

## CALL FOR PAPERS | *Decision Making: Neural Mechanisms*

# Discriminating evidence accumulation from urgency signals in speeded decision making

Guy E. Hawkins,<sup>1</sup> Eric-Jan Wagenmakers,<sup>1</sup> Roger Ratcliff,<sup>2</sup> and Scott D. Brown<sup>3</sup>

<sup>1</sup>Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands; <sup>2</sup>Department of Psychology, The Ohio State University, Columbus, Ohio; and <sup>3</sup>School of Psychology, University of Newcastle, Callaghan, New South Wales, Australia

Submitted 27 January 2015; accepted in final form 17 April 2015

**Hawkins GE, Wagenmakers EJ, Ratcliff R, Brown SD.** Discriminating evidence accumulation from urgency signals in speeded decision making. *J Neurophysiol* 114: 40–47, 2015.—First published April 22, 2015; doi:10.1152/jn.00088.2015. The dominant theoretical paradigm in explaining decision making throughout both neuroscience and cognitive science is known as “evidence accumulation”—the core idea being that decisions are reached by a gradual accumulation of noisy information. Although this notion has been supported by hundreds of experiments over decades of study, a recent theory proposes that the fundamental assumption of evidence accumulation requires revision. The “urgency gating” model assumes decisions are made without accumulating evidence, using only moment-by-moment information. Under this assumption, the successful history of evidence accumulation models is explained by asserting that the two models are mathematically identical in standard experimental procedures. We demonstrate that this proof of equivalence is incorrect, and that the models are not identical, even when both models are augmented with realistic extra assumptions. We also demonstrate that the two models can be perfectly distinguished in realistic simulated experimental designs, and in two real data sets; the evidence accumulation model provided the best account for one data set, and the urgency gating model for the other. A positive outcome is that the opposing modeling approaches can be fruitfully investigated without wholesale change to the standard experimental paradigms. We conclude that future research must establish whether the urgency gating model enjoys the same empirical support in the standard experimental paradigms that evidence accumulation models have gathered over decades of study.

decision-making; response time; mathematical model; urgency gating; evidence accumulation

THE STUDY OF DECISION-MAKING has over 50 years of history that crosses disciplinary lines, with particularly important contributions from cognitive science and neuroscience. The dominant theoretical paradigm explains decision-making using “accumulator” or “diffusion” models, which assume that noisy information is gradually sampled from the environment (Laming 1968; Link and Heath 1975; Ratcliff 1978; Stone 1960). This information is accumulated in an evidence counter that tracks support for one choice option over another. The process continues until support for one option or another reaches a threshold amount, triggering a choice. In hundreds of human studies

with thousands of participants, diffusion models have described many aspects of the data and provided insight into many theoretical and practical research areas, including decisions about motion stimuli, consumer goods, sleep deprivation, aging, and psychopharmacology (e.g., Krajchich and Rangel 2011; Palmer et al. 2005; Ratcliff et al. 2004; Ratcliff and Van Dongen 2011; Van Ravenzwaaij et al. 2012). The same models have been used to understand the neural structures that underpin decision-making, in both humans (Ditterich 2006; Forstmann et al. 2008, 2010; O’Connell et al. 2012; Ratcliff et al. 2009; Schurger et al. 2012) and other primates (Ding and Gold 2012; Hanes and Schall 1996; Kiani and Shadlen 2009; Maimon and Assad 2006; Purcell et al. 2010, 2012; Ratcliff et al. 2003a, 2007; Roitman and Shadlen 2002).

Recently, this understanding of decision-making based on evidence accumulation models (EAMs) has been radically revised by an interesting proposal from Cisek et al. (2009) and Thura et al. (2012). The core of the revision is the “urgency gating model” (UGM), which drops the central component of the EAMs, by assuming that environmental evidence is not accumulated over time. Instead, the UGM passes novel sensory information, which varies from moment-to-moment, through a low-pass filter. The low-pass filtered information is multiplied by an urgency signal that grows with decision time, and then these multiplied samples are monitored until any sample exceeds a decision threshold. The UGM is an original and insightful proposal that has already had important impacts on the field (for similar approaches see Hockley and Murdock 1987, and accompanying critique from Gronlund and Ratcliff 1991).

A critical question is why the standard EAMs have had such a long history in fitting real data, if they are so fundamentally wrong. An explanation for this was given by Thura et al. (2012), who gave a mathematical proof that the EAM and UGM are identical, whenever the stimulus environment does not change during the course of a decision. That is, as long as the decision stimulus is relatively constant, such as a static image, a stationary sound, or a random dot kinematogram with constant coherence, then the EAM’s success can be explained by its perfect mimicry of the UGM. This explanation is particularly powerful because almost all experiments used to support the EAM framework over the past 50 years have used

Address for reprint requests and other correspondence: G. E. Hawkins, Dept. of Psychology, Univ. of Amsterdam, Nieuwe Achtergracht 129, 1018 WS Amsterdam, The Netherlands (e-mail: guy.e.hawkins@gmail.com).

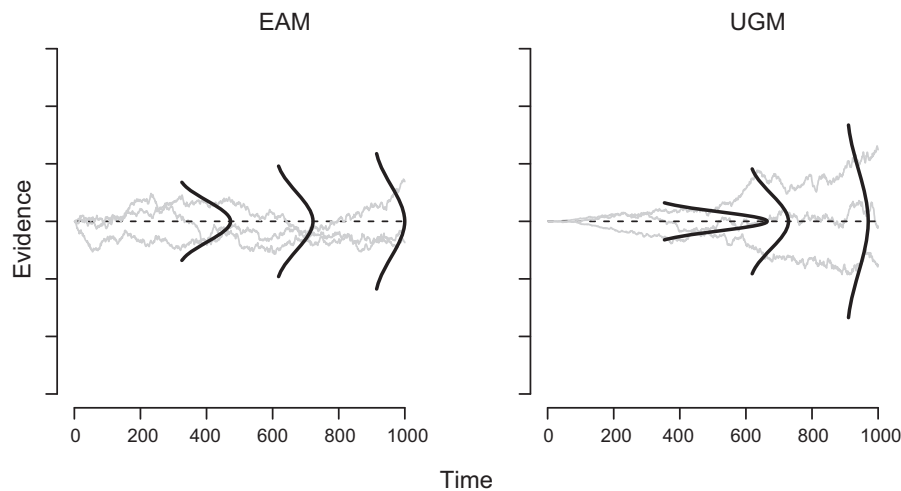


Fig. 1. Sample paths (gray lines) for an evidence accumulation model (left panel) and an urgency gating model (right panel), both with zero drift. The three black distribution lines on each panel show the distribution of sample paths across trials at times  $t = 300, 600$ , and  $900$ . There are no units on the y-axes because the models' evidence units are arbitrary. EAM, evidence accumulation model; UGM, urgency gating model.

