

Revisiting the Evidence for Collapsing Boundaries and Urgency Signals in Perceptual Decision-Making

Guy E. Hawkins,¹ Birte U. Forstmann,² Eric-Jan Wagenmakers,³ Roger Ratcliff,⁴ and Scott D. Brown¹

¹School of Psychology, University of Newcastle, Callaghan, NSW 2308, Australia, ²Amsterdam Brain and Cognition and ³Department of Psychology, University of Amsterdam, Amsterdam 1018WS, The Netherlands, and ⁴Department of Psychology, The Ohio State University, Columbus, Ohio 43210

For nearly 50 years, the dominant account of decision-making holds that noisy information is accumulated until a fixed threshold is crossed. This account has been tested extensively against behavioral and neurophysiological data for decisions about consumer goods, perceptual stimuli, eyewitness testimony, memories, and dozens of other paradigms, with no systematic misfit between model and data. Recently, the standard model has been challenged by alternative accounts that assume that less evidence is required to trigger a decision as time passes. Such “collapsing boundaries” or “urgency signals” have gained popularity in some theoretical accounts of neurophysiology. Nevertheless, evidence in favor of these models is mixed, with support coming from only a narrow range of decision paradigms compared with a long history of support from dozens of paradigms for the standard theory. We conducted the first large-scale analysis of data from humans and nonhuman primates across three distinct paradigms using powerful model-selection methods to compare evidence for fixed versus collapsing bounds. Overall, we identified evidence in favor of the standard model with fixed decision boundaries. We further found that evidence for static or dynamic response boundaries may depend on specific paradigms or procedures, such as the extent of task practice. We conclude that the difficulty of selecting between collapsing and fixed bounds models has received insufficient attention in previous research, calling into question some previous results.

Key words: decision-making; diffusion model; human; nonhuman primate; response time

Introduction

Over the past 50 years, psychology and neuroscience have contributed to a deeper understanding of decision-making based on “diffusion models” (Stone, 1960; Laming, 1968; Ratcliff, 1978). Diffusion models assume that noisy information is gradually sampled from the environment. The process continues until the balance of evidence reaches one of two decision boundaries, triggering a choice. In hundreds of human studies with thousands of participants, diffusion models have described data very accurately, providing insight into many theoretical and practical research areas, including decisions about consumer goods, memories, motion stimuli, clinical populations, aging, sleep deprivation, and psychopharmacology (Ratcliff, 1978; Krajbich et al., 2010; Krajbich and Rangel, 2011; Ratcliff and Van Dongen, 2011). More recently, the same models have been correlated with various components of neuroimaging and neurophysiological data in humans (Forstmann et al., 2008, 2010; O’Connell et al., 2012; Ratcliff et al., 2009; Schurger et al., 2012). In *Macaca mulatta*, the firing rates of some neurons seem to behave like processes from diffusion models (Bollimunta et al., 2012; Ding

and Gold, 2010, 2012a; Hanes and Schall, 1996; Heitz and Schall, 2012; Kiani and Shadlen, 2009; Pouget et al., 2011; Purcell et al., 2010; Purcell et al., 2012; Ratcliff et al., 2003; Roitman and Shadlen, 2002; Woodman et al., 2008): during decision-making, the firing rate of these neurons increases until it reaches a threshold value and a behavioral response is initiated. The notion that firing rate in some cells might represent a decision threshold has even received independent support from neural studies of saccade production (Brown et al., 2008; Hanes et al., 1998; Paré and Hanes, 2003).

Diffusion models typically assume fixed decision boundaries; the amount of evidence required to trigger a decision does not change with time (dashed lines, left panel, Fig. 1). Recently, however, a more complicated assumption has gained popularity: collapsing boundaries (solid curved lines, Fig. 1), sometimes interpreted as urgency signals, where decisions are triggered by less and less evidence as time passes (Bowman et al., 2012; Cisek et al., 2009; Ditterich, 2006a, 2006b; Drugowitsch et al., 2012; Milosavljevic et al., 2010; Thura et al., 2012). Figure 1 demonstrates how models with fixed and collapsing bounds make different predictions for response times. The collapsing bounds reduce the number of slow decisions, making response time distributions less skewed.

For decades, theories using fixed bounds have provided precise accounts of many aspects of the data. It is not clear whether collapsing bounds and urgency signal models fit data quite as well. Nevertheless, some researchers have asserted that the non-stationary assumption is true; for example, “It turns out that the

Received June 12, 2014; revised Oct. 3, 2014; accepted Nov. 18, 2014.

Author contributions: G.E.H., B.U.F., E.-J.W., R.R., and S.D.B. designed research; G.E.H. and S.D.B. performed research; R.R. contributed unpublished reagents/analytic tools; G.E.H. and S.D.B. analyzed data; G.E.H., B.U.F., E.-J.W., R.R., and S.D.B. wrote the paper.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Guy E. Hawkins, Amsterdam Brain and Cognition Center, University of Amsterdam, Nieuwe Achtergracht 129, 1018 W Amsterdam, The Netherlands. E-mail: guy.e.hawkins@gmail.com.
DOI:10.1523/JNEUROSCI.2410-14.2015

Copyright © 2015 the authors 0270-6474/15/352476-09\$15.00/0

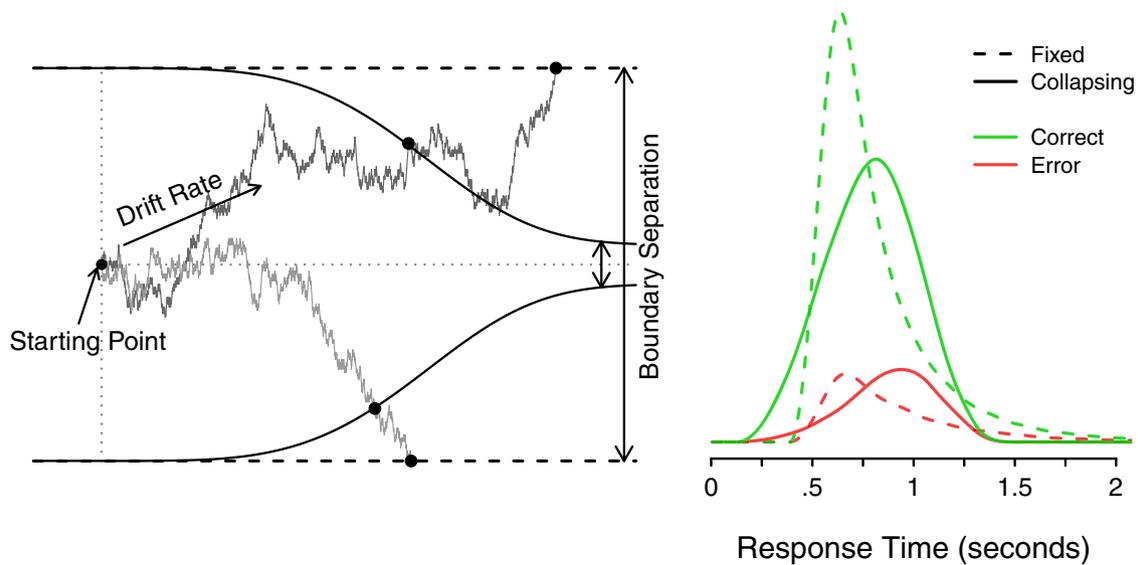


Figure 1. Left, Diffusion models with fixed (dashed) or collapsing (solid) decision boundaries. A model with collapsing boundaries can terminate the evidence accumulation process earlier than a model with fixed boundaries, resulting in faster decisions. Right, Graph showing how the models lead to different predictions for response time distributions.

prediction was misguided. There is no reason to assume the terminating bounds are flat (i.e., constant as a function of elapsed decision time)” (Shadlen and Kiani, 2013). We addressed these problems in a large-scale survey and found, overall, evidence in favor of the fixed bound approach.

