A parameter recovery assessment of time-variant models of decision-making

Nathan J. Evans^{1,2} · Jennifer S. Trueblood¹ · William R. Holmes³

Published online: 28 March 2019
© The Psychonomic Society, Inc. 2019

Abstract

Evidence accumulation models have been one of the most dominant modeling frameworks used to study rapid decision-making over the past several decades. These models propose that evidence accumulates from the environment until the evidence for one alternative reaches some threshold, typically associated with caution, triggering a response. However, researchers have recently begun to reconsider the fundamental assumptions of how caution varies with time. In the past, it was typically assumed that levels of caution are independent of time. Recent investigations have however suggested the possibility that levels of caution decrease over time and that this strategy provides more efficient performance under certain conditions. Our study provides the first comprehensive assessment of this newer class of models accounting for time-varying caution to determine how robustly their parameters can be estimated. We assess five overall variants of collapsing threshold/urgency signal models based on the diffusion decision model, linear ballistic accumulator model, and urgency gating model frameworks. We find that estimation of parameters, particularly those associated with caution/urgency modulation are most robust for the linearly collapsing threshold diffusion model followed by an urgency-gating model with a leakage process. All other models considered, particularly those with ballistic accumulation or nonlinear thresholds, are unable to recover their own parameters adequately, making their usage in parameter estimation contexts questionable.

Keywords Decision-making · Collapsing thresholds · Urgency gating · Parameter recovery · Probability density approximation · Bayesian parameter estimation

Introduction

Over the past several decades, the study of rapid decision-making has been used as a tool to investigate the properties

of the psychological processes responsible for decisions. One of the distinct benefits of using rapid decision-making in this context is the ability to develop process level models that link proposed theories to data. Within this field, more than five decades of research have been dedicated to the development and study of evidence accumulation models (EAMs), the dominant modeling paradigm used to study these decisions (Stone, 1960). However, researchers, particularly those in the neuroscience and primate research communities (Cisek, Puskas, & El-Murr, 2009; Thura, Beauregard-Racine, Fradet, & Cisek, 2012; Ditterich, 2006a; Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012), have begun to reconsider some of the most fundamental assumptions of this modeling framework. Whereas five decades of research have centered around the idea that accumulation of evidence over time is the basis of decisions, more recent work has suggested that temporal modulation of caution or a sense of urgency plays a prominent role in decisions. This has led to the recent proliferation of assessments of the role and importance of these factors in decision-making, as encoded in response time (RT) models (Hawkins, Forstmann, Wagenmakers,

All authors were supported by NSF grant SES-1556415. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agency. All associated code can be found at: <https://osf.io/xd8p2/>

✉ Jennifer S. Trueblood
jennifer.s.trueblood@vanderbilt.edu

✉ William R. Holmes
william.holmes@vanderbilt.edu

¹ Department of Psychology, Vanderbilt University, Nashville, TN, USA

² Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

³ Department of Physics and Astronomy, Department of Mathematics, Quantitative Systems Biology Center, Vanderbilt University, Nashville, TN, USA

Ratcliff, & Brown, 2015; Hawkins, Wagenmakers, Ratcliff, & Brown, 2015; Evans, Hawkins, Boehm, Wagenmakers, & Brown, 2017; Evans, Hawkins, & Brown, 2018; Evans & Hawkins, 2019; Ditterich, 2006a; Dutilh et al., 2018; Drugowitsch et al., 2012; Cisek et al., 2009; Thura et al., 2012; Winkel, Keuken, van Maanen, Wagenmakers, & Forstmann, 2014; Ditterich, 2006b; Trueblood et al., 2018; Carland, Marcos, Thura, & Cisek, 2015). However, while the mathematical and statistical properties of classical EAMs have been thoroughly investigated, to date this new class of models (e.g., collapsing bound or urgency based models) has not. Here, we provide a thorough analysis of the basic properties of these models and address questions such as, “can these models be reliably fit to data?” and, “can they be used for parameter inference?”. Our purpose here is not to compare these different models. Rather, it is to assess the basic properties of these models and to determine how they can and should be used in the future.

Over the past several decades, evidence accumulation models (EAMs) have served as an important tool to investigate the properties of rapid decision-making research (Hawkins et al., 2014; Matzke, Dolan, Logan, Brown, & Wagenmakers, 2013; Forstmann et al., 2011; Gomez, Ratcliff, & Perea, 2007; Ho et al., 2014; Ratcliff, Thapar, & McKoon, 2011; Ratcliff, Thapar, & McKoon, 2010; Evans & Brown, 2017; Evans, Rae, Bushmakin, Rubin, & Brown, 2017). Specifically, EAMs propose that evidence is accumulated for each of the decision alternatives until the evidence for one of the alternatives reaches a threshold level of evidence, which triggers a decision. Although several different specific EAMs have been proposed, all EAMs contain two critical parameters that explain the process described above: the “drift rate”, which is the rate of evidence accumulation for an alternative, and the “threshold”, which is the amount of evidence required to trigger a decision for an alternative (see (Ratcliff, Smith, Brown, & McKoon, 2016) and (Donkin & Brown, 2017) for reviews). Recently, researchers have begun to reconsider the psychological assumptions associated with response thresholds. In the past, it was typically assumed that response thresholds are constant over time, reflecting the assumption that a person’s level of caution is fixed within a decision and does not change over time. However, recent studies have begun to investigate whether those thresholds change (e.g., “collapse” or decrease) over time, reflecting the assumption that “urgency” leads to a reduction in caution over time (Cisek et al., 2009; Drugowitsch et al., 2012).

Investigation of these time-varying decision mechanisms has taken two basic forms. The first involves the presence of a “collapsing threshold”, which encodes the assumption that a decreasing amount of evidence is required to trigger a decision as the time spent on the decision increases

(Drugowitsch et al., 2012; Ditterich, 2006a). Alternatively, there have also been proposals of a conceptually similar “urgency signal”¹, where the thresholds remain fixed, but the evidence signal is multiplied by a value that continues to increase as the time spent on the decision increases (Cisek et al., 2009; Thura et al., 2012). These time-varying mechanisms are appealing for several normative reasons. Specifically, they allow greater efficiency (e.g., maximizing the time discounted rate of return) than fixed thresholds when experimental difficulty varies between trials (Drugowitsch et al., 2012; Thura et al., 2012), and allow fast deadlines to be met (Frazier & Yu, 2007).