just these time-constant stimuli (although there have been exceptions, reviewed in DISCUSSION).

We investigate more carefully the claim that the EAM and UGM cannot be distinguished. First, we show that the mathematical proof of this equivalence provided by Thura et al. (2012) is incorrect, and hence that the UGM and EAM make different predictions when the decision stimulus is constant. Second, we report simulation studies and an application to real data that illustrate the models can be easily distinguished in practice, even in data sets of practical sizes, and even with time-constant decision stimuli.

*The proof of equivalence between the evidence accumulation model and the urgency gating model is incorrect.* The key elements of this proof of equivalence occur in Equations 28 and 29 on page 2918 from Thura et al. (2012). Those equations provide distributions for sample paths from an EAM and UGM, respectively, in the absence of decision thresholds.<sup>1</sup> The two equations are identical except for the last term of each, which for the EAM (Equation 28) is  $\int_0^t \theta$  and for the UGM (Equation 29) is  $t \times \int_0^t \theta$ , where  $\theta$  represents a Gaussian distributed noisy stimulus representation. Thura et al. (2012) note that the integral in this term “quickly goes to zero in both cases because the noise has a mean of zero,” and conclude that the two models “behave nearly identically.”

The integral in question ( $\int_0^t \theta$ ) is an Itô integral of identically and independently distributed normal increments. The properties of this integral have been studied extensively (Feller 1968), originally for application to Brownian motion, but later in many different variants for psychological application (for a mathematically-focused review, see Smith 2000). This integral defines a family of normal distributions that change with time  $t$ . As Thura et al. (2012) note, the mean of these distributions is exactly zero, for all  $t$ . However, and crucially, the variance of the distributions grows linearly with  $t$ .

Returning to the argument of Thura et al. (2012) about the EAM and UGM, the two terms that discriminate Equations 28 and 29 are not identical. The term from Equation 28 produces normal distributions whose variance grows with  $t$ , while the term from Equation 29 produces normal distributions whose

variance grows with  $t^3$ . The difference between  $t$  and  $t^3$  becomes very large, very quickly. Indeed, rather than “quickly [going] to zero in both cases,” these final terms will quickly come to dominate Equations 28 and 29, producing very different model predictions.<sup>2</sup>

Figure 1 illustrates this effect. The left panel shows sample paths from the noise process of a diffusion model, from Equation 28 of Thura et al. (2012). Those sample paths were generated from a random walk simulation of a Brownian motion process, in which each time step of size  $\Delta$  adds a random Gaussian increment with zero mean and standard deviation  $\sqrt{\Delta}$ . The right panel shows samples from the noise process of an urgency gating model, from Equation 29 of Thura et al. (2012), with zero drift; sample paths from this model are multiplied by an urgency signal linearly related to time. The variance parameters of the two models were set such that they lead to identical distributions of sample paths for the two models at time  $t = 600$  (so the middle of the three distributions shown in each panel are identical for the EAM and UGM). The differences between the models are apparent in the distributions of sample paths before and after that time. The standard deviation of the EAM grows with the square root of time, while for the UGM it grows with  $t^{3/2}$ . This leads to larger variance for the EAM at early time points, with the UGM taking over at later times.

## METHODS

Even though the above argument makes it clear that the EAM and UGM are not mathematically identical, it still may be the case that they are similar enough to be practically identical. That is, perhaps Cisek et al. (2009) and Thura et al. (2012) were correct in the sense that the models are impossible to distinguish in finite data samples from realistic experiments with time-constant stimuli. We investigated this question using a model recovery simulation (Navarro et al. 2004; Wagenmakers et al. 2004). This procedure involves simulating data sets from one model (e.g., the EAM) and fitting those simulated data sets with both the EAM and UGM, to see which model fits best. If the models are discriminable, then data generated by the EAM will be better fit by the EAM than by the UGM, and vice versa.

<sup>1</sup> Beyond the error in Equation 29 we describe here, Equation 29 does not follow from the model specified by Thura et al. (2012), because the UGM assumes a finite time constant for its low-pass filter. We thank an anonymous reviewer for pointing out this additional error.

<sup>2</sup> The finite time constant used in the UGM, not considered in Equations 28 and 29 of Thura et al. (2012), influences these rates of growth further. We take this into account carefully in the simulation analyses below.

Table 1. Parameters used to generate data sets from the EAM and UGM for the model recovery simulation study

Model	$v_{1-4}$	$aL, aU$	$z$	$t_{er}$	$s$	Time Scale	Time Constant
EAM	<b>0.03, 0.08, 0.14, 0.23</b>	<b>0, 130</b>	$aU/2$	<b>300</b>	3.16	milliseconds	—
UGM	<b>0.04, 0.11, 0.18, 0.28</b>	$-aU, 126$	0	<b>300</b>	3.16	milliseconds	100

Boldface indicates parameters that were freely estimated from data when the models were fit to the synthetic data sets; regular face indicates parameters that were fixed (not estimated from data). Abbreviations:  $v$ , drift rate;  $aL, aU$ , lower and upper boundary;  $z$ , start point;  $t_{er}$ , nondecision time. Diffusion constant ( $s$ ) refers to the SD of the within-trial Gaussian noise process,  $\sim N(0, s)$ , which was fixed at  $\sqrt{1,000}$  multiplied by the standard value of 0.1 to accommodate time scale in ms. Time constant (in ms) of the low-pass filter implemented in the urgency gating model (UGM) relates to the temporal integration period. The dash (—) indicates the absence of a time constant parameter in the evidence accumulation model (EAM).