Materials and Methods

Model details. We study diffusion processes through discrete-state approximations. The basic assumption of the diffusion process is that a decision-maker accumulates noisy evidence from the environment over time. The accumulated evidence evolves toward one of two decision criteria that correspond to the two response alternatives and when the process reaches one of these boundaries a response is triggered. The predicted response time is the sum of the time taken to reach the boundary and an offset time required for non decision-related components of choice, such as encoding the stimulus and executing a motor response. The predicted response corresponds to the boundary that was crossed.

The basic diffusion model (as just described) is governed by five parameters: the average rate at which the process drifts toward one boundary (drift rate, ν); noise in the diffusion process (s , a scaling parameter not typically estimated from data and fixed in all model fits herein); the separation of the boundaries (a); the starting position of the diffusion process (z); and the time taken for nondecision processes (t_{er}). This basic diffusion model for decision-making is over 50 years old Stone (1960). Modern accounts assume variability in three model parameters from decision to decision (reflecting, e.g., fluctuations in attention). Such variability addresses well known deficiencies in the basic diffusion model—most prominently, if the boundaries are equidistant from the starting point, then the predicted correct and error response times are identical Feller (1968). The three parameters assumed to vary from decision to decision are the start point (uniform on the interval $z - \frac{sz}{2}, z + \frac{sz}{2}$), the drift rate (normally distributed with mean ν and SD η), and the nondecision time (uniform on the interval $t_{er} - \frac{st}{2}, t_{er} + \frac{st}{2}$).

Collapsing bounds and urgency signal models. The collapsing bounds model we analyzed was identical to the fixed bounds model just described except that the upper boundary decreased from its initial value (a) to an asymptotic value that was not forced to be the halfway point ($\frac{1}{2}a$), but was allowed to be some amount (a') away from that (Fig. 1). The lower boundary increased similarly from its initial value of zero to an asym-

metotic the same distance below the halfway point ($\frac{1}{2}a - a'$). The rate and form of the boundary collapse can be instantiated in multiple ways, corresponding to different assumptions about the decision process. For example, boundaries might collapse at an early stage in processing (Ratcliff and Frank, 2012) or throughout processing to gradually impose a deadline on decision time (Bowman et al., 2012). We assumed that the boundary changes followed a function that accommodates a wide range of possible shapes, including early, late, and gradually collapsing bounds, and also nests the more constrained fixed bounds model: the Weibull cumulative distribution function (Fig. 2). The flexibility of the Weibull distribution function allows it to approximate a range of candidate functions that might be proposed for collapsing bounds. For example, Figure 3 shows four potential boundary functions—fixed, exponential, hyperbolic, and logistic—and the best-fitting Weibull function for each of these generating functions. That the Weibull can closely mimic various forms of time variant boundaries suggests that the results reported below are not dependent on our particular choice of function for the collapsing boundary.

In the collapsing bounds model, we assumed the upper threshold u at time t after onset of evidence accumulation would be as follows:

$$u(t) = a - \left(1 - \exp\left(-\left(\frac{t}{\lambda}\right)^k\right)\right)\left(\frac{1}{2}a - a'\right) \quad (1)$$

To all datasets, we fit four variants of the collapsing bounds model that corresponded to different psychological assumptions about the decision process. These assumptions informed choices about which parameters were freely estimated from data and which were fixed (not estimated from data). Our results primarily focus on the “late collapse” models in Figure 2, which freely estimated the stage of the decision process when boundaries began to collapse (λ , scale parameter of the Weibull distribution) and the extent to which the boundaries collapsed (a' , asymptotic boundary setting, Eq. 1). The shape parameter was fixed at $k = 3$ —the value used in the middle right panel of Figure 2—to impose a “late collapse” decision strategy, and was also a representative value from preliminary model fitting when no constraint was placed on the shape parameter. The second model variant permitted freedom in whether the decision boundaries collapsed early or late in processing and in the rate of collapse (shape and scale parameters freely estimated from data) and fixed the asymptotic boundary setting to instantiate complete collapse, with the upper and lower boundaries meeting ($a' = 0$). The third variant fixed the shape at $k = 3$ and imposed complete collapse on the asymptotic

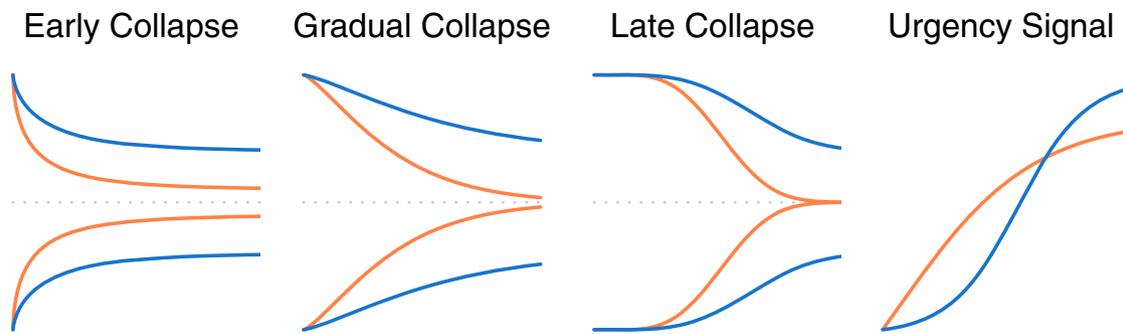


Figure 2. Various forms of dynamic diffusion models. The Weibull cumulative distribution function can generate decision boundaries that collapse at early (left) or late (middle right) stages of processing or gradually throughout processing (middle left). Orange lines indicate strong (left) and complete (middle left, middle right) collapse of the upper and lower boundaries, which imposes a hard deadline on processing. Blue lines represent a milder form of collapsing boundary that do not meet at the midpoint ($\frac{1}{2}a$). Right, Two illustrative urgency signal paths, which are applied as a gain (i.e., multiplicative) function on drift rates according to Equation 2. The orange and blue paths show rapid and delayed urgency signals, respectively, which are functionally similar to early and late collapsing boundaries.