The previous assessments of collapsing thresholds models have primarily relied on model comparison, e.g., comparing a fixed threshold EAM and a collapsing threshold EAM to determine how well each describes the process that underlies the data. The use of model comparison has resulted in mixed findings regarding whether humans generally seem to implement these time-varying mechanisms. Earlier studies focused on qualitative trends, which favored the presence of a time-varying mechanism (Cisek et al., 2009; Thura et al., 2012). However, studies using quantitative model selection have mostly showed evidence against these time-varying mechanisms (Hawkins et al., 2015; Hawkins et al., 2015; Voskuilen, Ratcliff, & Smith, 2016; Evans et al., 2017). Although, some recent quantitative studies (Palestro, Weichart, Sederberg, & Turner, 2018; Evans et al., 2018; Evans & Hawkins, 2019) have found evidence in favor of time-varying mechanisms in specific paradigms.

However, another important goal that has largely been ignored within the collapsing thresholds literature is parameter estimation, which involves measuring the latent parameters from data to test hypotheses (see (Kruschke & Liddell, 2015) for a discussion of the importance of estimation). With this approach, rather than formulating multiple models and comparing them, one formulates a single model whose specific parameter values determine its properties. For example, one could consider a collapsing thresholds model and measure the extent to which the thresholds collapse. A rate of collapse near 0 would be a strong indicator of the absence of collapsing thresholds, while a rate that deviates significantly from 0 would indicate its importance. However, this approach requires that the crucial parameters of the models being considered (the “collapse rate” in this example) can be reliably estimated from data. While parameter recovery studies for classical

¹In the case of an additive urgency signal, the urgency signal is mathematically identical to a collapsing threshold. However, in the more commonly applied multiplicative urgency signal, the urgency signal and collapsing threshold are very similar, though their exact mathematical forms differ.

EAMs have been carried out to validate their use for parameter inference (e.g., van Ravenzwaaij & Oberauer, 2009; Donkin, Brown, Heathcote, & Wagenmakers, 2011; Lerche & Voss, 2016; White, Servant, & Logan, 2018), this has not been performed for this new class of collapsing threshold/urgency models.

Our study aims to provide the first comprehensive assessment of parameter estimation with collapsing thresholds and urgency signal models by performing a large-scale parameter recovery study utilizing state-of-the-art Bayesian methods. Importantly, if a model cannot reliably estimate the correct parameter values from data generated by those very parameters, then the estimated parameters from the model when applied to real data are of little meaning. This has proven to be an issue for some existing cognitive models (e.g., see (Miletić, Turner, Forstmann, & van Maanen, 2017) for a parameter recovery study on the leaky-competing accumulator [LCA; (Usher & McClelland, 2001)] that shows an inability to recover several key parameters of the model) and more generally for models of complex biological processes (Holmes, 2015; Gutenkunst et al., 2007). We assess the parameter identifiability of three distinct model types based on the diffusion decision model framework (DDM; (Ratcliff, 1978)), linear ballistic accumulator framework (LBA; Brown & Heathcote 2008), and the urgency gating model framework (UGM; Cisek et al. 2009; Thura et al. 2012), which are schematically outlined in Fig. 1. Additionally, we consider two variants of collapse bound diffusion and urgency gating model formulations for completeness. Finally, based on prior suppositions (e.g., Cisek et al. 2009; Thura et al. 2012) that the type of data that is used when working with these models is of critical importance, we perform this assessment with two types of data, constant-evidence paradigms and changing-evidence paradigms.

Models

We first outline the collection of models that we perform the parameter recovery assessment on within this article. Five specific models are considered: two versions of a collapse threshold DDM ($CTDDM$), a collapse threshold variant of the LBA model ($CTLBA$), and two variants of the UGM. The two versions of the DDM differ only in the form of the collapsing threshold: in one case, we consider a linear collapsing threshold and in the other a non-linear. The two versions of the UGM differ only slightly in their mathematical formulation and interpretation: one is formulated in terms of a leakage process and the other in terms of a low-pass filtering process. For the collapse threshold LBA, we consider only a version with a linearly collapsing threshold.

We provide these assessments for both constant-evidence and changing evidence paradigms since (1) both of these paradigms have been used in previous assessments of time-varying models, and (2) as discussed by Cisek and colleagues (Cisek et al., 2009; Thura et al., 2012), time-varying decisions may be more appropriate to assess models that encode time-varying hypotheses. Therefore, the potentially richer data obtained from a changing-evidence paradigm *may* be more suitable for probing the importance of collapsing thresholds / urgency.

DDM

The DDM (Ratcliff, 1978; Ratcliff & Rouder, 1998), the most commonly used EAM used to study rapid, binary decisions, proposes that evidence stochastically accumulates until one of two thresholds, corresponding to one of the binary alternatives, is reached (Fig. 1, middle panel). The accumulation of evidence between these binary alternatives can be expressed as:

$$dE = vdt + \sigma dW \quad (1)$$

where E is the evidence state. This model contains the following parameters: drift rate (v), starting point (z , which represents response bias), non-decision time (ter), and the response threshold (a). In this implementation of the DDM, the response thresholds are respectively at $\pm a$. Furthermore,