Following Thura et al. (2012), we simulated synthetic data sets from the Stone (1960) version of the EAM. We implemented the UGM with an urgency function  $\mu(t)$  that was linear in time with a zero intercept,  $\mu(t) = \beta t$ , where  $\beta = 1$  is a scalar gain, and a low-pass filter with a time constant of 100 ms (for details see Thura et al. 2012).<sup>3</sup> For consistency with previous applications of the EAM, and with standard parameter estimation software, we fixed the standard deviation of the moment-to-moment variability in evidence strength to  $s = 0.1$ , for both models. To provide consistency with parameters reported by Thura et al. (2012) and Cisek et al. (2009), we report the final parameter estimates in time units of milliseconds. This transformation has no effect on model predictions: a distance of 100/2 units at a drift rate of 0.1 per millisecond is the same distance as 0.1/2 units at a drift rate of 0.1 per second. This results in the fixed parameter for moment-to-moment variability taking the value  $s = 0.1 \times \sqrt{1,000} = 3.16$ .

We simulated data sets that mimicked a common experimental design in the perceptual decision-making literature, with a single experimental factor (e.g., difficulty level) manipulated within subjects across a number of levels (we chose four). We selected parameter values for the EAM and UGM that led to approximately equivalent predictions in the limit of large samples: a mean response time 667 ms, and mean accuracy 60%, 75%, 87%, and 95% in the four difficulty conditions. The parameter values used to generate the data are shown in Table 1 and the predicted response time distributions, in the limit of large samples, are shown as crosses in the quantile-probability plots in Fig. 2A. Using the parameter values in Table 1, separately for the EAM and UGM we simulated data from 100 synthetic participants. Each participant's data set was of a plausible sample size for the study of perceptual decision-making, with 200 trials per difficulty condition for a total of 800 trials. The realistic sample sizes per condition meant that synthetic data sets were subject to noise relative to the predicted distributions shown as crosses in Fig. 2A.

We estimated parameters for the EAM and UGM separately for each synthetic data set using quantile maximum products estimation (QMPE; Heathcote et al. 2002; Heathcote and Brown 2004). Each synthetic data set was summarized with five quantiles of the distribution of response times (10th, 30th, 50th, 70th, 90th), calculated separately for correct and incorrect responses. The QMP statistic quantifies agreement between model and data by comparing the observed and predicted proportions of data falling into each inter-quantile bin, similarly to  $G^2$  and  $\chi^2$ . We evaluated model predictions by Monte Carlo simulation, using 10,000 replicates per experimental condition during parameter estimation, and 50,000 replicates per condition to precisely evaluate predictions at the search termination point. All Monte Carlo simulations (i.e., generating synthetic data sets and estimating model parameters) used Euler's method to approximate diffusion processes as stochastic diffusion equations, with a step size of 1 ms.

We optimized goodness of fit by adjusting the model parameters using differential evolution (Ardia et al. 2013; Mullen et al. 2011). For each synthetic data set and model, we independently estimated six model parameters (shown in boldface in Table 1). We set wide boundaries on all parameters and ran 100 particles for 500 search iterations. To avoid searches terminating in local maxima, we repeated this parameter estimation exercise five times, independently, for each model fit to each synthetic data set, and chose the best set of parameters overall. The EAM and UGM had the same number of parameters freely estimated from data, so we compared their goodness of fit to each synthetic data set using the maximized value of the QMP statistic, which approximates log-likelihood.

## RESULTS

Model recovery was accurate for every single synthetic data set: data generated from the EAM was always best fit by the EAM and vice versa for the UGM. This result is shown as distributions of the difference in goodness of fit in Fig. 2B. Values above zero (vertical gray line) indicate that the EAM provided a better fit than the UGM (i.e., higher QMP result for the EAM than the UGM) and vice versa. The reason for the complete separation of the distributions in Fig. 2B can be seen in Fig. 2A: there was considerable misfit between model predictions and data when the data-generating model did not match the fitting model—a poor fit when the UGM was fit to data generated from the EAM and when the EAM was fit to data generated from the UGM (top right and bottom left panels, respectively). In contrast, there was an excellent fit to data when the fitting model matched the data-generating model (top left and bottom right panels).

The model recovery simulation study provided unequivocal results. The distributions of differences in goodness of fit were perfectly separated (Fig. 2B), which occurred because there was considerable misfit when the fitting model did not match the data-generating model (Fig. 2A). These results confirm that the EAM and UGM do not make identical predictions when environmental input is constant.

The model recovery results were clear because the EAM and UGM predict qualitatively distinct patterns of response times. As shown in Fig. 2A, the EAM predicts positively skewed response time distributions, as typically observed in human data (Luce 1986). With no further modifications, it also predicts correct and error response times that are identical (Feller 1968). In contrast, the UGM predicts a pattern of response times that are approximately normally distributed where error responses are slower than correct responses (Fig. 2A). Even more unusually, the UGM predicts a monotonic relationship between the probability of making a particular response and the associated response time; that is, the quantile probability plots predicted by the UGM are monotonically increasing from right

<sup>3</sup> Cisek et al. (2009) studied a low-pass filter with a time constant of 200 ms. The results presented here remain the same whether one assumes a shorter (100 ms) or slightly longer (200 ms) time constant.