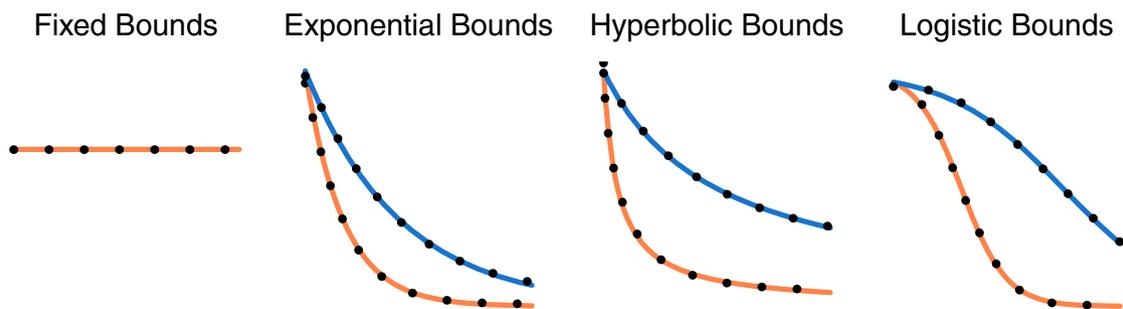


Figure 3. The Weibull cumulative distribution function mimics parametric forms that might be proposed for collapsing bounds. Orange and blue lines show examples of candidate stationary and dynamic boundaries of various parametric functions: fixed, exponential, hyperbolic, and logistic. Overlaid black dotted lines indicate the best fitting Weibull function, which in all cases closely mimics the generating function.

boundary ($a' = 0$) and freely estimated the time at which the boundaries commenced their trajectory toward complete collapse. The final model freely estimated all three parameters, which allowed freedom in the rate and form of the collapsing boundary and in the extent of the collapse. The results described below were from the first model and they generalized across the four variants of the collapsing bounds model used.

We also considered a different kind of collapsing bounds model, instantiated via an “urgency signal” (Churchland et al., 2008; Cisek et al., 2009; Ditterich, 2006a; Hanks et al., 2014; Thura et al., 2012). In this model, the boundaries were fixed throughout the decision, but the evidence accumulation process was subject to a gain parameter with a value that increased with the duration of the decision (Fig. 2). The gain parameter was implemented as a three-parameter logistic function following Ditterich (2006a). The urgency signal γ at time t after the onset of evidence accumulation is given by the following:

$$\gamma(t) = \frac{s_y \exp(s_x(t - d))}{1 + \exp(s_x(t - d))} + \frac{1 + (1 - s_y) \exp(-s_x d)}{1 + \exp(-s_x d)} \quad (2)$$

where d is a delay parameter and s_x and s_y are shape parameters (Fig. 2).

We also ran our analyses twice: once with the fixed and collapsing bounds models described above and once with the data modeled as a mixture of those models (at 98%) and a random uniform process across the observed range of response times (at 2%). This mixture represents the idea that a small fraction of responses might be contaminants unrelated to the stimulus or the regular decision process. The two analyses agreed closely and below we discuss results for the contaminant mixture model, which can provide more stable parameter estimates in the presence of outlying data (Ratcliff and Tuerlinckx, 2002).

Datasets for modeling. In a classic experiment, previously interpreted as supporting a fixed-bound diffusion model because the neural firing rates

for fast and slow responses appeared to hit the same maximum level to trigger a decision, Roitman and Shadlen (2002) had two monkeys make decisions about random dot motion (RDM; Fig. 4). A 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 monkeys’ task was to indicate the direction of coherent motion by making eye movements. The percentage of coherently moving dots was varied from trial to trial across six levels (0%, 3.2%, 6.4%, 12.8%, 25.6% and 51.2%). One rhesus monkey completed 2614 trials (Monkey B) and another completed 3534 trials (Monkey N).

Palmer et al. (2005), Experiment 1, replicated the experiment of Roitman and Shadlen (2002), but used six human participants. Methodological details were almost identical except for the duration of data collection sessions, the precise timing of feedback information, and the nature of rewards, all of which were tailored to human needs. The human participants each completed ~560 trials.

Ratcliff and McKoon (2008), Experiment 1, replicated the experiment of Palmer et al. (2005) with 15 human participants. Methodological details were similar, except that motion coherence was varied from trial to trial across six different levels (5%, 10%, 15%, 25%, 35%, and 50%), responses were indicated by button presses rather than saccades, and feedback on correct versus incorrect decisions was provided immediately rather than after a short delay. The human participants each completed ~960 trials.

Ratcliff et al. (2007) had two monkeys make decisions about patches of pixels that varied in brightness (Fig. 4). Monkeys were shown a square of black and white pixels that ranged across six levels in the proportion of black pixels (“bright”: 2%, 35%, 45%; “dark”: 55%, 65%, 98%). Responses were indicated by saccades. One rhesus monkey completed 12,021 trials and another completed 7632 trials.

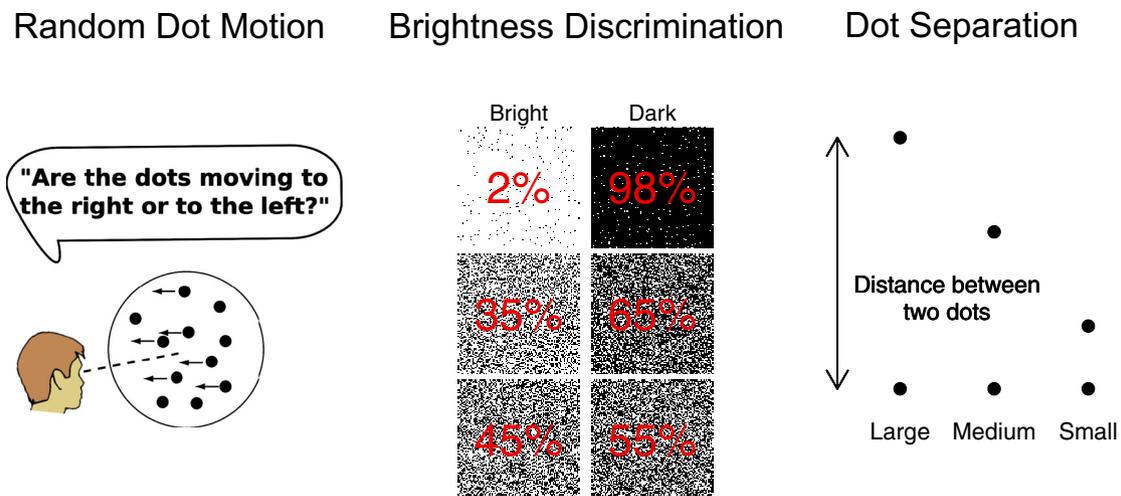


Figure 4. The three decision paradigms from the human and nonhuman primate studies: random dot motion, brightness discrimination, and dot separation.

Middlebrooks and Schall (2014) had two monkeys and eight humans make decisions about square 10×10 checkerboards that contained more cyan or magenta checkers, similar to the Ratcliff et al. (2007) brightness discrimination task. The percentage of cyan to magenta checkers was varied randomly from trial to trial and ranged across seven levels, determined separately for each monkey (Monkey B: 41%, 45%, 48%, 50%, 52%, 55%, 59%; Monkey X: 35%, 42%, 47%, 50%, 53%, 58%, 65%) and all humans completed the same set (35%, 42%, 46%, 50%, 54%, 58%, 65%). Responses were indicated by saccades. Perceptual categorization trials were randomly interleaved with stop-signal trials in which a signal would appear at a variable time after stimulus onset that indicated the participant must inhibit a prepared response. We only analyzed the no-stop trials for consistency with the experimental designs of the remaining seven datasets in our analysis. This gave 10,212 trials for Monkey B, 10,762 for Monkey X, and an average of 483 trials per human participant.

