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

Guy E. Hawkins[a], Birte U. Forstmann[b], Eric-Jan Wagenmakers[c], Roger Ratcliff[d], and Scott D. Brown[a]

[a] School of Psychology, University of Newcastle, Australia
[b] Amsterdam Brain and Cognition, University of Amsterdam, The Netherlands
[c] Department of Psychology, University of Amsterdam, The Netherlands
[d] Department of Psychology, The Ohio State University, USA

Abstract

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 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 non-human 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.

Introduction

Over the past 50 years, psychology and neuroscience have contributed to a deeper understanding of decision-making based on "diffusion models" (Laming, 1968; Ratcliff, 1978; Stone, 1960). 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 (e.g., Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011; Ratcliff, 1978; Ratcliff & Van Dongen, 2011). More recently, the same models have been correlated with various components of neuroimaging and neurophysiological data in humans (e.g., Forstmann et al., 2008, 2010; O'Connell, Dockree, & Kelly, 2012; Ratcliff, Philiastides, & Sajda, 2009; Schurger, Sitt, & Dehaene, 2012). In *Macaca mulatta*, the firing rates of some neurons seem to behave like processes from diffusion models (e.g., Bollimunta, Totten, & Ditterich, 2012; Ding & Gold, 2010, 2012a; Hanes & Schall, 1996; Heitz & Schall, 2012; Kiani & Shadlen, 2009; Pouget et al., 2011; Purcell et al., 2010; Purcell, Schall, Logan, & Palmeri, 2012; Ratcliff, Cherian, & Segraves, 2003; Roitman & Shadlen, 2002; Woodman, Kang, Thompson, & Schall, 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, Hanes, Schall, & Stuphorn, 2008; Hanes, Patterson, & Schall, 1998; Paré & 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 Figure 1). Recently, however, a more complicated assumption has gained popularity: collapsing boundaries (solid curved lines, Figure 1), sometimes interpreted as urgency signals, where decisions are triggered by less and less evidence as time passes (Bowman, Kording, & Gottfried, 2012; Cisek, Puskas, & El-Murr, 2009; Ditterich, 2006a, 2006b; Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012; Milosavljevic, Malmaud, Huth, Koch, & Rangel, 2010; Thura, Beauregard–Racine, Fradet, & Cisek, 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 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 & Kiani, 2013). We addressed these problems in a large-scale survey and found, overall, evidence in favor of the fixed bound approach.
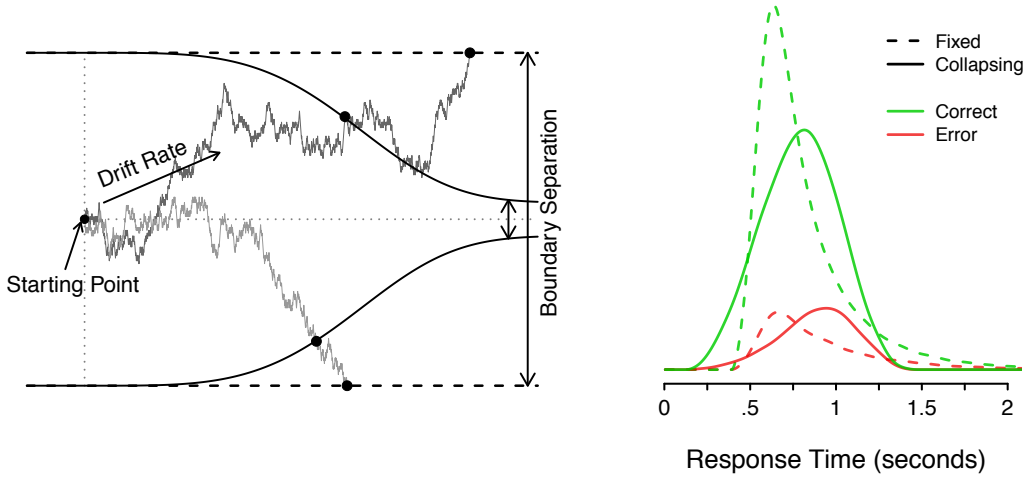
Figure 1.   The left panel illustrates diffusion models with fixed (dashed) or collapsing (solid) decision boundaries. A model with collapsing boundaries can terminate the evidence accumulation process earlier than a model with fixed boundaries, resulting in faster decisions. The right panel shows how the models lead to different predictions for response time distributions.