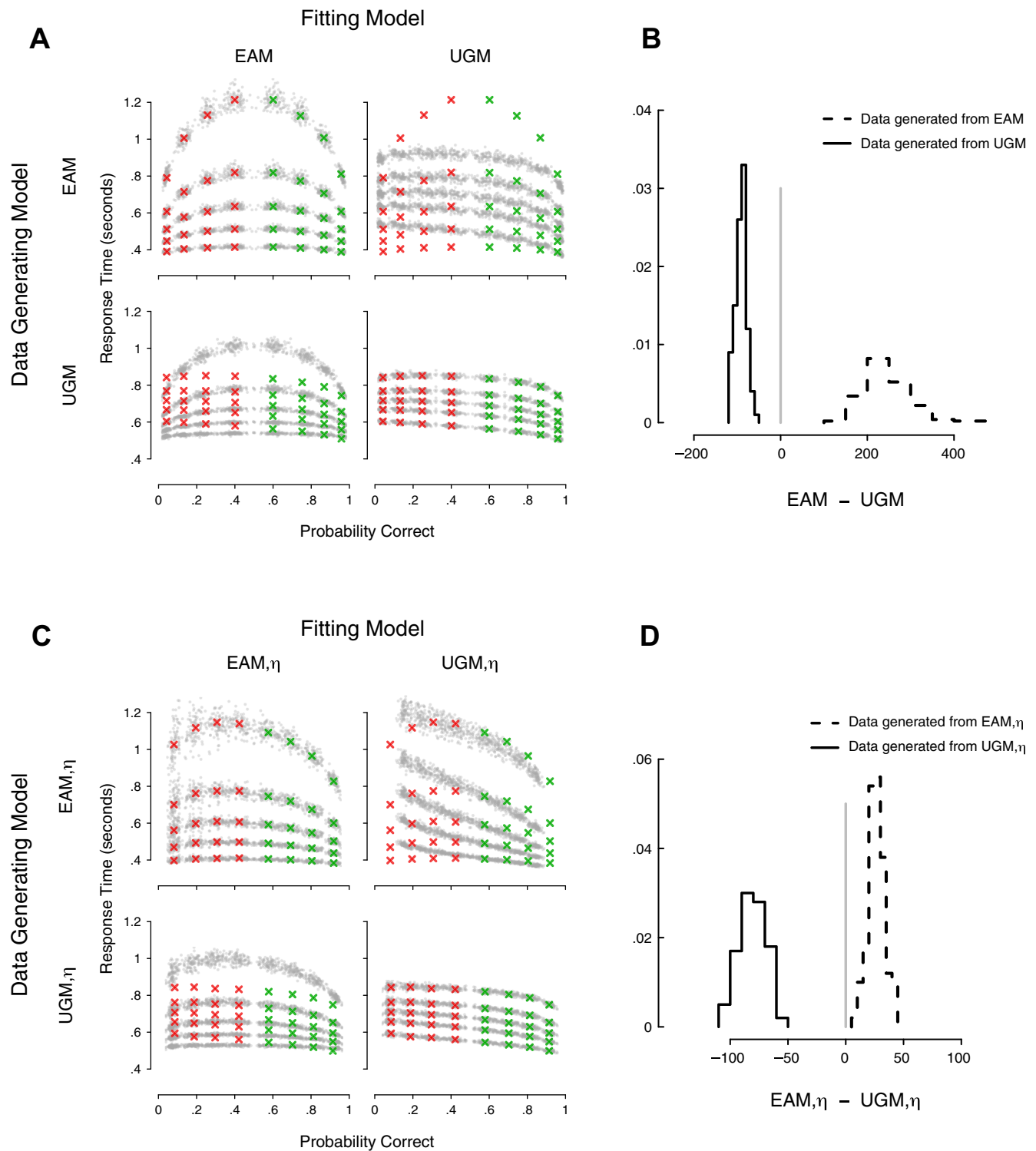
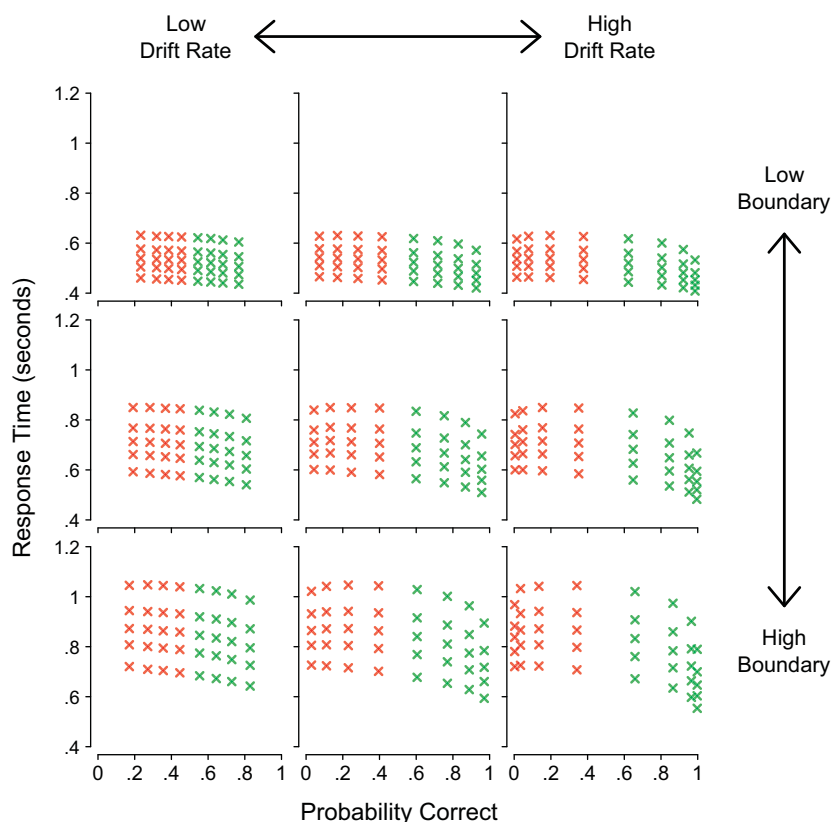


