Journal of Mathematical Psychology 73 (2016) 59–79

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Comparing fixed and collapsing boundary versions of the diffusion model

Chelsea Voskuilen [a,*], Roger Ratcliff [a], Philip L. Smith [b]

[a] The Ohio State University, United States
[b] The University of Melbourne, Australia

HIGHLIGHTS

- We compare fixed bound and collapsing bound versions of the diffusion model.
- An appropriate model selection method prefers the fixed bound model for our data.
- The two models produce response time distributions with similar shape.
- The estimated collapsing bounds were similar to estimated fixed bounds because the amount of collapse was small.
- Bounds that collapse over time do not provide an improvement over fixed bounds.

ARTICLE INFO

Article history:
Received 29 January 2015
Received in revised form
8 April 2016
Available online 24 May 2016

Keywords:
Response time models
Diffusion model
Collapsing boundaries
Model selection

ABSTRACT

Optimality studies and studies of decision-making in monkeys have been used to support a model in which the decision boundaries used to evaluate evidence collapse over time. This article investigates whether a diffusion model with collapsing boundaries provides a better account of human data than a model with fixed boundaries. We compared the models using data from four new numerosity discrimination experiments and two previously published motion discrimination experiments. When model selection was based on BIC values, the fixed boundary model was preferred over the collapsing boundary model for all of the experiments. When model selection was carried out using a parametric bootstrap cross-fitting method (PBCM), which takes into account the flexibility of the alternative models and the ability of one model to account for data from another model, data from 5 of 6 experiments favored either fixed boundaries or boundaries with only negligible collapse. We found that the collapsing boundary model produces response times distributions with the same shape as those produced by the fixed boundary model and that its parameters were not well-identified and were difficult to recover from data. Furthermore, the estimated boundaries of the best-fitting collapsing boundary model were relatively flat and very similar to those of the fixed-boundary model. Overall, a diffusion model with decision boundaries that converge over time does not provide an improvement over the standard diffusion model for our tasks with human data.

1. Introduction

Over the last 50 years, sequential sampling models such as the diffusion model have been applied extensively to a wide variety of tasks and participant populations. The tasks include recognition memory (Ratcliff, 1978; Ratcliff, Thapar, & McKoon, 2011; Starns, Ratcliff, & McKoon, 2012), lexical decisions (Ratcliff, Gomez, & McKoon, 2004; Ratcliff, Thapar, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008), perceptual discrimination (Ratcliff, 2014; Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999; Smith, Ratcliff, & Sewell, 2014), value-based decision making (Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011; Milosavljevic, Malmaud, Huth, Koch, & Rangel, 2010) go/no-go tasks (Gomez, Ratcliff, & Perea, 2007); simple reaction time (Ratcliff & Strayer, 2014; Ratcliff & Van Dongen, 2011; Smith, 1995), the response signal task (Ratcliff, 2006, 2008), and visual signal detection (Smith & Ratcliff, 2009; Smith, Ratcliff, & Wolfgang, 2004). The participant populations include older adults (Ratcliff, Thapar, & McKoon, 2001, 2003, 2004,

* Correspondence to: The Ohio State University, 291 Psychology Building, 1835 Neil Avenue, Columbus, OH 43210, United States. Tel.: +1 206 949 3086.
E-mail address: voskuilen.2@osu.edu (C. Voskuilen).

http://dx.doi.org/10.1016/j.jmp.2016.04.008
0022-2496/

2010; Ratcliff et al., 2011; Spaniol, Madden, & Voss, 2006), children and adolescents (Ratcliff, Love, Thompson, & Opfer, 2012), children with ADHD (Mulder et al., 2010), children with dyslexia (Zeguers et al., 2011), people undergoing sleep deprivation (Ratcliff & Van Dongen, 2009), people with induced hypoglycemia (Geddes et al., 2010), and people with anxiety or depression (White, Ratcliff, Vasey, & McKoon, 2010a,b). In sequential sampling models with fixed boundaries, noisy evidence about a stimulus is accumulated over time toward one of two decision boundaries, each of which represents a decision alternative. The distance between the boundaries may vary from trial to trial but they remain constant within trials, that is, from the time of stimulus presentation until the time of the response. Such models aim to account for all aspects of the experimental data, namely, the distributions of correct and error response times and the proportions of correct and error responses.