Ratcliff et al. (2001), Experiment 2 (young subjects), had humans decide whether the distance between two dots was small or large (Fig. 4). The separation of the dots ranged across 32 equal-sized steps from 11/16 inches (1.7 cm, “small”) to 1–5/16 inches (2.4 cm, “large”). Following Ratcliff et al. (2001), we collapsed data from the 32 levels with similar response time and accuracy performance into four conditions for model fitting.

In separate blocks, participants were instructed to respond as quickly as possible (speed-emphasis) or as accurately as possible (accuracy-emphasis). The 17 human participants each completed ~ 1000 accuracy-emphasis trials, which contributed to the primary analyses, and ~ 1000 speed-emphasis trials, which were used for a replication analysis. Responses were indicated by button presses.

Ratcliff et al. (2003) replicated Ratcliff et al.’s (2001) experiment but used two monkey participants. Methodological details were similar, except that, the duration of data collection was longer, dot separation ranged from 2 to 10° of visual angle in increments of 1° , there were no speed- or accuracy-emphasis instructions, and responses were indicated by saccades. One rhesus monkey completed 11,495 trials and another completed 5037 trials.

Experiment 1. Thirty-nine undergraduates from the University of Newcastle, Australia, completed speeded decisions in an RDM task. Methodological details were similar to Palmer et al. (2005) except that motion coherence was manipulated across six different levels (0%, 2.5%, 5%, 10%, 20%, 40%) and responses were indicated by button presses. The human participants each completed 432 trials.

An additional 74 undergraduates from the University of Newcastle completed RDM decisions for the replication analyses reported. All methodological details were the same as the first experiment except for a delayed feedback procedure similar to that of Palmer et al. (2005) and Roitman and Shadlen (2002). In particular, after a response, participants

did not receive feedback until at least 1 s (44 participants) or 2 s (30 participants) had elapsed since stimulus presentation.

Fitting diffusion models to data: parameter estimation and model selection. We estimated parameters for each model separately for each individual participant using quantile maximum products estimation (QMPE) (Heathcote et al., 2002; Heathcote and Brown, 2004; similar to χ^2 and multinomial maximum-likelihood estimation). The observed data were characterized by nine deciles, calculated separately for correct and incorrect responses. The QMP statistic is used to quantify agreement between the model and data by comparing the observed and predicted proportions of data falling into each interdecile bin. Because the urgency signal and collapsing bounds models do not have closed-form analytic solutions for their predicted distributions, we evaluated the predictions by Monte Carlo simulation using 10,000 replicates per experimental condition during parameter estimation and 50,000 replicates per condition for precisely evaluating predictions at the search termination point. To keep the model comparison fair, we used the same method for the fixed-bound diffusion (even though closed-form solutions are readily available). To simulate the models, we used Euler’s method to approximate their representation as stochastic differential equations, with a step size of 0.01 s fixed everywhere. In initial tests, we confirmed that our choice of step size made no difference other than slight linear inflation of the predicted response time distribution and that making the step size very small led to perfect agreement with the closed-form analytic solutions for the fixed-bound model (Brown et al., 2006).

We adjusted the model parameters to optimize goodness-of-fit using differential evolution methods (Ardia et al., 2013; Mullen et al., 2011). We also tested parameter optimization via particle swarm and simplex algorithms, but found that both approaches gave poorer model recovery performance. The model parameters for each participant and model were estimated independently. The mean drift rate parameter (ν) was estimated freely for each difficulty level (coherence level in random dot motion tasks, or brightness level in brightness discrimination). For example, the Ratcliff and McKoon (2008) experiment had six coherence levels, so we estimated six drift rate parameters. Roitman and Shadlen (2002), Palmer et al. (2005), and Experiment 1 had RDM stimuli with a 0% coherence condition, for which we fixed the drift rate to zero. We similarly fixed the drift rate to zero for the 50% cyan/magenta stimuli in Middlebrooks and Schall’s (2014) data. All other model parameters were estimated once for each participant and model combination (i.e., constant across difficulty levels): boundary height (a), start point of evidence accumulation (z), start-point variability (sz), nondecision time (t_{er}), variability in nondecision time (st), and drift rate variability (η). For the urgency signal and collapsing bounds models, we estimated the three additional parameters described in the previous section. We set wide bounds on all parameters and ran 120 particles for 500 search iterations.

We repeated this parameter estimation exercise three times independently for each model fit to each participant's data and chose the best set of parameters overall.

We used three different approaches to model selection to compare the goodness-of-fit for the fixed bounds, urgency signal, and collapsing bounds models. The fixed bound model is nested within both the other models, so it must always fit more poorly than either. The three model selection methods evaluated whether the improvement in fit observed for the urgency signal or collapsing bounds models was sufficient to justify their extra complexity. We used the Akaike Information Criterion (AIC) (Akaike, 1974), the Bayesian Information Criterion (BIC) (Schwarz, 1978), and nested model likelihood ratio tests (Edwards, 1992). The results of the three model selection methods were very similar, so we report only the BIC results below.

We used the estimated BIC values to approximate posterior model probabilities, which account for uncertainty in the model selection procedure. Assuming a uniform prior across the $m = 3$ models under consideration and that the data-generating model was one of those under consideration, the BIC-based approximation is given by Wasserman (2000) as follows:

$$P_{\text{BIC}}(M_i | \text{Data}) = \frac{\exp\left(-\frac{1}{2}\text{BIC}(M_i)\right)}{\sum_{j=1}^m \exp\left(-\frac{1}{2}\text{BIC}(M_j)\right)} \quad (3)$$

We also repeated the estimation of approximate posterior model probabilities comparing all four versions of the collapsing bounds model with the fixed bound and urgency signal models (i.e., comparison of $m = 6$ models). This did not change the conclusions regarding which model structure provided the best account of each dataset (i.e., fixed bounds, urgency signal, or collapsing bounds), so we do not report it further.

Model recovery. We observed that reliably discriminating between synthetic data generated by fixed bound versus collapsing bound models was not always easy. Prompted by this, we performed extensive model recovery analyses to ensure that our results were robust, sensitive, and unbiased and not due to artifacts of the parameter estimation routines. We used a model recovery procedure (Navarro et al., 2004; Wagenmakers et al., 2004) that involves simulating many datasets from one model (the fixed bounds model) and fitting those simulated datasets with both the fixed and collapsing bounds models. We then assessed the difference in BIC between the two fitted models for each simulated dataset. If model recovery is easy, then the fit of the data-generating model should yield a better BIC than its competitor model. This process was then repeated using simulated data from the collapsing bounds model. We did not repeat this procedure using the urgency signal model because this model was rarely preferred over the fixed or collapsing bounds models in data (3 of 93 subjects; Fig. 5) and the heavy computational burden.

For the simulation study, we used the first variant of our collapsing bounds models, which form the primary focus of our results where the shape parameter was fixed ($k = 3$). For each experiment under examination and separately for the fixed and collapsing bounds models, we averaged across the best fitting parameter estimates for each subject and used these to simulate 100 synthetic datasets from each model, with sample size equal to the sample size in the real datasets (with very large simulated sample sizes, model recovery was close to perfect). We ran the parameter estimation routine three times independently for each simulated dataset and chose the best set of parameters for each simulated dataset (i.e., the parameter set with the best value of the objective function). In total, the model recovery procedure required $\sim 10,000$ parameter estimation exercises.