Fig. 1 Examples of the three different model variants considered within our manuscript. In all cases, the *green lines* display evidence accumulation in a constant-evidence paradigm, and the *blue lines* display evidence accumulation in a changing-evidence paradigm with a piecewise extension of the models. The *red vertical lines* display the time at which the evidence changes in the environment (t_{switch}), with the accumulation of both models being identical until the evidence change. **a** Diagram of the LBA. The *solid colored lines* display the accumulation of evidence for the first alternative, and the *dashed colored lines* display the accumulation for the second alternative. The *black line* displays the standard fixed threshold within the model that evidence for each alternative accumulates to, whereas the *gray line* displays the linear collapsing thresholds that we attempt to measure within this study. In addition, there is a switch time delay parameter (t_{delay}) included, which is an offset from when the evidence change in the environment begins to effect the accumulation process. **b** Diagram of the DDM. The *solid black lines* display the standard fixed thresholds, the *solid gray lines* display a linear collapsing threshold, and the *dashed gray lines* display a Weibull collapsing threshold. In addition, there is a switch time delay parameter (t_{delay}) included, which is an offset from when the evidence change in the environment begins to effect the accumulation process. **c** A diagram of the UGM

we consider only the (v, ter, a) parameters and fix $z=0$. Fixing the bias is done since most studies of time-varying caution do not include it. The DDM can also be extended to included between decision variability in drift rate, starting point, and/or non-decision time, though we do not consider these factors here as these parameters are known to have recovery issues (Lerche & Voss, 2016) and are rarely the primary focus of parameter inference.

We provide two time-varying extensions to the DDM. Firstly, we extend the DDM to a collapsing thresholds model with a linear collapsing function:

$$a_{upper}(t) = a_0 - a_{slope}t \quad (2)$$

The lower boundary is given by $a_{lower}(t) = -a_{upper}(t)$. This adds a single extra parameter to the DDM framework: the linear rate of collapse (a_{slope}) from the initial threshold. Secondly, we utilize a non-linear collapse function that has been applied within previous literature (Hawkins et al., 2015; Evans et al., 2018; Evans & Hawkins, 2019), where the collapse takes the form of a Weibull function:

$$a_{upper}(t) = a_0 - pWeibull(t; shape, scale) \times (a_0 - a_{asympt}) \quad (3)$$

The lower boundary again satisfies $a_{lower}(t) = -a_{upper}(t)$. The Weibull function contains three parameters corresponding to the shape of the collapse ($shape$), the scale of the collapse ($scale$), and the asymptote of the Weibull function (a_{asympt}) at $t \rightarrow \infty$. For simplicity, we fixed the minimum collapse point to be very small ($1e^{-3}$), resulting in two extra parameters in addition to the standard DDM parameters: the shape ($shape$) and scale ($scale$) of the Weibull collapse function.

LBA

The LBA (Fig. 1, top panel) proposes that evidence for different alternatives in a decision accumulate independently and deterministically (e.g., no within-trial noise), where the accumulation of evidence for each alternative can be expressed as:

$$dE_a = v_a dt, \quad (4)$$

where E is the evidence state, and the subscript a indexes the alternative. In this framework, each alternative is associated with a drift rate (v_a) whose value corresponds to the strength of evidence for that alternative. This model contains the following parameters for a binary choice: two drift rates v_1, v_2 for the two alternatives, the threshold (b), the start point variability A , the non-decision time ter , and trial to trial variability in drift rates s (this is critical to the model due to the absence of within trial variability). This gives the LBA six parameters: v_1, v_2, s, b, A , and t_0 .

We extend the LBA to a collapsing thresholds model with a linear collapsing function:

$$b(t) = b_0 - b_{slope}t \quad (5)$$

Importantly, a linear collapsing function maintains the computational simplicity that the LBA originally was designed for, as the time that the threshold is crossed (i.e., the response time) can easily be obtained by the point of intersection between the lines. As our goal is to maintain the simplicity of the LBA framework, we do not include a more computationally taxing Weibull collapsing function. This adds a single extra parameter to the LBA framework: the linear rate of collapse from the initial threshold (b_{slope}).

UGM

The UGM (Fig. 1, bottom panel) proposes a similar process of dependent evidence accumulation as the DDM. However, the UGM proposes that evidence is barely accumulated at all, with a focus on novel evidence (Cisek et al., 2009). The essential assumption of this model is that evidence, rather than being accumulated, is first smoothed by some form of leakage or filtering process, and then multiplied by a time-varying urgency signal.

We assess two versions of the UGM within our parameter recovery. Firstly, we assess the UGM previously fit in Hawkins et al. (2015) and Evans et al. (2017) and described previously (e.g., Carland et al. 2015), which assumes evidence is filtered prior to being weighted by the urgency signal. This version of the model has the following five parameters: v, a, ter, u , and τ , and can be expressed as:

$$x_{(t+\Delta)} = utE_{(t+\Delta)}, \quad E_{(t+\Delta)} = E_{(t)} + \frac{\Delta}{\tau + \Delta}(-E_{(t)} + vdt + \sigma dW). \quad (6)$$

where the two additional parameters beyond the diffusion model, u and τ , are the urgency signal and time constant of the filtering process, respectively. Since this is a non-standard random walk that does not have a clear continuous analogue, we express it in its discrete form rather than a stochastic differential equation. Specifically, Δ is the step-size of the process, ut is the time-dependent urgency signal, and $x_{(t+\Delta)}$ is the urgency-transformed evidence value that is compared against the threshold at each step to determine if a decision is made.²

We note that the variable E here has a slightly different meaning than in the prior models. In prior models, E was

²Note that within this definition, as used within previous literature, the urgency is not incorporated into the accumulation process, and instead is applied to the accumulated evidence at each step

the decision variable whose accumulation to a threshold triggers a response. The UGM theory, however, posits that the evidence signal is weighted by an urgency signal to produce an effective decision variable. Here, E represents the evidence signal, ut is the urgency signal, and x is the decision variable.

The second version we consider differs slightly in its mathematical form. It assumes that rather than being subjected to a filter with an associated time constant, evidence is subject to leakage with an associated rate (Carland, Thura, and Cisek, 2015). This version of the model has the following four parameters: v , a , ter , and L , and can be expressed as:

$$x = utE, \quad dE = (-LE + v)dt + \sigma dW, \quad (7)$$

where u is fixed at 1 (see Appendix C for further detail on this choice). We stress again that this is not a fundamentally different model but rather a slightly different mathematical formulation of the UGM. In particular, the leakage in this formulation has the effect of a smoothing filter with a time constant related to $1/L$. We include both of these mathematical instantiations of the UGM for completeness. We note that for this version of the UGM, the urgency rate is mathematically unidentifiable and thus we do not attempt its recovery here, see the Appendix for further details.