These fixed or constant boundary models have been able to successfully fit a wide variety of data (see previous paragraph). However, some researchers have instead proposed models with decision boundaries that converge over time during a trial, such that progressively less evidence is required to trigger a response as the trial progresses (Bowman, Kording, & Gottfried, 2012; Churchland, Kiani, & Shadlen, 2008; Cisek, Puskas, & El-Murr, 2009; Ditterich, 2006a,b; Hanks, Mazurek, Kiani, Hopp, & Shadlen, 2011; Rao, 2010; Sanders & Ter Linden, 1967; Thura, Beauregard-Racine, Fradet, & Cisek, 2012; Thura & Cisek, 2014; Viviani, 1979a,b; Viviani & Terzuolo, 1972).

Support for these collapsing or converging boundary models has come from several sources. In the neurophysiological literature, firing rates in the lateral interparietal area of macaques appear to reflect the accumulation of evidence for a particular choice in a dot-motion task (Churchland et al., 2008). Because the increase in firing rate is also found on trials on which there is no coherent motion, the authors proposed that the time-dependent increase in firing rates represents an urgency-signal, which they argued is equivalent to a collapsing boundary (but this is not true for other urgency-signal implementations, e.g. Thura et al., 2012; Thura & Cisek, 2014). Models with collapsing boundaries may also be able to represent certain physiological properties such as the refractory period after a neuron has fired (Kryukov, 1976) or certain dynamics of the basal ganglia (Ratcliff & Frank, 2012).

In the behavioral literature, numerous researchers have demonstrated that subjects are willing to make decisions based on less evidence as a trial progresses in expanded judgment or deferred decision-making tasks (Busemeyer & Rapoport, 1988; Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012; Rapoport & Burkheimer, 1971; Sanders & Ter Linden, 1967). However, these kinds of tasks are quite different from the standard two-choice tasks used in most speeded decision-making studies. In most expanded judgment tasks, evidence is presented at a slower rate (e.g., a new piece of information every 2 s) and subjects' response times can be much longer (e.g., 20–40 s). Some of these tasks also include an explicit cost associated with acquiring new information. Given the differences between these tasks and the standard two-choice tasks, evidence for collapsing boundary models in expanded judgment tasks cannot be taken as evidence of their appropriateness for standard rapid two-choice tasks.

Some researchers have also argued in favor of models with collapsing boundaries on the grounds of optimality (Busemeyer & Rapoport, 1988; Deneve, 2012; Ditterich, 2006a; Drugowitsch et al., 2012; Rapoport & Burkheimer, 1971; Thura et al., 2012). If subjects are trying to achieve a short mean response time for a given level of accuracy and a single fixed stimulus strength, then a model with fixed boundaries is statistically optimal, assuming that there is no variability in drift rate across trials (Moran, 2015; Wald & Wolfowitz, 1948). However, if there is variability in drift rate across trials, then a model with boundaries that change over time is statistically optimal if subjects are trying to maximize their reward rate (i.e., make the most correct responses per unit of time; Ditterich, 2006a; Thura et al., 2012) or minimize mean response time for a given level of accuracy (Moran, 2015). A model with boundaries that collapse over time has also been argued to be optimal if there is a cost associated with time spent on a decision (Busemeyer & Rapoport, 1988; Drugowitsch et al., 2012; Rapoport & Burkheimer, 1971). It has also been argued that collapsing boundary models are optimal in response-deadline tasks in which subjects are trying to find a balance between being accurate and still making a response before a deadline (Frazier & Yu, 2008), although in practice, subjects do not appear to behave optimally in these kinds of tasks (Balci et al., 2011; Karsilar, Simen, Papadakis, & Balci, 2014), and this type of data can be accounted for with a forced decision at the deadline (Ratcliff, 1988, 2006). Karsilar et al. (2014) examined accuracy as a function of response time in a response deadline task and found that subjects do show a decrease in accuracy before a response deadline, but the size of the decrease is more consistent with the predictions of a fixed boundary diffusion model than with the predictions of a collapsing boundary model with boundaries selected to optimize reward rate.