Results

Evidence for stationary and dynamic decision-making models

We conducted the largest survey of data and performed the most intensive analyses of fixed versus collapsing decision boundaries to date. We gathered nine datasets from eight studies of the sort

that have been used to support dynamic decision-making models. We examined three distinct perceptual decision-making paradigms—random dot motion, brightness discrimination, and dot separation (Fig. 4)—with experiments conducted in four independent laboratories. In total, we used data from eight nonhuman primates (*M. mulatta*) and 85 humans, with $>116,000$ decisions in all.

For an overwhelming majority of human participants (72 of 85), the data supported the fixed bounds model over both the collapsing bounds and urgency signal models. Of the remaining 13 participants, nine were best described by the collapsing bounds model and results for the other four were inconclusive. Data from the eight nonhuman primates painted a different picture, with four clearly supporting the collapsing bounds model and two monkeys each supporting the fixed bounds and urgency signal models.

The results can be further understood by examination of the differences between experiments, shown as stacked bar plots of cumulative posterior model probabilities separately for each subject in the top row of Figure 5. Four of the five experiments with human participants provided evidence in favor of the fixed bounds model, with just one experiment providing evidence in favor of the collapsing bounds model. That single experiment Palmer et al. (2005), with just six participants, resulted in the same number of participants classified as having collapsing bounds than all the other human experiments combined, despite those other experiments having 79 participants in total. It is more difficult to draw such conclusions from the monkey data, because only two monkeys participated in each experiment.

The second row of panels in Figure 5 shows the average boundaries estimated for the fixed and collapsing bound models in each experiment using parameters averaged across subjects. In the datasets that supported the dynamic models, the estimated boundaries for the fixed and collapsing bounds models were very different; the collapsing bounds partially truncate the slow right tails of the predicted response time distributions, reducing their skew. This is shown in the quantile-probability plots in Figure 5: those datasets that supported the use of urgency signals or collapsing bounds had response time distributions that were less skewed than those experiments that did not support the dynamic models.

In addition to the data reported above, we replicated Experiment 1 by collecting data from an additional 74 participants under different experimental conditions (withholding feedback for a short delay) and by reanalyzing data from 17 participants in Ratcliff et al. (2001) with different instructions (emphasizing decision speed, rather than accuracy). Of these 91 additional human datasets, 64 showed very strong evidence for the fixed bounds model (i.e., $P_{\text{BIC}}(M_{\text{fixed}} | D) \geq .99$), and an additional 12 showed strong support (i.e., $P_{\text{BIC}}(M_{\text{fixed}} | D) \approx .95 - .99$). Of the remaining 15 subjects, one showed strong preference for the collapsing bounds model and the other 14 showed no strong preference for any model.

Collapsing boundaries and urgency signals do not replace between-trial variability in parameters

The modern fixed bounds model assumes that the drift rate, start point of evidence accumulation, and nondecision time vary randomly across trials (Ratcliff and Tuerlinckx, 2002; Ratcliff et al., 1999). The inclusion of these variability parameters has a demonstrated history of improving the fit of the fixed bounds model to data. For example, Ratcliff (1978) demonstrated that between-trial variability in drift rate allows the fixed bound model to