Accounting for changes of evidence

While changing-evidence paradigms have been of great interest in decision-making literature (Diederich, 1997; Diederich & Busemeyer, 1999; Usher & McClelland, 2001; Diederich & Busemeyer, 2006; Diederich, 2008; Kiani, Hanks, & Shadlen, 2008; Tsetsos, Usher, & McClelland, 2011; Tsetsos, Gao, McClelland, & Usher, 2012; Evans et al., 2017), models extended to these paradigms have rarely been the focus of parameter recovery studies (though see (Holmes, Trueblood, & Heathcote, 2016) for a parameter recovery study on the piecewise LBA and Holmes and Trueblood (2018) for a parameter recovery study on the piecewise DDM). In the current parameter recovery study, we assess the recoverability of parameters in changing-evidence paradigms where evidence changes at a single, known discrete point in time (t_{switch}). For example, in a random dot motion task, this change would correspond to a change in the motion of dots at a specific point in time (e.g., from initially moving to the left to moving to the right after 200 ms). Accounting for this change of evidence requires augmentation of the models. For the LBA, we utilize a piecewise extension (pLBA, Holmes et al. 2016) where it is assumed that after the change of evidence, drift rates change to reflect the new evidence. We further assume

there is an unknown delay between the change of stimulus and the change of drift t_{delay} , which adds an additional parameter to the model (see Fig. 1). This parameter captures the time between the objective change in stimulus and the time it takes an individual to adapt to and encode the new information. A similar piecewise extension is applied to the DDM (as in Holmes & Trueblood 2018), which again adds a parameter t_{delay} .

For the UGM, we similarly assume that the evidence parameter changes in response to the change of evidence. In this case, however, we do not include a delay between the change of stimulus and change of parameter. The addition of the delay parameter to the DDM and LBA frameworks is to account for the fact that the change of stimulus likely does not result in an instantaneous change in drift rate parameters. Rather, there would likely be a nonlinear response to the change of evidence that changes the drift rates from v_{before} to v_{after} over a period of time. The inclusion of this delay parameter is essentially an approximation that accounts for the time it takes for that non-linear process to generate the new drift rates. In the UGM, however, the filtering of evidence (or alternatively leakage) already provides a mechanism to smooth the transition from before to after the change of stimulus, with the filtering time constant (τ) or leakage rate (L) setting the timescale over which the transition occurs. More specifically, recall that in the UGM, it is not the sensory evidence that is multiplied by the urgency signal, but rather a smoothed (by filtering or leakage) version of it. That smoothing is what the t_{delay} in the DDM and LBA models is intended to account for, and thus the t_{delay} is not mechanistically needed in UGM. Thus, additional parameters are not required to account for changes of evidence in UGM.

Methods

For all recoveries, we used the same general method, though the exact details differed based on the number and identity of the parameters. To assess the identifiability of parameters of each of the models, we performed a large-scale simulation/parameter recovery study utilizing state of the art Bayesian methods. For each model, a large number of parameter sets were chosen, synthetic data was simulated based on those parameters, and parameter recovery was performed for each parameter set using that simulated data. We also used two types of input evidence in our simulations: constant evidence, where the evidence being integrated from the environment remains constant over time, and changing evidence, where the evidence being integrated from the environment changes at a known, fixed time during the trial. This was motivated by the suggestions of Cisek et al. (2009) that time-variant models should be most

distinguishable from time-variant models in changing-evidence paradigms.

To generate the simulated datasets, we used a Latin hypercube sampling design (McKay, Beckman, and Conover, 1979). Importantly, this method allows for the most efficient method of sampling of the parameter space of interest. We generated 4000 parameter sets using the Latin hypercube design. For each, a single synthetic data set was generated, with both a fixed- and changing-evidence condition, each with 1000 trials. For the fixed-evidence assessments, we only fit the models to the fixed evidence condition of the data.

For the changing-evidence condition, we simulated data with a single change in evidence that occurred after 250 ms, corresponding to a direct swap in the drift rates for the two accumulators in the LBA and a swap of the positive/negative sign in the DDM/UGM (i.e., both are symmetric changes in evidence). Only parameter sets that produced data sets with reasonable response time distributions were kept for further analysis. Our criteria for reasonable response time distributions were the same for all simulations for all models, and can be found in the [Appendix](#) along with the number of data sets that failed to meet these criteria. Note that it is possible that our choice of exclusion criteria could have influenced our recovery findings, as models that successfully recover parameter sets used in this study may not be able to recover parameter sets excluded. In order to focus on the practical use case of these models, however, we chose to only study the region of “data space” that is relevant

to applications and thus do not consider RT distributions that differ dramatically from those observed empirically.

We fit datasets using Bayesian parameter estimation with differential evolution Markov chain Monte Carlo (DE-MCMC; Ter Braak, 2006; Turner, Sederberg, Brown, & Steyvers, 2013) to estimate the posterior distributions. Since these models do not have tractable analytic likelihood functions, we utilized the recently developed probability density approximation (PDA) method to computationally approximate the needed likelihood functions (Turner & Sederberg, 2014; Holmes, 2015; Evans, Holmes, & Trueblood, 2019), which we detail further in the [Appendix](#). To assess the accuracy of the parameter recovery, we compared the mean posterior estimated parameter values to the generating parameter values across all datasets for each parameter. Our fits were performed to (1) the constant evidence data (labeled “constant” in the figures of the results) and (2) the constant and changing evidence data simultaneously (labeled “changing” in the figures of the results) with all parameters constrained over conditions (e.g., threshold parameters are the same for both types of trials). See the [Appendix](#) for additional details regarding the generation of synthetic data and parameter recovery methods.