Fig. 2. Results of the model recovery analyses. *A*: quantile probability plots of goodness of fit of the EAM and UGM (left and right columns, respectively) to data generated from the EAM and UGM (top and bottom rows, respectively). Panels show the probability of a correct response on the  $x$ -axes and response time on the  $y$ -axes. Green and red crosses represent correct and error responses, respectively, across experimental conditions, simulated from the parameter values described in Table 1. Gray dots represent the predictions of the best fitting model to each synthetic data set. Vertical placement of the crosses and dots show, for each condition, the 10th, 30th, 50th (i.e., median), 70th, and 90th percentiles of the response time distribution. *B*: distributions of the difference in QMP statistics for data generated from the EAM (dashed histogram) and UGM (solid histogram). Distributions that fall to the left of zero indicate the UGM provided the best fit to data and those to the right of zero indicate the EAM provided the best fit to data. The gray vertical line shows the point where both models provided an equally good fit to data. *C* and *D*: results of the model recovery analysis that allowed trial-to-trial variability in drift rate ( $\eta$ ), using identical formatting to *A* and *B*, respectively.



Fig. 3. Quantile probability plots of UGM model predictions across a range of parameter settings. The center panel shows the UGM predictions of the parameter settings described in Table 1. Relative to the generating values used in the model recovery study, the left and right columns, respectively, show UGM predictions when drift rates are 50% smaller and 50% larger. The top and bottom rows, respectively, show UGM predictions when the boundary settings are 50% lower and 50% higher. All other details are as described in Fig. 2A. The pattern of slow errors and near-normal distributions is robust across parameter settings.



to left. This means, for example, that incorrect responses in very easy conditions (low probability responses) are slower than incorrect responses in difficult conditions. This pattern is not typically observed in data, but was observed in predictions from the UGM for all parameter settings that we investigated.

Figure 3 demonstrates this robustness to parameter settings in the UGM. The pattern of increasingly slow errors and near-normal distributions is observed across a broad range of drift rates and boundary settings.

Next, we investigated the claim that the UGM can mimic the EAM. We did this by allowing structural parameters of the UGM to be estimated as free parameters, including the time constant of the low-pass filter and the scaling parameter of the urgency function. We simulated data from the EAM using parameter settings described in Table 1, with a very large number of observations per cell. The simulated data were fit with the UGM using identical methods to the model recovery study, except that we now freely estimated the time constant of the low-pass filter and the scaling parameter ( $\beta$ ) of the urgency function, both with wide bounds. The  $\beta$  parameter was free to vary across the interval [0,5] units, and the time constant was free to vary from [0,2] seconds, which was larger than any response time in the (simulated EAM) data. Figure 4 shows that when the time constant and urgency signal parameters of the UGM are freely estimated from data, it makes highly similar predictions to the EAM: equal mean response times, when collapsed across correct and error responses (left panel, though the UGM always predicts that errors are slower than corrects), and equal mean accuracy (position along the x-axis, right panel). However, even in this very free version of the UGM, the models once again make different predictions for the shape of the response time distributions (y-axis position, right

panel). This finding further confirms that the EAM and UGM can be distinguished in standard experimental paradigms.

As a final simulation-based test of the discriminability of the EAM and UGM, we further explored the UGM prediction that error responses are slower than correct responses (Fig. 3). This is the same qualitative relationship predicted by an EAM that

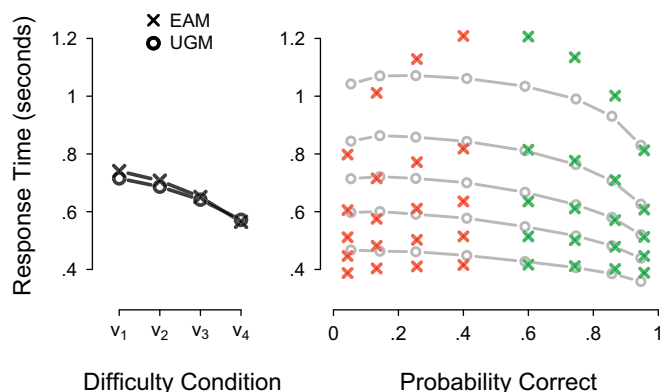


Fig. 4. The UGM closely approximates mean response time and mean accuracy predictions of the EAM when its time constant and urgency signal parameters are estimated as free parameters. Data were simulated from the EAM under the parameter settings described in Table 1, with a very large number of observations per condition. The simulated data were fit with a UGM model that freely estimated from data the time constant of the low-pass filter and the scaling parameter of the urgency signal. The left panel shows that predicted mean response times of the EAM (crosses) and UGM (circles) were almost equal across four simulated difficulty conditions (i.e., drift rates). The right panel shows quantile probability plots of data simulated from the EAM (green and red crosses) and predictions of the best-fitting UGM model (gray circles connected with lines). Under this less constrained parameterization the UGM more closely mimics the EAM (i.e., contrast the right panel here with the top right panel of Fig. 2A).

is modified with trial-to-trial variability in drift rate (Ratcliff 1978). It is therefore possible that the EAM and UGM might not be discriminable in time-constant stimuli paradigms when both are endowed with across-trial variability in drift rate. To this end, we conducted a second model recovery study with identical methods to the first, where the only difference was that both models also included trial-to-trial variability in drift rate in the form of a Gaussian distribution. In the simulated data, the means of the Gaussian were the drift rates described in Table 1 and the standard deviation (denoted  $\eta$ ) was set to approximately half the largest drift rate ( $\eta = .1$  for the EAM,  $\eta = .13$  for the UGM), a common value in the EAM literature (cf. Table 3, Matzke and Wagenmakers 2009). The crosses in Fig. 2C confirm that the value of  $\eta$  in the EAM was sufficient to predict the qualitative pattern of error responses that are slower than correct responses. When fitting the models to the simulated data, we freely estimated the  $\eta$  parameter for both the EAM and UGM, as well as the six parameters that were estimated from data in the first model recovery analysis (i.e., parameters shown in boldface in Table 1).