Although collapsing boundary models have become increasingly popular in both the behavioral and neurophysiological literature, many researchers using sequential-sampling models in these domains have successfully used fixed boundary versions of these models (Bode et al., 2012; Brown, Hanes, Schall, & Stuphorn, 2008; Ding & Gold, 2010, 2012; Forstmann et al., 2010, 2008; O'Connell, Dockree, & Kelly, 2012; Ramakrishnan & Murthy, 2013; Ramakrishnan, Sureshbabu, & Murthy, 2012; Ratcliff, Philiastides, & Sajda, 2009; Salinas & Stanford, 2013; Schall, 2003; Schurger, Sitt, & Dehaene, 2012; Smith & McKenzie, 2011; Usher & McClelland, 2001; Wang, 2002; Wong & Wang, 2006). Hawkins, Forstmann, Wagenmakers, Ratcliff, and Brown (2015) recently conducted a large-scale analysis of data from both humans and non-human primates from several different tasks and found that support for fixed or collapsing boundary versions of the model appears to depend on the specific tasks and procedures used. Overall, Hawkins et al. found that a fixed-boundary model was preferred for most of the human subjects, but a collapsing boundary model was preferred for most of the non-human primates. This may be a result of practice effects, given that the non-human primates in these studies received much more extensive training on these tasks (e.g., months) than the human subjects.

The results we present in this article complement and extend those of Hawkins et al. (2015) in several ways. First, unlike Hawkins et al., who used Monte Carlo simulation to obtain predictions for the collapsing boundary model, we used exact predictions based on the integral equation methods described in Smith (2000). These methods make it possible to derive exact response time distribution and accuracy predictions for diffusion models with time-varying drift rates or boundaries or both. This ensures that our model predictions are reliable and were not dependent on our simulation method or sample size. Second, the collapsing boundary model we examined had a different functional form for the boundary collapse than the one used by Hawkins et al. We used a hyperbolic ratio function, which Shadlen and colleagues (Churchland et al., 2008; Hanks et al., 2011) previously used to quantify the hypothetical urgency signal in physiological data, whereas Hawkins et al. used a Weibull function (which was chosen for its ability to mimic a wide range of possible boundary collapse functions). The hyperbolic ratio function we used can be well-approximated by a Weibull function, but has fewer parameters than the Weibull function and is more constrained in terms of the variety of shapes it can take. Third, we applied our methods to a different set of tasks and manipulations than the ones considered by Hawkins et al., so our

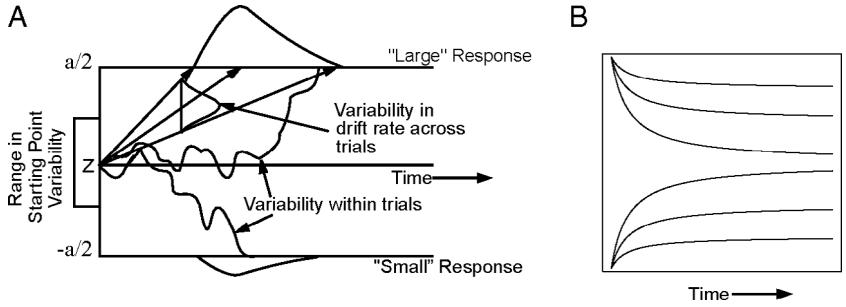

Fig. 1. Diffusion model with fixed or collapsing boundaries. A: Diffusion model with fixed boundaries. Evidence accumulation begins at starting point z and continues until one of the boundaries ($a/2$ and $-a/2$) is reached. B: Collapsing boundaries for several values of κ (0.3, 0.6, and 0.9) and one value of $t_{0.5}$ (0.15).