Results

Figure 2 shows results of the recovery for the LBA model, Figs. 3 (linear) and 4 (Weibull) show recovery results for

Fig. 2 Parameter recovery plots for the LBA with linear collapsing thresholds. The *columns* display the different parameters of the model (drift difference, drift mean, threshold, and collapse rate, respectively), and the *rows* display the different conditions (constant evidence, and

both changing and collapsing evidence, respectively). Each plot displays the generated values on the *x*-axis, and the mean of the estimated posteriors on the *y*-axis, with the generated/estimated combinations as *black points*, and the *red line* represents perfect recovery

Fig. 3 Parameter recovery plots for the DDM with linear collapsing thresholds. The *columns* display the different parameters of the model (drift rate, threshold, and collapse rate, respectively), and the *rows* display the different conditions (constant evidence, and both changing and collapsing evidence, respectively). Each plot displays the generated values on the *x*-axis, and the mean of the estimated posteriors on the *y*-axis, with the generated/estimated combinations as *black points*, and the *red line* representing perfect recovery

Fig. 4 Parameter recovery plots for the DDM with Weibull collapsing thresholds. The *columns* display the different parameters of the model (drift rate, threshold, the shape of the Weibull collapse, and the scale of the Weibull collapse, respectively), and the *rows* display the different conditions (constant evidence, and both changing and collapsing

evidence, respectively). Each plot displays the generated values on the *x*-axis, and the mean of the estimated posteriors on the *y*-axis, with the generated/estimated combinations as *black points*, and the *red line* representing perfect recovery

the DDM models, and Figs. 5 (filter) and 6 (leakage) show recovery results for the UGM models. In each case, the horizontal axes show the true value of the generating parameters while the vertical axes show the mean of the resulting posterior distributions. Thus, a quality recovery is characterized by a cloud of points that lie close to the diagonal (true parameter = fit parameter).

In general, there appears to be three consistent trends in the parameter recoveries across all of the models. Firstly, the difference in drift rate between alternatives (which is just the drift rate in the DDM and UGM models) recovers extremely well, regardless of the time-variant function. Thus the discriminability or strength of evidence is recoverable in all cases, regardless of the model or type of data used. Secondly, there appears to be a clear link between the recovery of the initial threshold and the recovery of the time-variant parameters (i.e., the threshold collapse, or urgency signal), where time-variant and initial threshold parameters are either both recovered well, or both recovered poorly. Specifically, both recover well in the linear $CTDDM$ while neither recover well in the non-linear $CTDDM$ or $CTLBA$. This suggests that a key tradeoff exists between the initial threshold and these time-variant parameters, and that estimates of initial thresholds are only reliable within certain time-variant models. Lastly, the addition of

a changing evidence condition appears to aide the recovery of the parameters for all models, though in most cases the improvement is fairly minor.

At a specific level, while the $CTLBA$ accurately recovers the difference between drift rates, it cannot recover the values of the drift rates themselves, the initial threshold, or the rate of threshold collapse. While the non-linear $CTDDM$ can very precisely recover the drift rate, it cannot recover any of the threshold parameters adequately, with either fixed or changing evidence. The linear $CTDDM$ on the other hand precisely recovers all parameters including the drift rate, the initial threshold, and the rate of threshold collapse. This is true when either fixed or changing evidence is considered, though the addition of changing evidence does somewhat improve the quality of recovery.

The UGM results are somewhat more nuanced. Recovery of the evidence strength parameter (v) is precise in both cases. Estimation of the time constant/leakage rate is reasonable for both models as well, in the appropriate regime. Small time constants and large leakage rates both recover well, while large time constants/small leakage rates do not. In retrospect, this is sensible because large time constants/small leakage rates correspond to times that are longer than decision times themselves. More importantly, while the threshold is recoverable in the

Fig. 5 Parameter recovery plots for the UGM with a filter process. The *columns* display the different parameters of the model (drift rate, threshold, the urgency multiplier, and the time-constant of the filter process, respectively), and the *rows* display the different conditions (constant evidence, and both changing and collapsing evidence,

respectively). Each plot displays the generated values on the x-axis, and the mean of the estimated posteriors on the y-axis, with the generated/estimated combinations as *black points*, and the *red line* representing perfect recovery

Fig. 6 Parameter recovery plots for the UGM with a leakage process. The *columns* display the different parameters of the model (drift rate, threshold, and the leakage rate, respectively), and the *rows* display the different conditions (constant evidence, and both changing and collapsing evidence, respectively). Each plot displays the generated values on the *x*-axis, and the mean of the estimated posteriors on the *y*-axis, with the generated/estimated combinations as *black points*, and the *red line* representing perfect recovery

leakage formulation, neither the threshold nor the urgency value (u) are recoverable in the filter version, with either fixed or changing information.

These results indicate that the only two models (of those considered) for which accurate recovery of parameters is possible are the linear c_T DDM and the leakage formulated UGM. Importantly however, the leakage formulated UGM does not contain an estimable “strength of urgency parameter” as it was shown mathematically to not be identifiable. Thus the only one of the five considered models containing a parameter estimating the relative importance of time-varying caution/urgency that can be accurately inferred is the c_T DDM.

An important distinction is needed here though. Just because a model does not recover its parameters does not preclude its use in modeling. The inability of a model to recover its own parameters is typically due to some form of parameter degeneracy or indeterminacy, otherwise referred to as model “sloppiness” (Holmes, 2015; Gutenkunst et al., 2007). In such cases, this simply means that more than one parameter set can account equally, or almost equally well for the same data. That is, the ability of a model to accurately fit a data set, and the ability of that model to precisely recover parameters are different. In the case of the three models that do not recover parameters, the “best fit” parameters do still fit the generated data very well (results not shown). It is simply the case that for those models, many parameter sets account for the data very well and the fitting process

converged onto a different set than those that generated the data.