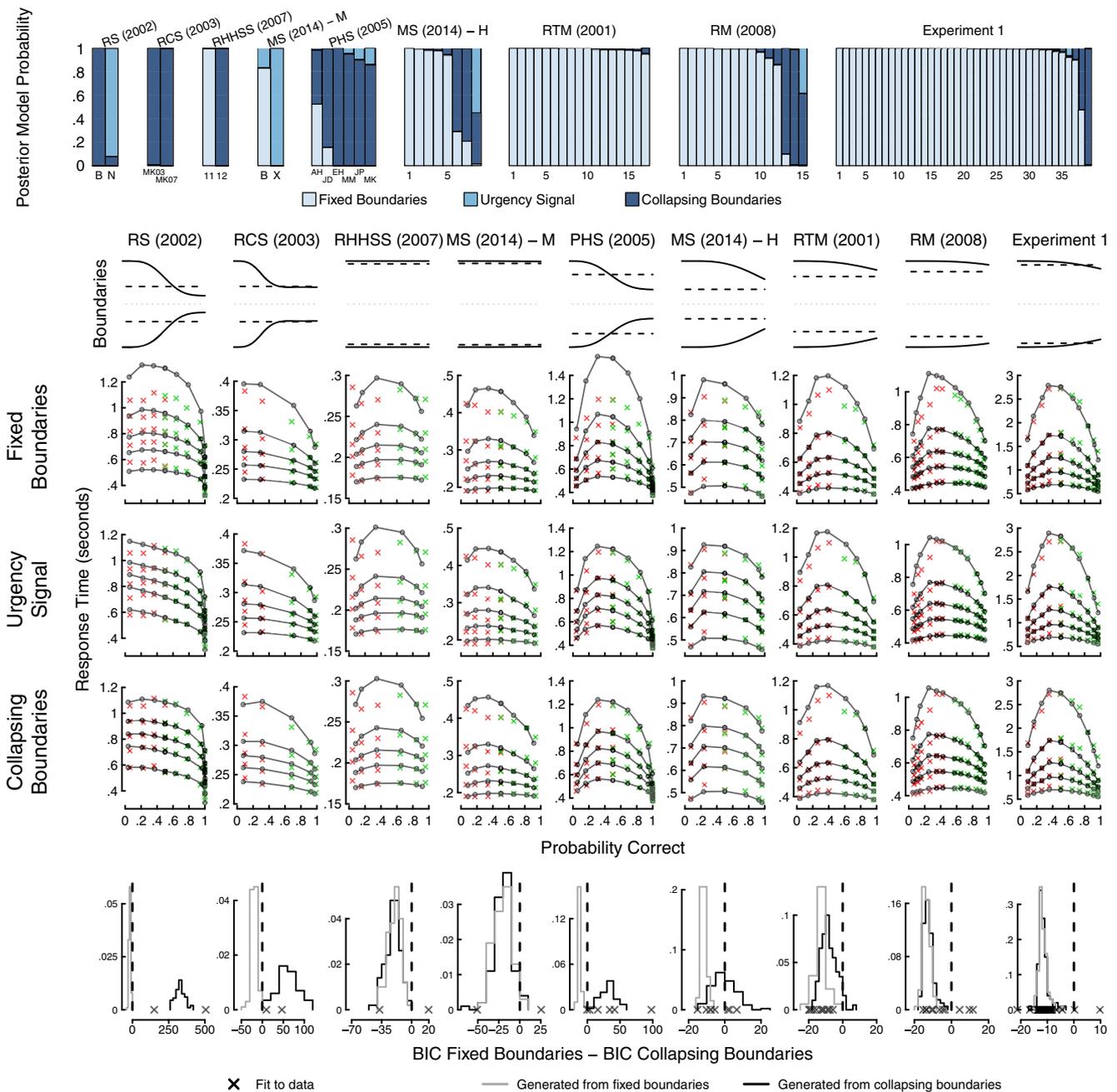


Figure 5. Top row, BIC-based approximations to posterior model probabilities in favor of the fixed bounds, urgency signal, and collapsing bounds models separately for the nine datasets. Shade represents the three models and columns represent individual subjects, with subject labels from the original reports below. Second row, Average decision boundaries for fixed and collapsing models for each experiment using parameters averaged over subjects. Third, fourth, and fifth rows, Choice probabilities and response times for both data and model fits, as quantile-probability plots. Panels show the probability of a correct response on the x-axes and response time (in seconds) on the y-axes. Green and red crosses represent correct and error responses, respectively, across experimental conditions. Vertical placement of the crosses show, for each condition, the 10th, 30th, 50th (i.e., median), 70th, and 90th percentiles of the response time distribution, aggregated across subjects. Predictions of the fixed bounds, urgency signal, and collapsing bounds models are overlaid on data as black lines. Bottom row, Results of model recovery simulation study. Histograms represent distributions of the difference in BIC values from data simulated from the fixed (gray histograms) and collapsing bounds (black histograms) models. Distributions that fall to the left of zero support the fixed bounds model and those to the right of zero support the collapsing bounds model. Crosses represent the corresponding BIC difference values from the model fits to data. Heading acronyms refer to the nine datasets: RS (2002), Roitman and Shadlen (2002); RCS (2003), Ratcliff et al. (2003); RHHSS (2007), Ratcliff et al. (2007); MS (2014)–M, macaques from Middlebrooks and Schall (2014); PHS (2005), Palmer et al. (2005); MS (2014)–H, humans from Middlebrooks and Schall (2014); RTM (2001), Ratcliff et al. (2001); RM (2008), Ratcliff and McKoon (2008); Experiment 1, Reported in this manuscript.

account for erroneous responses that are slower than correct responses. In contrast, slow errors naturally emerge from time-variant models with collapsing boundaries or urgency signals (Ditterich, 2006b) and therefore might provide an alternative conceptual grounding to the between-trial variability parameters of the conventional fixed bounds model.

We tested this assertion by comparing the fixed bounds model with between-trial variability in drift rate (η), start-point (sz), and nondecision time (st) to two models with no between-trial variability parameters (i.e., $\eta = 0, sz = 0, st = 0$): a collapsing bounds model with free parameters for the shape (k), scale (λ), and asymptotic boundary setting (a') and an urgency signal

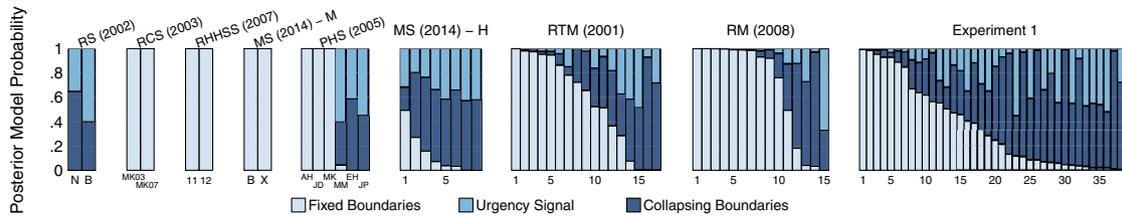


Figure 6. Approximations to posterior model probabilities in favor of the fixed bounds model with between trial variability parameters and the urgency signal and collapsing bounds models without between trial variability parameters. All details are as described for the top row of Figure 5.

model with delay (d) and two shape parameters (s_x , s_y). In this comparison, the three models have the same number of free parameters, so both AIC and BIC are equivalent to comparison directly on log-likelihood.

The stacked bar plots in Figure 6 show cumulative posterior model probabilities separately for each subject. Although the results are less clear than in the comparison above, the overall conclusion is unchanged by omission of between-trial variability parameters from the collapsing bounds models. As before, more human participants were best described by the fixed bounds model than by either the collapsing bounds or urgency signal models. Breaking the results for human participants down by experiment reveals that one experiment provides clear evidence in favor of the collapsing bounds or urgency signal model (Middlebrooks and Schall, 2014), two experiments provide mixed evidence (Experiment 1; Palmer et al., 2005), and the other two provide majority evidence in favor of the fixed bounds model (Ratcliff et al., 2001; Ratcliff and McKoon, 2008).

For data collected from nonhuman primates, the analysis without between-trial variability for the collapsing bound and urgency signal models provided strong support for the fixed bound model. For six of the eight monkeys, the fixed bound model was preferred in this analysis. As for the humans, this reinforces the conclusion that our decision to include the between-trial variability parameters in the primary analysis did not unfairly penalize the collapsing bounds and urgency signal models.

Model recovery

We took care to ensure that our methods for evaluating evidence in favor of collapsing versus fixed bounds models were robust, sensitive, and unbiased. We independently conducted the parameter optimization routines for the fit of all three models to each subject three times. We used an identical model fitting procedure to examine model recovery for each experiment using 100 datasets generated from each of the fixed and collapsing bounds models with the averaged parameter values across subjects within each experiment. We did not run model recovery simulations for the urgency signal model because it received so little support from the data.

The bottom row of Figure 5 shows that, for many parameter settings, model recovery was excellent even for the sample sizes matching the real data. Gray and black histograms represent data simulated from the fixed and collapsing bounds models, respectively, using the mean parameter estimates from the fits of the models to data. Distributions that fall below zero (dashed vertical line) support the fixed bounds model and vice versa for values above zero.

Model recovery was perfect when the data-generating parameters—taken from the fits to data—differed substantially between the fixed and collapsing models. For example, with simulated

datasets modeled after the experiments that supported the use of the collapsing bounds model (second row, Fig. 5; Palmer et al., 2005; Ratcliff et al., 2003; Roitman and Shadlen, 2002), the data-generating model was correctly identified 100% of the time (i.e., the gray distributions were below zero and black distributions above zero).

As expected, for those experiments in which the estimated parameters for the collapsing bounds model were similar to the fixed bounds model, datasets simulated from the collapsing bounds model are frequently classified as fixed bounds because the estimated boundaries for the collapsing bounds (second row, Fig. 5) were close to constant. The result is that the distributions of differences in BIC should all favor the fixed bounds model (i.e., fall below zero), because the collapsing bounds model is penalized due to its additional parameters, complexity that is not justified by the data generated from models with close-to-constant boundaries.

Discussion

Diffusion models using decision boundaries that do not change during a decision (fixed bounds) have provided detailed accounts of many aspects of decision-making data (Albantakis and Deco, 2009, 2011; Basten et al., 2010; Boucher et al., 2007; Fetsch et al., 2014; Hanes and Schall, 1996; Ho et al., 2009; Kiani and Shadlen, 2009; Lo and Wang, 2006; Mulder et al., 2012; Purcell et al., 2010; Purcell et al., 2012; Ramakrishnan et al., 2012; Ratcliff et al., 2003; Ratcliff et al., 2007; Ratcliff and Starns, 2013; Resulaj et al., 2009; Shankar et al., 2011). More complex nonstationary models with collapsing boundaries or increasing urgency signals have recently become popular, especially in some (Churchland et al., 2008; Ditterich, 2006a; Drugowitsch et al., 2012), but not all (Purcell et al., 2010; Purcell et al., 2012), neurophysiological studies of primates. The dynamic models implement a constantly changing decision strategy in which the quantity of evidence required to trigger a decision decreases with time. We conducted the first extensive investigation of the evidence for models of speeded decision-making with static versus dynamic response boundaries. Overall, data from nine experiments provided most support for the conventional, fixed bound model. We found evidence for collapsing boundaries or urgency signals for a small proportion of human subjects, but for most of the nonhuman primate participants (six of eight). A follow-up analysis using an alternative, simpler, specification for the collapsing boundary and urgency signal models did not reverse the conclusions for human participants, but did provide much stronger support for the fixed bounds model from the monkeys' data (Fig. 6).