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 towards one boundary (drift rate, v); noise in the diffusion process (s, a scaling parameter not typically estimated from data, and fixed in all model fits in this paper); the separation of the boundaries (a); starting position of the diffusion process (z); and the time taken for non-decision processes (t_{er}). This basic diffusion model for decision-making is more than 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, 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 v and standard deviation η), and the non-decision time (uniform on the interval $t_{er} - \frac{st}{2}$, $t_{er} + \frac{st}{2}$).
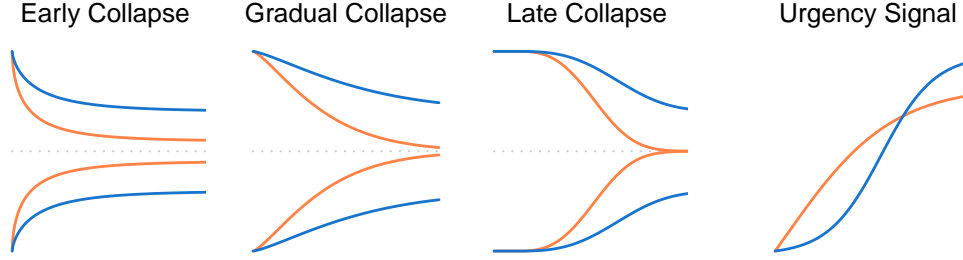
Figure 2. Various forms of dynamic diffusion models. The Weibull cumulative distribution function can generate decision boundaries that collapse at early or late stages of processing (left and middle right panels, respectively), or gradually throughout processing (middle left panel). Orange lines indicate 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 ($\frac{1}{2}a$). The right panel shows 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.

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 half-way point ($\frac{1}{2}a$), but was allowed to be some amount (a') away from that (see Figure 1). The lower boundary increased similarly from its initial value of zero to an asymptote the same distance below the half-way 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 (e.g., Ratcliff & Frank, 2012) or throughout processing to gradually impose a deadline on decision time (e.g., Bowman et al., 2012). We assumed 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 (Figure 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:

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

To all data sets, we fit four variants of the collapsing bounds model that corresponded to different psychological assumptions about the decision process. These assumptions informed

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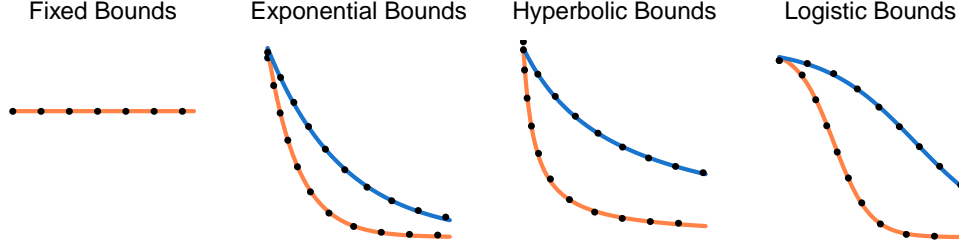
Figure 3. The Weibull cumulative distribution function mimics parametric forms that might be proposed for collapsing bounds. Orange and blue lines show examples of candidate stationary and dynamic boundaries of various parametric functions; fixed, exponential, hyperbolic, and logistic. Overlaid black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

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, Equation 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 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 employed.

We also considered a different kind of collapsing bounds model, instantiated via an "urgency signal" (Churchland, Kiani, & Shadlen, 2008; Cisek et al., 2009; Ditterich, 2006a; Hanks, Kiani, & Shadlen, 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 whose value increased with the duration of the decision (Figure 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:

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

where d is a delay parameter, and s_x and s_y are shape parameters (see Figure 2).
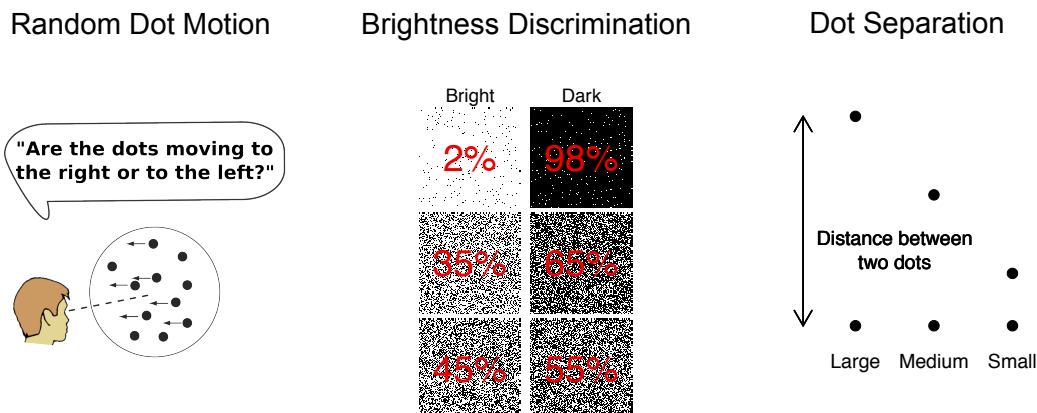
Figure 4. The three decision paradigms from the human and non-human primate studies: random dot motion, brightness discrimination, and dot separation.

We also ran our analyses twice: once with the fixed and collapsing bounds models described above, and once where the data were 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 & Tuerlinckx, 2002).

Data Sets for Modeling

Roitman & Shadlen (2002).

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; Figure 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 (*Macaca mulatta*) completed 2,614 trials (monkey B) and another completed 3,534 trials (monkey N).

Palmer, Huk, & Shadlen (2005).

Palmer, Huk, and Shadlen's (2005) Experiment 1 replicated Roitman and Shadlen's (2002) experiment 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 approximately 560 trials.

Ratcliff & McKoon (2008).

Ratcliff and McKoon's (2008) Experiment 1 replicated Palmer et al.'s (2005) experiment 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 approximately 960 trials.

Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves (2007).

Ratcliff, Hasegawa, Hasegawa, Smith, and Segraves (2007) had two monkeys make decisions about patches of pixels that varied in brightness (Figure 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%; and "dark" – 55%, 65%, 98%). Responses were indicated by saccades. One rhesus monkey (*Macaca mulatta*) completed 12,021 trials and another completed 7,632 trials.

Middlebrooks & Schall (2014).

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 Ratcliff et al.'s (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 (B: 41%, 45%, 48%, 50%, 52%, 55%, 59%, 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, where a signal would appear at a variable time following 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 data sets 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, Thapar, & McKoon (2001).

Ratcliff, Thapar, and McKoon's (2001) Experiment 2 (young subjects) had humans decide whether the distance between two dots was small or large (Figure 2). The separation of the dots ranged across 32 equal-sized steps from $\frac{11}{16}$in. (1.7cm, "small") to $1\frac{5}{16}$in. (2.4cm, "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 approximately 1,000 accuracy-emphasis trials, which contributed to the primary analyses, and approximately 1,000 speed-emphasis trials, which were used for a replication analysis. Responses were indicated by button presses.

Ratcliff, Cherian, & Segraves (2003).

Ratcliff et al. (2003) replicated Ratcliff et al.'s (2001) experiment but used two monkey participants. Methodological details were similar, except for the duration of data collection, dot separation that ranged from $2° - 10°$ of visual angle in increments of $1°$, no speed- or

accuracy-emphasis instructions, and responses indicated by saccades. One rhesus monkey (*Macaca mulatta*) completed 11,495 trials and another completed 5,037 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 Palmer et al.'s (2005) and Roitman and Shadlen's (2002). In particular, following a response participants did not receive feedback until at least one second had elapsed since stimulus presentation (44 participants) or two seconds (28 participants).

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, Brown, & Mewhort, 2002; Heathcote & 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 inter-decile bin. Since 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-bounds 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 0.01sec 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, Ratcliff, & Smith, 2006).