Discussion

Our study aimed to provide the first comprehensive assessment of whether response time models that included time-variant collapsing thresholds or urgency signals can be used as “measurement tools” to reliably estimate the latent parameter values of interest. Previously within the collapsing thresholds literature, the goal of parameter estimation has mostly been ignored in favor of model inference, although both goals should be viewed as important. We assessed five different models through parameter recovery simulations, which looked at whether the drift rate, threshold, and time-varying threshold/urgency parameters could be recovered for each. We performed these parameter recoveries within a constant-evidence only paradigm—the standard within rapid decision-making—and a paradigm with both constant-evidence and changing-evidence conditions.

Our results demonstrate that where parameter inferences are considered, there are substantial differences between the caution/urgency modulation models assessed here. Specifically, only two of these models, the diffusion decision model with a linearly collapsing threshold and the urgency gating model formulated with a leakage process,

are able to accurately recover their own parameters. Of those two, however, only the $CTDDM$ has an estimable parameter reflecting the relative importance of time-varying caution/urgency. While the leakage formulated UGM is recoverable, this is only the case once the “strength of urgency” parameter (u) is removed, since it is mathematically unidentifiable based on response time data. While other models are capable of recovering some of their parameters, only the $CTDDM$ is capable of recovering the critical parameters associated with caution.

One other interesting finding of our recovery study was that changing evidence conditions appear to aide the recovery of parameters for all models. Importantly, this finding may have potential implications beyond the time-varying models that we assess within this study, to other decision-making models that have previously shown recovery issues. For example, the study of Miletic et al. (2017) found that the leaky-competing accumulator (LCA; Usher and McClelland 2001), a neurally motivated model of decision-making, exhibits parameter recovery issues. As the LCA naturally extends to changing-evidence paradigms, the additional constraint may improve the recovery of the parameters, as it did with the models in our study. However, it should be noted that the improvements we observed were fairly minor, and no model parameters qualitatively shifted from “non-recoverable” to “recoverable” due to the addition of changing evidence. Regardless, future research could potentially investigate the recovery of the parameters of the LCA—as well as other difficult to recover models—within changing-evidence paradigms.

Based on our findings, we make the following recommendations to researchers who wish to investigate or apply models with time-variant components. Firstly, we recommend that if parameter estimation is the goal, then researchers use either the $CTDDM$ with a linear collapse, or the UGM with a leakage process. If the relative strength of time-varying caution/urgency are of primary importance, then the $CTDDM$ should be used. Alternatively, if another type of time-variant model is of key theoretical interest, like a $CTDDM$ with a Weibull collapse, we recommend that researchers stick to the goal of model inference, that is contrasting different models based on how well they fit data, rather than basing conclusions on parameters, since the parameter estimates from these models do not appear to be robust. In all cases, we strongly recommend that proper parameter recovery assessments are performed *before* using the parameters of any time-varying model for inference, as we have shown that small changes to the formulation of a model can result in a lack of robustness of the parameter estimates.

Appendix A: Criteria for inclusion of a data set

All generated datasets were required to meet the following criteria to be considered a reasonable representation of a typical RT data set, and therefore, be included in the recovery assessment:

In both conditions :

$$0.4 < \text{Mean RT} < 2.5$$

$$0.4 < \text{Median RT} < 2.5$$

$$0.1 < \text{IQR of RT} < 2$$

$$\text{Minimum RT} < 1.5$$

$$1.5 < \text{Maximum RT}$$

In constant evidence :

$$0.05 < \text{Maximum RT} < 1$$

In changing evidence :

$$0.05 < \text{Maximum RT} < 0.95$$

For the 4000 generated data sets for each model, 2166 met the criteria for the $CTLBA$ (i.e., 54.2% included), 1239 met the criteria for the $CTDDM$ with a linear collapse (31%), 1030 met the criteria for the $CTDDM$ with a Weibull collapse (25.8%), 892 met the criteria for the UGM with a filter process (22.3%), and 437 met the criteria for the UGM with a leakage process (11%).

Appendix B: Specific details for the Latin hypercube and Bayesian parameter estimation for each model

For the Latin hypercube design, standard parameters of the models were sampled from uniform distributions, and parameters relating to the threshold collapse, urgency, or leakage were sampled from an exponential distribution. Our reason for choosing an exponential distribution for these parameters was to ensure equal density sampling of small and large values over multiple orders of magnitude of these parameters. Uniform sampling will significantly oversample large values of parameters at the expense smaller values of a parameter. For leakage for example, small/large values correspond to weak/strong leakage and

thus it is critical to test recovery in both regimes. Similarly, small/large values of urgency and leakage parameters have psychologically different meanings. To sample from the exponential distribution, we sampled from a $U[0, 1]$, multiplied the values by the natural logarithm of the maximum of the parameter range divided by the minimum of the parameter range, added the logarithm of the minimum of the range, and took the exponent (i.e., $\exp(\log(\frac{\max}{\min}) + \log(\frac{\min}{\max}) \times U[0, 1])$).

DDM

For the $CTDDM$ with a linear collapse, we used the following parameter ranges for the Latin hypercube design:

v	[0.01, 0.6]
$a_{initial}$	[0.01, 0.6]
a_{slope}	[0.001, 0.5]
ter	0.3
t_{delay}	0.2
z	0
s	0.1

and the following prior distributions for the Bayesian parameter estimation:

$$\begin{aligned} v &\sim N(.3, .3) \\ a_{initial} &\sim TN(.2, .2, 0, \infty) \\ a_{slope} &\sim TN(.3, .3, 0, \infty) \\ ter &\sim TN(.6, .6, 0, \infty) \\ t_{delay} &\sim TN(.1, .1, 0, \infty) \end{aligned}$$

where z was fixed to 0, and s was fixed to 0.1.