Our results highlight the dangers of generalizing widely based on data from just one decision-making paradigm, species, or procedure, as has occurred from some proponents of dynamic diffusion models. Of the eight macaques that received very extensive practice, six exhibited strong support for dynamic bounds

models in our primary analysis. Conversely, the fixed bounds model was strongly supported in three experiments in which participants completed only a single session (Experiment 1; Ratcliff and McKoon, 2008) or two sessions (Ratcliff et al., 2001) of participation. Practice is known to influence boundary settings in decision-making models (Balci et al., 2011; Dutilh et al., 2009; Starns and Ratcliff, 2010), so it is plausible that collapsing bounds models provide good descriptions of data from extremely highly practiced participants; however, this remains an open question. Indeed, there are a number of datasets in the literature that we did not explore that demonstrate a good fit of fixed bounds models to data from highly practiced participants (Ding and Gold, 2012b; Purcell et al., 2010; Purcell et al., 2012; Ramakrishnan et al., 2012).

Task practice may also influence decision strategy through an interaction with the procedure used to administer rewards. Roitman and Shadlen (2002) and Palmer et al. (2005) withheld feedback information (and its associated reward for correct responses for nonhuman primates) until a response was registered or a minimum time poststimulus onset had elapsed (1 s), whichever came later. This procedure rewards the decision-maker for withholding their responses on some trials to avoid premature errors. Collapsing response boundaries might implement an appropriate withholding strategy: evidence accumulation can proceed as in a fixed bound model until the minimum delay time has passed, after which the boundary rapidly declines. An urgency signal model can similarly explain these data: the urgency signal is weak at trial onset, but as the delay period narrows, the signal increases to quickly elicit a response once the delay window has passed. These hypotheses are supported by the evidence in favor of dynamic boundary models from the two experiments using delayed reward timing (Palmer et al., 2005; Roitman and Shadlen, 2002).

It appears that subjects might learn the delayed response strategy after considerable task practice. In Roitman and Shadlen (2002), nonhuman primates completed thousands of decision trials and all of the participants in Palmer et al. (2005) had previous experience with similar psychophysical experiments before participation in the target experiment. This strategy was not observed in our replication experiments, in which 74 humans completed a single session under conditions similar to Palmer et al. (2005) with a minimum delayed feedback time of 1 or 2 s. Those subjects provided strong support for the fixed bounds model.

Overall, our survey suggests that humans and nonhuman primates use similar approaches in speeded decision-making tasks, but that both species can adopt different strategies as a function of the task constraints imposed by the experimenter. Many factors might differentially influence the use of decision strategies, including training, time-on-task, instructions, and the procedures used for reward timing; future research is required to understand the relative contributions of these factors. By applying computational cognitive models, we can disentangle the strategic effects of speeded decision-making from those that reflect the accumulation of information. Such a model-based decomposition of performance in humans and nonhuman primates maximizes the opportunity to learn about similarities and differences in cognitive processes across species.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723. [CrossRef](#)
- Albantakis L, Deco G (2009) The encoding of alternatives in multiple-choice decision making. *Proc Natl Acad Sci U S A* 106:10308–10313. [CrossRef Medline](#)
- Albantakis L, Deco G (2011) Changes of mind in an attractor network of decision-making. *PLoS Comput Biol* 7:e1002086. [CrossRef Medline](#)
- Ardia D, Mullen KM, Peterson BG, Ulrich J (2013) DEoptim: Differential evolution in R (R package version 2.2–2) [computer software manual]. Available from: <http://CRAN.R-project.org/package=DEoptim>.
- Balci F, Simen P, Niyogi R, Saxe A, Hughes JA, Holmes P, Cohen JD (2011) Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Atten Percept Psychophys* 73:640–657. [CrossRef Medline](#)
- Basten U, Biele G, Heekeren HR, Fiebach CJ (2010) How the brain integrates costs and benefits during decision making. *Proc Natl Acad Sci U S A* 107:21767–21772. [CrossRef Medline](#)
- Bollimunta A, Totten D, Ditterich J (2012) Neural dynamics of choice: Single-trial analysis of decision-related activity in parietal cortex. *J Neurosci* 32:12684–12701. [CrossRef Medline](#)
- Boucher L, Palmeri TJ, Logan GD, Schall JD (2007) Inhibitory control in mind and brain: an interactive race model of countermanding saccades. *Psychol Rev* 114:376–397. [CrossRef Medline](#)
- Bowman NE, Kording KP, Gottfried JA (2012) Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. *Neuron* 75:916–927. [CrossRef Medline](#)
- Brown JW, Hanes DP, Schall JD, Stuphorn V (2008) Relation of frontal eye field activity to saccade initiation during a countermanding task. *Exp Brain Res* 190:135–151. [CrossRef Medline](#)
- Brown S, Ratcliff R, Smith P (2006) Evaluating methods for approximating stochastic differential equations. *J Math Psychol* 50:402–410. [CrossRef](#)
- Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. *Nat Neurosci* 11:693–702. [CrossRef Medline](#)
- Cisek P, Puskas GA, El-Murr S (2009) Decisions in changing conditions: The urgency-gating model. *J Neurosci* 29:11560–11571. [CrossRef Medline](#)
- Ding L, Gold JI (2010) Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30:15747–15759. [CrossRef Medline](#)
- Ding L, Gold JI (2012a) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cereb Cortex* 22:1052–1067. [CrossRef Medline](#)
- Ding L, Gold JI (2012b) Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron* 75:865–874. [CrossRef Medline](#)
- Ditterich J (2006a) Evidence for time-variant decision making. *Eur J Neurosci* 24:3628–3641. [CrossRef Medline](#)
- Ditterich J (2006b) Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Netw* 19:981–1012. [CrossRef Medline](#)
- Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. *J Neurosci* 32:3612–3628. [CrossRef Medline](#)
- Dutilh G, Vandekerckhove J, Tuerlinckx F, Wagenmakers EJ (2009) A diffusion model decomposition of the practice effect. *Psychon Bull Rev* 16:1026–1036. [CrossRef Medline](#)
- Edwards AWF (1992) *Likelihood*. Baltimore: Johns Hopkins UP.
- Feller W (1968) *An introduction to probability theory and its applications, Vol 1*. New York: Wiley.
- Fetsch CR, Kiani R, Newsome WT, Shadlen MN (2014) Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83:797–804. [CrossRef Medline](#)
- Forstmann BU, Dutilh G, Brown S, Neumann J, von Cramon DY, Ridderinkhof KR, Wagenmakers EJ (2008) Striatum and pre-SMA facilitate decision-making under time pressure. *Proc Natl Acad Sci U S A* 105:17538–17542. [CrossRef Medline](#)
- Forstmann BU, Anwander A, Schäfer A, Neumann J, Brown S, Wagenmakers EJ, Bogacz R, Turner R (2010) Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proc Natl Acad Sci U S A* 107:15916–15920. [CrossRef Medline](#)
- Hanes DP, Schall JD (1996) Neural control of voluntary movement initiation. *Science* 274:427–430. [CrossRef Medline](#)
- Hanes DP, Patterson WF 2nd, Schall JD (1998) Role of frontal eye fields in countermanding saccades: visual, movement, and fixation activity. *J Neurophysiol* 79:817–834. [Medline](#)
- Hanks T, Kiani R, Shadlen MN (2014) A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *Elife* 3:e02260. [CrossRef Medline](#)
- Heathcote A, Brown S, Mewhort DJ (2002) Quantile maximum likelihood estimation of response time distributions. *Psychon Bull Rev* 9:394–401. [CrossRef Medline](#)