We adjusted the model parameters to optimize goodness-of-fit using differential evolution methods (Ardia, Mullen, Peterson, & Ulrich, 2013; Mullen, Ardia, Gil, Windover, & Cline, 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 (v) was estimated freely for each difficulty level (e.g., coherence level in random dot motion tasks, or brightness level in brightness discrimination). For example, Ratcliff and McKoon's (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), non-decision time (t_{er}), variability in non-decision 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)

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

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 data set (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 artefacts of the parameter estimation routines. We used a model recovery procedure (Navarro, Pitt, & Myung, 2004; Wagenmakers, Ratcliff, Gomez, & Iverson, 2004) that involves simulating many data sets from one model (e.g., the fixed bounds model) and fitting those simulated data sets with both the fixed and collapsing bounds models. We then assessed the difference in BIC between the two fitted models for each simulated data set. If model recovery is easy, 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 as this model was rarely preferred over the fixed or collapsing bounds models in data (3 of 93 subjects, see Figure 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 data sets from each model, with sample size equal to the sample size in the real data sets (with very large simulated sample sizes, model recovery was close to perfect). We ran the parameter estimation routine three times, independently, for each simulated data set and chose the best set of parameters for each simulated data set (i.e., the parameter set with the best value of objective function). In total, the model recovery procedure required approximately 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 data sets 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 (see Figure 4) – with experiments conducted in four independent laboratories. In total, we used data from eight non-human primates (*Macaca mulatta*) and 85 humans, with over 116,000 decisions in all.

For an overwhelming majority of people – 72 out of 85 human participants – the data supported the fixed bounds model over both the collapsing bounds and urgency signal models. Of the remaining 13 participants, 9 were best described by the collapsing bounds model, and results for the other 4 were inconclusive. Data from the 8 non-human primates painted a different picture, with 4 clearly supporting the collapsing bounds model, and 2 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 upper 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 show the average boundaries estimated for the fixed and collapsing bound models in each experiment, using parameters averaged across subjects. In the data sets 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 data sets 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 re-analyzing data from 17 participants in Ratcliff et al. (2001) with different instructions (emphasizing decision speed, rather than accuracy). Of these 91 additional human data sets, 64 showed very strong evidence for the fixed bounds model (i.e., $P_{\text{BIC}}(M_{fixed}|D) \geq .99$), and an additional 12 showed strong support (i.e., $P_{\text{BIC}}(M_{fixed}|D) \approx .95 - .99$). Of the remaining 15 subjects, 1 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 the drift rate, start-point of evidence accumulation, and non-decision time vary randomly across trials (Ratcliff & Tuerlinckx, 2002; Ratcliff, Van Zandt, & McKoon, 1999). The inclusion of these variability parameters has a demonstrated history of improving the fit of the fixed bounds model to data. For instance, Ratcliff (1978) demonstrated that between-trial variability in drift rate allows the fixed bound model to 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 thus 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 non-decision 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 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 & 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 & McKoon, 2008).