For the $CTDDM$ with a Weibull function collapse, we used the following parameter ranges for the Latin hypercube design:

v	[0.01, 0.6]
$a_{initial}$	[0.01, 0.6]
$shape$	[0.01, 10]
$scale$	[0.01, 10]
ter	0.3
t_{delay}	0.2
z	0
s	0.1

where $shape$ and $scale$ were sampled from an exponential distribution for the Latin hypercube, rather than the standard uniform. The following prior distributions for the Bayesian parameter estimation:

$$\begin{aligned} v &\sim N(.3, .3) \\ a_{initial} &\sim TN(.2, .2, 0, \infty) \\ shape &\sim TN(.5, .5, 0, \infty) \\ scale &\sim TN(.5, .5, 0, \infty) \\ ter &\sim TN(.6, .6, 0, \infty) \\ t_{delay} &\sim TN(.1, .1, 0, \infty) \end{aligned}$$

where z was fixed to 0, and s was fixed to 0.1.

LBA

For the $CTLBA$, we used the following parameter ranges for the Latin hypercube design:

v_{mean}	[2.5, 5.5]
$v_{difference}$	[0.05, 2]
$B_{initial}$	[0.4, 3]
B_{slope}	[0.05, 1.5]
A	1
t_0	0.3
t_{delay}	0.2
s	1

and the following prior distributions for the Bayesian parameter estimation:

$$\begin{aligned} v_{correct} &\sim TN(3, 3, 0, \infty) \\ v_{error} &\sim TN(1, 1, 0, \infty) \\ B_{initial} &\sim TN(2, 2, 0, \infty) \\ B_{slope} &\sim TN(.5, .5, 0, \infty) \\ A &\sim TN(2, 2, 0, \infty) \\ t_0 &\sim TN(.5, .5, 0, \infty) \\ t_{delay} &\sim TN(.1, .1, 0, \infty) \end{aligned}$$

where s was fixed to 1.

UGM

For the UGM with a filter process, we used the following parameter ranges for the Latin hypercube design:

v	[0.01, 0.6]
a_{initial}	[0.01, 0.6]
$urgency$	[0.5, 3]
$time - constant$	[0.1, 3]
ter	0.3
z	0
s	0.1

where *urgency* and *time – constant* were sampled from an exponential distribution for the Latin hypercube, rather than the standard uniform. The following prior distributions for the Bayesian parameter estimation:

$$\begin{aligned} v &\sim N(.3, .3) \\ a_{\text{initial}} &\sim TN(.2, .2, 0, \infty) \\ urgency &\sim TN(1, 1, 0, \infty) \\ time - constant &\sim TN(.3, .3, 0, \infty) \\ ter &\sim TN(.6, .6, 0, \infty) \end{aligned}$$

where z was fixed to 0, and s was fixed to 0.1.

For the UGM with a leakage process, we used the following parameter ranges for the Latin hypercube design:

v	[0.01, 0.6]
a_{initial}	[0.01, 0.6]
L	[0.1, 10]
ter	0.3
z	0
s	0.1

where L was sampled from an exponential distribution for the Latin hypercube, rather than the standard uniform. The following prior distributions for the Bayesian parameter estimation:

$$\begin{aligned} v &\sim N(.3, .3) \\ a_{\text{initial}} &\sim TN(.2, .2, 0, \infty) \\ L &\sim TN(2, 2, 0, \infty) \\ ter &\sim TN(.6, .6, 0, \infty) \end{aligned}$$

where z was fixed to 0, and s was fixed to 0.1.

Appendix C: Unidentifiability of urgency parameter in the leakage formulated UGM

Here we briefly discuss why the urgency parameter (u) in the leakage formulated UGM (7) is unidentifiable and thus why it is not considered in the recovery process here. The two equations comprising (7) can be combined using the product rule along with substitution of variables. The resulting single SDE describing the leakage formulated UGM becomes

$$dx = \left[\left(\frac{1}{t} - L \right) x + t(uv) \right] dt + t(u\sigma)dw, \quad (8)$$

where we have intentionally grouped the terms (uv) and $(u\sigma)$. Thus the parameter u effectively serves only to rescale the strength of evidence. This unidentifiability can be seen in two ways. Most directly, a change of variables $x = uv$ for this equation leads to

$$dy = \left[\left(\frac{1}{t} - L \right) y + tv \right] dt + t\sigma dw. \quad (9)$$

This simple change of variables completely removes u from the equation, indicating it is not identifiable. Second, the parameters v and σ typically represent the mean and moment-to-moment variance in the sensory evidence, which is assumed to take the form $N(v, \sigma^2)$. From this point of view, replacing v with (uv) and σ with $(u\sigma)$ simply leads to a rescaling of the strength of evidence to $N(uv, (u\sigma)^2)$. Based on this analysis, we conclude that u is not identifiable in this formulation of the UGM and we thus do not consider its recovery. Instead, we effectively work with the rescaled (9).

Appendix D: Fitting through probability density approximation (PDA)

Given the use of the changing information paradigm in this study, none of the models used here have an analytic likelihood. We thus fit each model using probability density approximation (PDA; Turner & Sederberg 2014; Holmes 2015). Specifically, PDA involves simulating a large number of trials (in our case 10,000) from the model of interest, and creating a pseudo-likelihood from these simulations using kernel density estimation. This method differs from more common methods of fitting through simulation, such as χ^2 (e.g., White et al. 2018) and QMLE (e.g., Hawkins et al. 2015, b), which attempt to minimize the discrepancy between response time quantiles for the

empirical data and simulated data. In contrast, the PDA approximates the actual likelihood of the observed data given the particular set of parameters. This allows the use of standard MCMC methods to perform Bayesian parameter estimation.