- Heathcote A, Brown SD (2004) Reply to Speckman and Rouder: a theoretical basis for QML. *Psychon Bull Rev* 11:577–578. [CrossRef](#)
- Heitz RP, Schall JD (2012) Neural mechanisms of speed-accuracy tradeoff. *Neuron* 76:616–628. [CrossRef](#) [Medline](#)
- Ho TC, Brown S, Serences JT (2009) Domain general mechanisms of perceptual decision making in human cortex. *J Neurosci* 29:8675–8687. [CrossRef](#) [Medline](#)
- Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324:759–764. [CrossRef](#) [Medline](#)
- Krajbich I, Rangel A (2011) A multi-alternative drift diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc Natl Acad Sci U S A* 108:13852–13857. [CrossRef](#) [Medline](#)
- Krajbich I, Armel C, Rangel A (2010) Visual fixations and comparison of value in simple choice. *Nat Neurosci* 13:1292–1298. [CrossRef](#) [Medline](#)
- Laming DRJ (1968) Information theory of choice-reaction times. London: Academic.
- Lo CC, Wang XJ (2006) Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci* 9:956–963. [CrossRef](#) [Medline](#)
- Middlebrooks PG, Schall JD (2014) Response inhibition during perceptual decision making in humans and macaques. *Atten Percept Psychophys* 76:353–366. [CrossRef](#) [Medline](#)
- Milosavljevic M, Malmaud J, Huth A, Koch C, Rangel A (2010) The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making* 5:437–449.
- Mulder MJ, Wagenmakers EJ, Ratcliff R, Boekel W, Forstmann BU (2012) Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *J Neurosci* 32:2335–2343. [CrossRef](#) [Medline](#)
- Mullen K, Ardia D, Gil D, Windover D, Cline J (2011) DEoptim: An R package for global optimization by differential evolution. *J Stat Softw* 40:1–26.
- Navarro DJ, Pitt MA, Myung IJ (2004) Assessing the distinguishability of models and the informativeness of data. *Cogn Psychol* 49:47–84. [CrossRef](#) [Medline](#)
- O'Connell RG, Dockree PM, Kelly SP (2012) A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci* 15:1729–1735. [CrossRef](#) [Medline](#)
- Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J Vis* 5:376–404. [Medline](#)
- Paré M, Hanes DP (2003) Controlled movement processing: Superior colliculus activity associated with countermanded saccades. *J Neurosci* 23:6480–6489. [Medline](#)
- Pouget P, Logan GD, Palmeri TJ, Boucher L, Paré M, Schall JD (2011) Neural basis of adaptive response time adjustment during saccade countermanding. *J Neurosci* 31:12604–12612. [CrossRef](#) [Medline](#)
- Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ (2010) Neurally constrained modeling of perceptual decision making. *Psychol Rev* 117:1113–1143. [CrossRef](#) [Medline](#)
- Purcell BA, Schall JD, Logan GD, Palmeri TJ (2012) From salience to saccades: Multiple-alternative gated stochastic accumulator model of visual search. *J Neurosci* 32:3433–3446. [CrossRef](#) [Medline](#)
- Ramakrishnan A, Sureshbabu R, Murthy A (2012) Understanding how the brain changes its mind: microstimulation in the macaque frontal eye fields reveals how saccade plans are changed. *J Neurosci* 32:4457–4472. [CrossRef](#) [Medline](#)
- Ratcliff R (1978) A theory of memory retrieval. *Psychol Rev* 85:59–108. [CrossRef](#)
- Ratcliff R, Frank MJ (2012) Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. *Neural Comput* 24:1186–1229. [CrossRef](#) [Medline](#)
- Ratcliff R, McKoon G (2008) The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput* 20:873–922. [CrossRef](#) [Medline](#)
- Ratcliff R, Starns JJ (2013) Modeling response times, choices, and confidence judgments in decision making. *Psychol Rev* 120:697–719. [CrossRef](#) [Medline](#)
- Ratcliff R, Tuerlinckx F (2002) Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon Bull Rev* 9:438–481. [CrossRef](#) [Medline](#)
- Ratcliff R, Van Dongen HP (2011) Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proc Natl Acad Sci U S A* 108:11285–11290. [CrossRef](#) [Medline](#)
- Ratcliff R, Van Zandt T, McKoon G (1999) Connectionist and diffusion models of reaction time. *Psychol Rev* 102:261–300.
- Ratcliff R, Thapar A, McKoon G (2001) The effects of aging on reaction time in a signal detection task. *Psychol Aging* 16:323–341. [CrossRef](#) [Medline](#)
- Ratcliff R, Cheria A, Segraves M (2003) A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. *J Neurophysiol* 90:1392–1407. [CrossRef](#) [Medline](#)
- Ratcliff R, Hasegawa YT, Hasegawa RP, Smith PL, Segraves MA (2007) Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J Neurophysiol* 97:1756–1774. [Medline](#)
- Ratcliff R, Philiastides MG, Sajda P (2009) Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc Natl Acad Sci U S A* 106:6539–6544. [CrossRef](#) [Medline](#)
- Resulaj A, Kiani R, Wolpert DM, Shadlen MN (2009) Changes of mind in decision-making. *Nature* 461:263–266. [CrossRef](#) [Medline](#)
- Roitman JD, Shadlen MN (2002) Responses of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22:9475–9489. [Medline](#)
- Schurger A, Sitt JD, Dehaene S (2012) An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc Natl Acad Sci U S A* 109:E2904–E2913. [CrossRef](#) [Medline](#)
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464. [CrossRef](#)
- Shadlen MN, Kiani R (2013) Decision making as a window on cognition. *Neuron* 80:791–806. [CrossRef](#) [Medline](#)
- Shankar S, Massoglia DP, Zhu D, Costello MG, Stanford TR, Salinas E (2011) Tracking the temporal evolution of a perceptual judgment using a compelled-response task. *J Neurosci* 31:8406–8421. [CrossRef](#) [Medline](#)
- Starns JJ, Ratcliff R (2010) The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychol Aging* 25:377–390. [CrossRef](#) [Medline](#)
- Stone M (1960) Models for choice-reaction time. *Psychometrika* 25:251–260. [CrossRef](#)
- Thura D, Beauregard-Racine J, Fradet CW, Cisek P (2012) Decision making by urgency gating: theory and experimental support. *J Neurophysiol* 108:2912–2930. [CrossRef](#) [Medline](#)
- Wagenmakers E.-J., Ratcliff R, Gomez P, Iverson GJ (2004) Assessing model mimicry using the parametric bootstrap. *J Math Psychol* 48:28–50. [CrossRef](#)
- Wasserman L (2000) Bayesian model selection and model averaging. *J Math Psychol* 44:92–107. [CrossRef](#)
- Woodman GF, Kang MS, Thompson K, Schall JD (2008) The effect of visual search efficiency on response preparation: Neurophysiological evidence for discrete flow. *Psychol Sci* 19:128–136. [CrossRef](#) [Medline](#)