For data collected from non-human 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
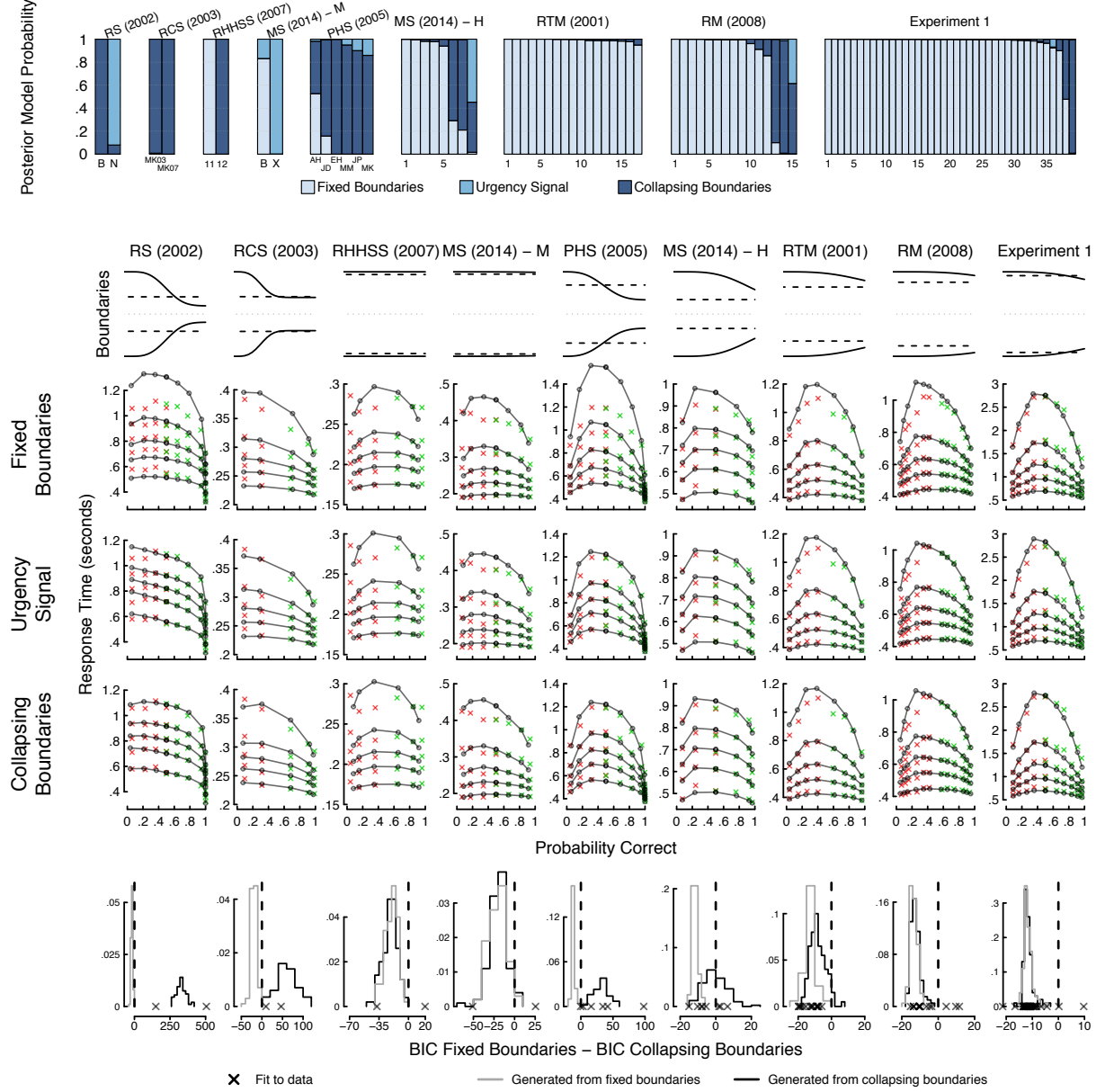
Figure 5. Results. **Top Row:** BIC-based approximations to posterior model probabilities in favor of the fixed bounds, urgency signal, and collapsing bounds models, separately for the nine data sets. Shade represents the three models, columns represent individual subjects, with subject labels from the 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 data sets: 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.
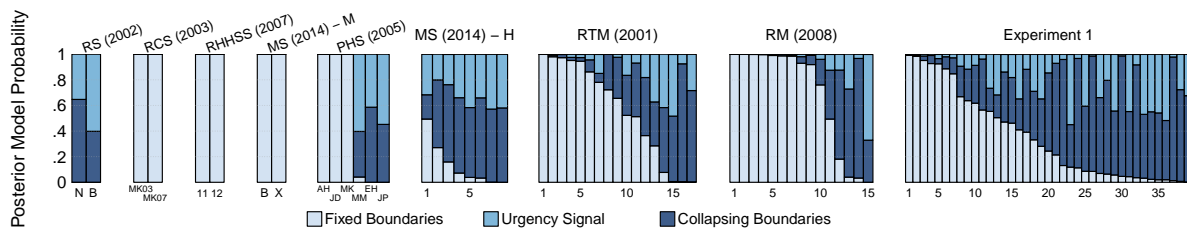
Figure 6. Approximations to posterior model probabilities in favor of the fixed bounds model with between trial variability parameters, and the urgency signal and collapsing bounds models without between trial variability parameters. All details are as described in the upper row of Figure 5.

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 data sets 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, since it received so little support from the data.