Model recovery was again accurate for every single synthetic data set (see Fig. 2, C and D). That is, even when endowed with across-trial noise in drift rate, data simulated from the EAM were always best fit by the EAM and vice versa for the UGM. Even though the EAM with drift rate variability predicts the qualitative pattern of slower errors than correct responses, just like the UGM, the models are perfectly discriminable because they still predict markedly different shapes for response time distributions. It is possible that the EAM and UGM may become less discriminable when across-trial noise in drift rate is much larger than within-trial noise. In some ways this is a foregone conclusion because when across-trial noise, which is external to the core model components under consideration here, dominates within-trial noise, a core component of the models, the EAM and UGM become models of “noise” and are thus less identifiable. These considerations are tangential to our arguments, as we use the standard settings for within-trial noise parameters, and we estimated a ratio of across- to within-trial noise that was representative of the broader literature.

*Application to experimental data.* Although the model recovery analyses provided clear results, it is still possible that the models are indistinguishable in real data. We provide a proof-of-concept against this possibility, by fitting the two models to data from two classic studies of decision-making with constant environmental input (Experiment 1, Ratcliff and McKoon 2008; Roitman and Shadlen 2002). Ratcliff and McKoon (2008) had 15 human participants, and Roitman and Shadlen (2002) had 2 Rhesus macaques, make decisions about random dot motion (RDM), a popular paradigm in the study of visual perceptual decision-making. An RDM decision is based on a cloud of dots, of which a certain percentage move coherently toward the left or right of the screen while the remaining dots move randomly. The task was to indicate the direction of coherent motion by making button responses (Ratcliff and McKoon 2008) or eye movements (Roitman and Shadlen 2002). The percentage of coherently-moving dots was varied from trial to trial across six levels that differed across the two studies: Ratcliff and McKoon (2008) 5%, 10%, 15%, 25%, 35%, and 50%; Roitman and Shadlen (2002) 0%, 3.2%, 6.4%, 12.8%, 25.6%, and 51.2%. Ratcliff

and McKoon (2008) had each participant complete approximately 960 decision trials. Roitman and Shadlen (2002) had one monkey complete 2,614 trials and the other complete 3,534 trials.

To demonstrate that the EAM and UGM accounts can be identified in data, we fit the models to individual participant data using identical methods as in our second model recovery analysis, which assumed a fixed time constant of 100 ms in the UGM and allowed for trial-to-trial variability in drift rate in both models ( $\eta$  parameter). Although we fit the models to individual participant data, we present the data and model predictions averaged over participants. The top panels of Fig. 5 show that the EAM provided a better fit to the human data of Ratcliff and McKoon (2008).<sup>4</sup> The bottom panels show the reverse result: the macaque data of Roitman and Shadlen (2002) were much better described by the UGM. Neither model provided a perfect account of the data—for example, the EAM predicted the 90<sup>th</sup> percentile was slower than observed in the data of Ratcliff and McKoon (2008), and the UGM tended to underestimate the observed accuracy rates and predicted the 10<sup>th</sup> percentile for error responses was slower than observed in the data of Roitman and Shadlen (2002). Nevertheless, the fits of the EAM and UGM clearly demonstrate that separation of the models is possible in real data. In particular, our model fits suggest that the principles of evidence accumulation may provide the best account of the human data of Ratcliff and McKoon (2008), and that the nonhuman primates of Roitman and Shadlen (2002) may have used urgency-related mechanisms in their decisions, which may or may not be implemented within an accumulation framework (e.g., see Ditterich 2006).

While this analysis demonstrates that the models can be discriminated in data, and that at least one data set exists whose data are consistent with each model, much more work is required to address the question of which model fits most data best. Extensive analysis of many experiments from multiple paradigms will be required to establish whether the UGM enjoys the same empirical support that the EAM has gathered in previous research.

## DISCUSSION

Evidence accumulation models have a long history in explaining decision-making, at neural and behavioral levels. Recent research has suggested that this success occurred despite the fundamental assumption of the models being wrong. In contrast, the urgency gating model (Cisek et al. 2009; Thura et al. 2012) supposes instead that decision-relevant information from the environment is not accumulated at all but is instead used in a moment-by-moment fashion in the form of low-pass filtering the sensory state. This interesting, and radical, proposal raises the question of why evidence accumulation models have provided a good fit to data from hundreds of experiments in dozens of paradigms, if their fundamental assumptions are erroneous.

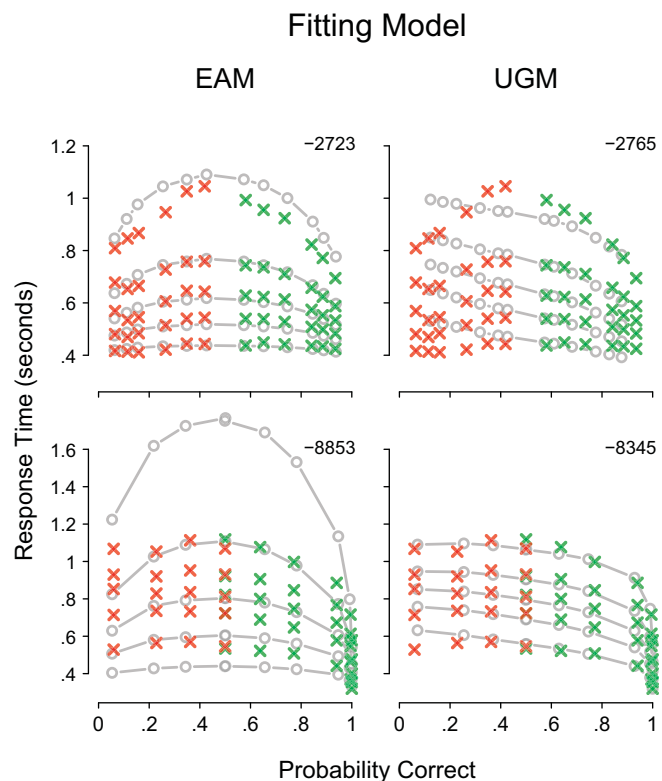
A proposed explanation for this discrepancy relies on a proof that the evidence accumulation model and the urgency gating

<sup>4</sup> The fit of the EAM to the data of Ratcliff and McKoon (2008) reported here is not as precise as the fit of the EAM shown in the original publication. This is because there were differences in the model assumptions and fitting methods. Our current set of assumptions and methods were chosen to best match the UGM for fair comparison.

