

Vidal, F., Burle, B., Grapperon, J., & Hasbroucq, T. (2011). An ERP study of cognitive architecture and the insertion of mental processes: Donders revisited. *Psychophysiology*, *48*(9), 1242–1251.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, *11*, 3571–3594.

Woodman, G. F., Kang, M.-S., Thompson, K., & Schall, J. D. (2008). The effect of visual search efficiency on response preparation neurophysiological evidence for discrete flow. *Psychological Science*, *19*(2), 128–136.