References

- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178.
- Carland, M. A., Marcos, E., Thura, D., & Cisek, P. (2015). Evidence against perfect integration of sensory information during perceptual decision making. *Journal of Neurophysiology*, 115(2), 915–930.
- Carland, M. A., Thura, D., & Cisek, P. (2015). The urgency-gating model can explain the effects of early evidence. *Psychonomic Bulletin & Review*, 22(6), 1830–1838.
- Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *The Journal of Neuroscience*, 29(37), 11560–11571.
- Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, 41(3), 260–274.
- Diederich, A. (2008). A further test of sequential-sampling models that account for payoff effects on response bias in perceptual decision tasks. *Perception & Psychophysics*, 70(2), 229–256.
- Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic-dominance principle of decision making. *Psychological Science*, 10(4), 353–359.
- Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics*, 68(2), 194–207.
- Ditterich, J. (2006). Evidence for time-variant decision making. *European Journal of Neuroscience*, 24, 3628–3641.
- Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks*, 19(8), 981–1012.
- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, 18(1), 61–69.
- Donkin, C., & Brown, S. D. (2017). Response time modeling. The Stevens' handbook of experimental psychology and cognitive neuroscience.
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience*, 32(11), 3612–3628.
- Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., & et al. (2018). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, 1–19.
- Evans, N. J., & Brown, S. D. (2017). People adopt optimal policies in simple decision-making, after practice and guidance. *Psychonomic Bulletin & Review*, 24, 597–606.
- Evans, N. J., & Hawkins, G. E. (2019). When humans behave like monkeys: Feedback delays and extensive practice increase the efficiency of speeded decisions. *Cognition*, 184, 11–18.
- Evans, N. J., Hawkins, G. E., Boehm, U., Wagenmakers, E. J., & Brown, S. D. (2017). The computations that support simple decision-making: A comparison between the diffusion and urgency-gating models. *Scientific Reports*, 7, 16433.
- Evans, N. J., Hawkins, G. E., & Brown, S. D. (2018). The role of passing time in decision-making. Retrieved from <https://psyarxiv.com/3wq6g/>.
- Evans, N. J., Holmes, W. R., & Trueblood, J. S. (2019). Response time data provides critical constraints on dynamic models of multi-alternative, multi-attribute choice. *Psychonomic Bulletin and Review*.
- Evans, N. J., Rae, B., Bushmakina, M., Rubin, M., & Brown, S. D. (2017). Need for closure is associated with urgency in perceptual decision-making. *Memory & Cognition*, 45, 1193–1205.
- Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E. J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed–accuracy tradeoff in the elderly brain: A structural model-based approach. *The Journal of Neuroscience*, 31(47), 17242–17249.
- Frazier, P. I., & Yu, A. J. (2007). Sequential hypothesis testing under stochastic deadlines. In *NIPS* (465–472).
- Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, 136(3), 389.
- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10), e189.
- Hawkins, G. E., Forstmann, B. U., Wagenmakers, E. J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience*, 35(6), 2476–2484.
- Hawkins, G. E., Marley, A., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive Science*, 38(4), 701–735.
- Hawkins, G. E., Wagenmakers, E. J., Ratcliff, R., & Brown, S. D. (2015). Discriminating evidence accumulation from urgency signals in speeded decision making. *Journal of Neurophysiology*, 114(1), 40–47.
- Ho, T. C., Yang, G., Wu, J., Cassey, P., Brown, S. D., Hoang, N., & et al. (2014). Functional connectivity of negative emotional processing in adolescent depression. *Journal of Affective Disorders*, 155, 65–74.
- Holmes, W. R. (2015). A practical guide to the probability density approximation (PDA) with improved implementation and error characterization. *Journal of Mathematical Psychology*, 68, 13–24.
- Holmes, W. R., & Trueblood, J. S. (2018). Bayesian analysis of the piecewise diffusion decision model. *Behavior Research Methods*, 50(2), 730–743.
- Holmes, W. R., Trueblood, J. S., & Heathcote, A. (2016). A new framework for modeling decisions about changing information: The piecewise linear ballistic accumulator model. *Cognitive Psychology*, 85, 1–29.
- Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12), 3017–3029.
- Kruschke, J. K., & Liddell, T. M. (2015). The Bayesian new statistics: Two historical trends converge. SSRN Electronic Journal.
- Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology*, 7, 1324.
- Matzke, D., Dolan, C. V., Logan, G. D., Brown, S. D., & Wagenmakers, E. J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. *Journal of Experimental Psychology: General*, 142(4), 1047.

- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239–245.
- Miletić, S., Turner, B. M., Forstmann, B. U., & van Maanen, L. (2017). Parameter recovery for the leaky competing accumulator model. *Journal of Mathematical Psychology*, 76, 25–50.
- Palestro, J. J., Weichart, E., Sederberg, P. B., & Turner, B. M. (2018). Some task demands induce collapsing bounds: Evidence from a behavioral analysis. *Psychonomic Bulletin & Review*, 25(4), 1225–1248.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347–356.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4), 260–281.
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, 60(3), 127–157.
- Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, 140(3), 464.
- Stone, M. (1960). Models for choice–reaction time. *Psychometrika*, 25, 251–260.
- 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.
- Thura, D., Beauregard-Racine, J., Fradet, C. W., & Cisek, P. (2012). Decision making by urgency gating: Theory and experimental support. *Journal of Neurophysiology*, 108(11), 2912–2930.
- Trueblood, J. S., Holmes, W. R., Seegmiller, A. C., Douds, J., Compton, M., & Szentirmai, E. (2018). The impact of speed and bias on the cognitive processes of experts and novices in medical image decision-making. *Cognitive Research: Principles and Implications*, 3(1), 28. <https://doi.org/10.1186/s41235-018-0119-2>
- Tsetos, K., Gao, J., McClelland, J. L., & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, 6, 79.
- Tsetos, K., Usher, M., & McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in Neuroscience*, 5, 63.
- Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, 21(2), 227–250.
- Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, 18(3), 368.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550.
- van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: Ez, fast-dm, and DMAT. *Journal of Mathematical Psychology*, 53(6), 463–473.
- Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, 73, 59–79.
- White, C. N., Servant, M., & Logan, G. D. (2018). Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic Bulletin & Review*, 25(1), 286–301.
- Winkel, J., Keuken, M. C., van Maanen, L., Wagenmakers, E. J., & Forstmann, B. U. (2014). Early evidence affects later decisions: Why evidence accumulation is required to explain response time data. *Psychonomic Bulletin & Review*, 21(3), 777–784.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.