results are important in establishing the generality or otherwise of their conclusions. To foreshadow our main conclusion, we find that most of our data were more parsimoniously described by a fixed-boundary than by a collapsing boundary model, in agreement with the findings of Hawkins et al. for their human data. Together with the results of Hawkins et al., our findings show that fixed-boundary models provide a good account of performance from a variety of speeded decision tasks and serve as a caution against overgeneralization from physiological data and expanded judgment tasks.

2. Model parameterization

In the diffusion model, noisy evidence is accumulated over time from a starting point toward one of two decision boundaries, as shown in Fig. 1(A). The starting point is denoted z and the boundaries are denoted $a/2$ and $-a/2$ such that the distance between the two boundaries is determined by the parameter a. This parameterization differs from the one used in the standard diffusion model, in which the boundaries are set to a and 0. We parameterize the model in this way for ease of comparison between the fixed-boundary and collapsing boundary models. The rate of evidence accumulation is called the *drift rate* (ν) and is determined by the quality of the information extracted from the stimulus. There is noise in the accumulation of evidence within each trial, represented by the infinitesimal standard deviation (s), which is the square-root of the diffusion coefficient. In our fits, s acts as a scaling parameter which is fixed across conditions.[1] Response time predictions are obtained by combining the decision time (the time taken for the accumulating evidence to reach one of the boundaries) with a uniformly distributed non-decision component. The non-decision component, which encompasses both encoding (i.e., the transformation of the stimulus representation to a decision-related representation) and response output processes, is assumed to be uniformly distributed with mean T_{er} and range s_t.

The values of drift rate and starting point (or, equivalently, the starting points of the decision boundaries) vary from trial to trial. The across-trial variability in drift rate is assumed to be normally distributed with a standard deviation of η. The variability in the starting point is assumed to be uniformly distributed with a range of s_z. These distributional assumptions are not critical and the use of alternative distributions does not significantly change the estimates of the diffusion model's other parameters (Ratcliff, 2013). The inclusion of these parameters, however, is necessary for the model to be able to produce different patterns of correct and error response times (Ratcliff & McKoon, 2008).

In the collapsing boundary version of the diffusion model, noisy evidence is accumulated toward boundaries that decrease over time (as in Fig. 1(B)). All aspects of this model are identical to the fixed boundary model, except for the addition of two parameters that govern the boundary collapse. The boundaries are parameterized as functions of the initial boundary separation (a), time (t), the amount of collapse (κ), and a semi-saturation constant ($t_{0.5}$). The expressions for the boundaries are:

$$a_1(t) = \frac{a}{2}\left[1 - \kappa\frac{t}{t + t_{0.5}}\right]$$

$$a_2(t) = -\frac{a}{2}\left[1 - \kappa\frac{t}{t + t_{0.5}}\right].$$

The semi-saturation constant is the value of time at which the boundaries have collapsed by half κ. Small values of $t_{0.5}$ represent a faster rate of collapse. This is the same parameterization as used by Hanks et al. (2011), except that the amount of collapse is expressed as a fraction of the total boundary separation rather than as an absolute value. This parameterization simplified fitting and led to more interpretable parameter estimates. Fig. 1(B) shows boundaries for a single value of $t_{0.5}$ (0.15) and three values of κ (0.3, 0.6, 0.9). We chose this functional form for the collapsing boundaries since it has previously been found to provide a reasonable fit to data from both humans and monkeys (Churchland et al., 2008; Hanks et al., 2011).