The lower 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 data sets modeled after the experiments that supported the use of the collapsing bounds model (see second row of Figure 5; Palmer et al., 2005; Ratcliff et al., 2003; Roitman & 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 where the estimated parameters for the collapsing bounds model were similar to the fixed bounds model, data sets simulated from the collapsing bounds model are frequently classified as fixed bounds, because the estimated boundaries for the collapsing bounds (second-top row of the figure) 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), since the collapsing bounds model is penalized due to its additional parameters; complexity which 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 (e.g., Albantakis & Deco, 2009, 2011; Basten, Biele, Heekeren, & Fiebach, 2010; Boucher, Palmeri, Logan, & Schall, 2007; Fetsch, Kiani, Newsome, & Shadlen, 2014; Hanes & Schall, 1996; Ho, Brown, & Serences, 2009; Kiani & Shadlen, 2009; Lo & Wang, 2006; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Purcell et al., 2010, 2012; Ramakrishnan, Sureshbabu, & Murthy, 2012; Ratcliff et al., 2003, 2007; Ratcliff & Starns, 2013; Resulaj, Kiani, Wolpert, & Shadlen, 2009; Shankar et al., 2011). More complex non-stationary models, with collapsing boundaries or increasing urgency signals, have recently become popular, especially in some neurophysiological studies of primates (e.g., Churchland et al., 2008; Ditterich, 2006a; Drugowitsch et al., 2012) but not all (e.g., Purcell et al., 2010, 2012). 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 non-human primate participants (6 of 8). 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 (see Figure 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 where participants completed only a single session of participation (Experiment 1; Ratcliff & McKoon, 2008) or two sessions (Ratcliff et al., 2001). Practice is known to influence boundary settings in decision-making models (e.g., Balci et al., 2011; Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Starns & Ratcliff, 2010), and so it is plausible that collapsing bounds models provide good descriptions of data from extremely highly practiced participants, though this remains an open question. Indeed, there are a number of data sets in the literature that we did not explore that demonstrate a good fit of fixed bounds models to data from highly practiced participants (e.g., Ding & Gold, 2012b; Purcell et al., 2010, 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 non-human primates) until a response was registered or a minimum time post-stimulus onset had elapsed (e.g., 1 second), 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 accu-

mulation 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 & Shadlen, 2002).

It appears that subjects might learn the delayed response strategy following considerable task practice. Roitman and Shadlen's (2002) non-human primates completed thousands of decision trials, and all of Palmer et al.'s (2005) participants had previous experience with similar psychophysical experiments, prior to participation in the target experiment. This strategy was not observed in our replication experiments, where 74 humans completed a single session under conditions similar to Palmer et al. (2005) with a minimum delayed feedback time of one or two seconds; those provided strong support for the fixed bounds model.

Overall, our survey suggests that humans and non-human 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 non-human 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 Transactions on Automatic Control*, *19*, 716–723.

Albantakis, L., & Deco, G. (2009). The encoding of alternatives in multiple-choice decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10308–10313.

Albantakis, L., & Deco, G. (2011). Changes of mind in an attractor network of decision–making. *PLoS Computational Biology*, *7*, e1002086.

Ardia, D., Mullen, K. M., Peterson, B. G., & Ulrich, J. (2013). *DEoptim: Differential evolution in R.* (R package version 2.2-2)

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception & Psychophysics*, *73*, 640–657.

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Science*, *107*, 21767–21772.

Bollimunta, A., Totten, D., & Ditterich, J. (2012). Neural dynamics of choice: Single–trial analysis of decision–related activity in parietal cortex. *Journal of Neuroscience*, *32*, 12684–12701.

Boucher, L., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2007). Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review*, *114*, 376–397.

Bowman, N. E., Kording, K. P., & Gottfried, J. A. (2012). Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. *Neuron*, *75*, 916–927.

Brown, J. W., Hanes, D. P., Schall, J. D., & Stuphorn, V. (2008). Relation of frontal eye field activity to saccade initiation during a countermanding task. *Experimental Brain Research*, *190*, 135–151.

Brown, S., Ratcliff, R., & Smith, P. (2006). Evaluating methods for approximating stochastic differential equations. *Journal of Mathematical Psychology*, *50*, 402–410.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency–gating model. *The Journal of Neuroscience*, *29*(37), 11560–11571.

Ding, L., & Gold, J. I. (2010). Caudate encodes multiple computations for perceptual decisions. *Journal of Neuroscience*, *30*, 15747–15759.

Ding, L., & Gold, J. I. (2012a). Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex*, *22*, 1052–1067.

Ding, L., & Gold, J. I. (2012b). Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron*, *75*, 865–874.

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

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

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience, 32*, 3612–3628.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review, 16*, 1026–1036.

Edwards, A. W. F. (1992). *Likelihood.* Baltimore, MD: The Johns Hopkins University Press.

Feller, W. (1968). *An introduction to probability theory and its applications: Vol. I.* New York: John Wiley & Sons.