Fig. 5. Quantile probability plots of the goodness of fit of the EAM (left panels) and UGM (right panels) to data from Ratcliff and McKoon (2008; top panels) and Roitman and Shadlen (2002; bottom panels). Green and red crosses, respectively, show mean correct and error responses in data. Model predictions averaged over individual participant fits are shown as gray circles connected with lines. The mean QMP goodness-of-fit statistic is shown in the top right of each panel, where larger (more positive) values indicate better fit to data. All other details are as described in Fig. 2A.

Ratcliff &  
McKoon (2008)

Roitman &  
Shadlen (2002)



model are mathematically identical in situations where the decision stimulus does not vary with time (Thura et al. 2012). This might explain the successes of the evidence accumulation models, as many experiments in which they are used involve time-constant stimuli. There are, however, notable exceptions where evidence accumulation models have provided a good account of decisions with time-varying information, such as dynamic stimuli that vary in strength (e.g., Brown and Heathcote 2005; Pietsch and Vickers, 1997; Usher and McClelland, 2001), and briefly presented stimuli (e.g., Huk and Shadlen 2005; Ratcliff 2002; Ratcliff and Rouder 2000; Ratcliff et al. 2003b; Smith and Ratcliff 2009; Smith et al. 2004; Thapar et al. 2003). In the latter paradigms, letters, patches of black and white pixels, or Gabor patches were presented briefly and then masked. If evidence used in the decision process tracks stimulus information, as the UGM suggests, then drift rate should rise at stimulus onset and then fall to zero. When stimulus durations are short enough that the decision process has not terminated before the mask, then this is equivalent to moving the starting point toward the correct decision boundary. Ratcliff and Rouder (2000) showed that this predicts errors that are much slower than correct responses (because the diffusion process must travel a long way to the incorrect boundary), but this did not occur in data. The data provided most support for a model that assumed sensory evidence was encoded in a short-term representation that provided a constant drift rate, which suggests that the EAM can provide a good account of data from tasks that use nonstationary decision stimuli. It is not immediately clear how the UGM may account for this pattern of results as its low-pass filtering of sensory evidence leads to a gradual decrease in information over time. This series of experiments might therefore provide a critical test of the UGM.

More generally, the EAM and UGM examined here can be considered special cases drawn from the broader class of sequential sampling models (e.g., Ditterich 2006). For example, both models assume arbitrary time constants (perfect in the case of the EAM, relatively short for the UGM) and time-variant gain (constant in the EAM, increasing over time in the UGM). Neither approach can provide a perfect account of data, and it might be most fruitful to consider the models examined here as particular model instantiations drawn from a larger family of possible models. Careful comparison of general model features, such as urgency signals, allows bounds to be placed on models of the cognitive processes that are more (or less) active under various task constraints (for an example of this approach, see Hawkins et al. 2015). We believe that consideration of data from a large range of experimental paradigms, including briefly presented stimuli and time-invariant and time-variant sensory information, will be instructive in this process.

Here, we have shown that the proof of Thura et al. (2012) is incorrect, and that evidence accumulation models are not identical to the urgency gating model, even when the decision stimulus is constant. We also showed that the two models could be perfectly distinguished in simulated data—in which the true, data-generating model was known—even when the simulated stimulus was constant, and when the sample sizes were realistic. Finally, we demonstrated that the two models can be discriminated in real data. Our results therefore show that the urgency gating model and evidence accumulation models can be discriminated in standard paradigms with time-constant stimuli. This work opens the possibility of fruitful investigations of the relative merits of the two opposing models considered here, and the family of sequential sampling models in general, without wholesale change to the standard experimental paradigms.



## GRANTS

This work was supported in part by Australian Research Council Future Fellowship FT120100244 to S. D. Brown and a European Research Council grant to E.-J. Wagenmakers.

## DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

## AUTHOR CONTRIBUTIONS

Author contributions: G.E.H., E.-J.W., R.R., and S.D.B. conception and design of research; G.E.H., E.-J.W., R.R., and S.D.B. analyzed data; G.E.H., E.-J.W., R.R., and S.D.B. interpreted results of experiments; G.E.H. and S.D.B. prepared figures; G.E.H. and S.D.B. drafted manuscript; G.E.H., E.-J.W., R.R., and S.D.B. edited and revised manuscript; G.E.H., E.-J.W., R.R., and S.D.B. approved final version of manuscript.