Fetsch, C. R., Kiani, R., Newsome, W. T., & Shadlen, M. N. (2014). Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron, 83*, 797–804.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., Bogacz, R., & Turner, R. (2010). Cortico–striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Science, 107*, 15916–15920.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre–SMA facilitate decision–making under time pressure. *Proceedings of the National Academy of Science, 105*, 17538–17542.

Hanes, D. P., Patterson, W. F., II, & Schall, J. D. (1998). Role of frontal eye fields in countermanding saccades: Visual, movement, and fixation activity. *Journal of Neurophysiology, 79*, 817–834.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*, 427–430.

Hanks, T., Kiani, R., & Shadlen, M. N. (2014). A neural mechanism of speed–accuracy tradeoff in macaque area LIP. *eLife, 3*, doi: 10.7554/eLife.02260.

Heathcote, A., & Brown, S. D. (2004). Reply to Speckman and Rouder: A theoretical basis for QML. *Psychonomic Bulletin & Review, 11*, 577–578.

Heathcote, A., Brown, S. D., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9*, 394–401.

Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed–accuracy tradeoff. *Neuron, 76*, 616–628.

Ho, T., Brown, S., & Serences, J. (2009). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience, 29*, 8675–8687.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*, 759–764.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and comparison of value in simple choice. *Nature Neuroscience, 13*, 1292–1298.

Krajbich, I., & Rangel, A. (2011). A multi–alternative drift diffusion model predicts the relationship between visual fixations and choice in value–based decisions. *Proceedings of the National Academy of Science, 108*, 13852–13857.

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

Lo, C.-C., & Wang, X.-J. (2006). Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience, 9*, 956–963.

Middlebrooks, P. G., & Schall, J. D. (2014). Response inhibition during perceptual decision making in humans and macaques. *Attention, Perception & Psychophysics, 76*, 353–366.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reactime of value–based choices under high and low time pressure. *Judgment and Decision Making*, *5*, 437–449.

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, *32*, 2335–2343.

Mullen, K., Ardia, D., Gil, D., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, *40*, 1–26.

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, *49*, 47–84.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation–to–bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729–1735.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*, 376–404.

Paré, M., & Hanes, D. P. (2003). Controlled movement processing: Superior colliculus activity associated with countermanded saccades. *Journal of Neuroscience*, *23*, 6480–6489.

Pouget, P., Logan, G. D., Palmeri, T. J., Boucher, L., Paré, M., & Schall, J. D. (2011). Neural basis of adaptive response time adjustment during saccade countermanding. *Journal of Neuroscience*, *31*, 12604–12612.

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*, 1113–1143.

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

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. *Journal of Neuroscience*, *32*, 4457-4472.

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

Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two–choice decisions. *Journal of Neurophysiology*, *90*, 1392–1407.

Ratcliff, R., & Frank, M. (2012). Reinforcement–based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. *Neural Computation*, *24*(1186–1229).

Ratcliff, R., Hasegawa, Y. T., Hasegawa, Y. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single–cell recording data from the superior colliculus in a brightness–discrimination task. *Journal of Neurophysiology*, *97*, 1756–1774.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two–choice decision tasks. *Neural Computation*, *20*, 873–922.

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

Ratcliff, R., & Starns, J. J. (2013). Modeling response times, choices, and confidence judgments in decision making. *Psychological Review*, *120*, 697–719.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*, 323–341.

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

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

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *102*, 261–300.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision–making. *Nature*, *461*, 263–266.

Roitman, J. D., & Shadlen, M. N. (2002). Responses of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*, 9475–9489.

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self–initiated movement. *Proceedings of the National Academy of Science*, *109*, 2904–2913.

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

Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*, 791–806.

Shankar, S., Massoglia, D. P., Zhu, D., Costello, M. G., Stanford, T. R., & Salinas, E. (2011). Tracking the temporal evolution of a perceptual judgment using a compelled–response task. *Journal of Neuroscience*, *31*, 8406–8421.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, *25*, 377–390.

Stone, M. (1960). Models for choice–reaction time. *Psychometrika*, *25*, 251–260.

Thura, D., Beauregard–Racine, J., Fradet, C. W., & Cisek, P. (2012). Decision making by urgency gating: Theory and experimental support. *Journal of Neurophysiology*, *108*, 2912–2930.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*, 28–50.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*, 92–107.

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