## REFERENCES

- Ardia D, Mullen KM, Peterson BG, Ulrich J. *DEoptim: Differential Evolution in R* [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=DEoptim> (R package version 2.2-2), 2013.
- Brown S, Heathcote A. Practice increases the efficiency of evidence accumulation in perceptual choice. *J Exp Psychol: Human Percept Perform* 31: 289–298, 2005.
- Cisek P, Puskas GA, El-Murr S. Decisions in changing conditions: the urgency-gating model. *J Neurosci* 29: 11560–11571, 2009.
- Ding L, Gold JI. Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cereb Cortex* 22: 1052–1067, 2012.
- Ditterich J. Stochastic models of decisions about motion direction: behavior and physiology. *Neural Netw* 19: 981–1012, 2006.
- Feller W. *An Introduction to Probability Theory and Its Applications: Vol. I*. New York: Wiley, 1968.
- Forstmann BU, Anwander A, Schäfer A, Neumann J, Brown S, Wagenmakers EJ, Turner R. Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proc Natl Acad Sci USA* 107: 15916–15920, 2010.
- Forstmann BU, Dutilh G, Brown S, Neumann J, von Cramon DY, Ridderinkhof KR, Wagenmakers EJ. Striatum and pre-SMA facilitate decision-making under time pressure. *Proc Natl Acad Sci USA* 105: 17538–17542, 2008.
- Gronlund SD, Ratcliff R. Analysis of the Hockley and Murdock decision model. *J Math Psychol* 35: 319–344, 1991.
- Hanes DP, Schall JD. Neural control of voluntary movement initiation. *Science* 274: 427–430, 1996.
- Hawkins GE, Forstmann BU, Wagenmakers EJ, Ratcliff R, Brown SD. Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *J Neurosci* 35: 2476–2484, 2015.
- Heathcote A, Brown SD. Reply to Speckman and Rouder: A theoretical basis for QML. *Psychon Bull Rev* 11: 577–578, 2004.
- Heathcote A, Brown SD, Mewhort DJK. Quantile maximum likelihood estimation of response time distributions. *Psychon Bull Rev* 9: 394–401, 2002.
- Hockley WE, Murdock BB. A decision model for accuracy and response latency in recognition memory. *Psychol Rev* 94: 341–358, 1987.
- Huk AC, Shadlen MN. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J Neurosci* 25: 10420–10436, 2005.
- Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324: 759–764, 2009.
- Krajchich I, Rangel A. A multi-alternative drift diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc Natl Acad Sci USA* 108: 13852–13857, 2011.
- Laming DRJ. *Information Theory of Choice-Reaction Times*. London: Academic, 1968.
- Link SW, Heath RA. A sequential theory of psychological discrimination. *Psychometrika* 40: 77–105, 1975.
- Luce RD. *Response Times*. New York: Oxford Univ. Press, 1986.
- Maimon G, Assad JA. A cognitive signal for the proactive timing of action in macaque LIP. *Nat Neurosci* 9: 948–955, 2006.
- Matzke D, Wagenmakers EJ. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon Bull Rev* 16: 798–817, 2009.
- Mullen K, Ardia D, Gil D, Windover D, Cline J. DEoptim: An R package for global optimization by differential evolution. *J Stat Software* 40: 1–26, 2011.
- Navarro DJ, Pitt MA, Myung IJ. Assessing the distinguishability of models and the informativeness of data. *Cogn Psychol* 49: 47–84, 2004.
- O'Connell RG, Dockree PM, Kelly SP. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci* 15: 1729–1735, 2012.
- Palmer J, Huk AC, Shadlen MN. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J Vision* 5: 376–404, 2005.
- Pietsch A, Vickers D. Memory capacity and intelligence: Novel techniques for evaluating rival models of a fundamental information-processing mechanism. *J Gen Psychol* 124: 231–339, 1997.
- Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ. Neurally constrained modeling of perceptual decision making. *Psychol Rev* 117: 1113–1143, 2010.
- Purcell BA, Schall JD, Logan GD, Palmeri TJ. From salience to saccades: Multiple-alternative gated stochastic accumulator model of visual search. *J Neurosci* 32: 3433–3446, 2012.
- Ratcliff R. A theory of memory retrieval. *Psychol Rev* 85: 59–108, 1978.
- Ratcliff R. A diffusion model account of reaction time and accuracy in a brightness discrimination task: fitting real data and failing to fit fake but plausible data. *Psychon Bull Rev* 9: 278–291, 2002.
- Ratcliff R, Cheria A, Segraves M. A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. *J Neurophysiol* 90: 1392–1407, 2003a.
- Ratcliff R, Hasegawa YT, Hasegawa YP, Smith PL, Segraves MA. Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J Neurophysiol* 97: 1756–1774, 2007.
- Ratcliff R, McKoon G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput* 20: 873–922, 2008.
- Ratcliff R, Philastides MG, Sajda P. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc Natl Acad Sci USA* 106: 6539–6544, 2009.
- Ratcliff R, Rouder JN. A diffusion model account of masking in two-choice letter identification. *J Exp Psychol Human Percept Perform* 26: 127–140, 2000.
- Ratcliff R, Thapar A, Gomez P, McKoon G. A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychol Aging* 19: 278–289, 2004.
- Ratcliff R, Thapar A, McKoon G. A diffusion model analysis of the effects of aging on brightness discrimination. *Percept Psychophys* 65: 523–535, 2003b.
- Ratcliff R, Van Dongen HP. Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proc Natl Acad Sci USA* 108: 11285–11290, 2011.
- Roitman JD, Shadlen MN. Responses of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22: 9475–9489, 2002.
- Schurger A, Sitt JD, Dehaene S. An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc Natl Acad Sci USA* 109: 2904–2913, 2012.
- Smith PL. Stochastic dynamic models of response time and accuracy: A foundational primer. *J Math Psychol* 44: 408–463, 2000.
- Smith PL, Ratcliff R. An integrated theory of attention and decision making in visual signal detection. *Psychol Rev* 116: 283–317, 2009.
- Smith PL, Ratcliff R, Wolfgang DJ. Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. *Vision Res* 44: 1297–1320, 2004.
- Stone M. Models for choice-reaction time. *Psychometrika* 25: 251–260, 1960.
- Thapar A, Ratcliff R, McKoon G. A diffusion model analysis of the effects of aging on letter discrimination. *Psychol Aging* 18: 415–429, 2003.
- Thura D, Beauregard-Racine J, Fradet CW, Cisek P. Decision making by urgency gating: theory and experimental support. *J Neurophysiol* 108: 2912–2930, 2012.
- Usher M, McClelland JL. On the time course of perceptual choice: the leaky competing accumulator model. *Psychol Rev* 108: 550–592, 2001.
- Van Ravenzwaaij D, Dutilh G, Wagenmakers EJ. A diffusion model decomposition of the effects of alcohol on perceptual decision making. *Psychopharmacology* 219: 1017–1025, 2012.
- Wagenmakers EJ, Ratcliff R, Gomez P, Iverson GJ. Assessing model mimicry using the parametric bootstrap. *J Math Psychol* 48: 28–50, 2004.