Our version of the collapsing boundary model differs slightly from others in the literature in that it includes between-trial variability in non-decision time, starting point, and drift rates. In the fixed boundary version of the diffusion model, these parameters are necessary to produce fast and slow errors and RT distributions with the correct shape (Ratcliff, 1981; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Ratcliff et al., 1999; Wagenmakers et al., 2008). Proponents of the collapsing boundary model have often used simpler versions of the diffusion model which do not include these parameters or include only some of them (Ditterich, 2006a,b; Palmer, Huk, & Shadlen, 2005; Shadlen & Kiani, 2013) and have argued that the collapsing boundary model is capable of producing slow errors without these sources of variability (fast errors are generally not addressed, and variability in non-decision time is only sometimes used; Ditterich, 2006a). The model without between-trial variability is a limiting case of the model we tested and if there was no between-trial variability in model parameters then the fitting method would have produced estimates of these parameters that were near zero. However, we did not find this to be the case in fits to our data. Both the fixed and collapsing boundary models were unable to capture the behavior of the error response time distributions without these between-trial sources of variability. Our goal in carrying out these models comparisons was to determine if one of these versions of the diffusion model is better than the other, meaning both that it provides a closer fit to the observed data and also that it provides meaningful and interpretable parameter values.

[1] Constraining s across all of the conditions is not strictly necessary from a scaling perspective—it technically only needs to be fixed for one of the conditions (Donkin, Brown, & Heathcote, 2009). However, in all of our fits the more constrained model (with s fixed across all conditions) was able to fit the observed data.

3. Parametric bootstrap cross-fitting method

It is well established that any reasonable model selection process should involve a trade-off between the goodness-of-fit (GOF) of the models in question and their parsimony (Myung, 2000; Pitt & Myung, 2002; Wagenmakers, Ratcliff, Gomez, & Iverson, 2004). The complexity of a model is determined by both the number of parameters in the model and the functional form of the model, and is equivalent to model flexibility (i.e., the ability of the model to fit diverse patterns of data). However, many simple model selection methods, such as the likelihood ratio test (LRT; Wilks, 1938), Akaike information criterion (AIC; Akaike, 1973; Burnham & Anderson, 2002; Parzen, Tanabe, & Kitagawa, 1998), and Bayesian information criterion (BIC; Raftery, 1995; Schwarz, 1978), treat model complexity only as a function of the number of parameters in a model. This type of approach ignores differences in the functional forms of the models being compared and does not adequately address model mimicry. To address these problems, we used a general resampling procedure known as the parametric bootstrap cross-fitting method (PBCM; Wagenmakers et al., 2004).

The parametric bootstrap cross-fitting method (PBCM) is a procedure designed to quantify the model mimicry of a pair of models (i.e., the ability of one model to account for data generated from another model). There are several steps in this procedure. First, the two models being compared are fit to the data, resulting in a set of parameter values for each model.[2] Second, simulated data are generated from each of the models using the parameter values from the fits. Third, each model is fit to each of the simulated data sets, resulting in four GOF values (2 models $\times$ 2 sets of simulated data). These latter two steps are then repeated multiple times to obtain distributions of GOF values. The difference in the GOF of the two models given the data is then examined to compare how well each model is able to fit data generated from the other model. Ideally, each model should provide a closer fit to its own data than to data generated from the other model. If we plot histograms of the differences in GOF values from two models A and B, we should see something like the distributions in the top panel of Fig. 2, with the crossover point of the two distributions at or around zero. When the data are generated from model A, then model A produces smaller GOF values and so the distribution of GOF_A–GOF_B values is mostly negative. When the data are generated from model B, then model B produces smaller GOF values and the distribution of differences is mostly positive. Researchers commonly fit two or more models to their data, possibly with some kind of parameter-counting correction (as in BIC), and then choose the model with the smallest GOF as the best-fitting model.

However, this simple model comparison approach does not sufficiently take into account differences in model flexibility and the fact that adjusting parameters of the model may have differential effects on the functional form of the model predictions (e.g., changes in mean non-decision time will change the position of all of the RT distributions whereas changes in the variability in non-decision time will affect the leading edges of the RT distributions and also spread the distributions out slightly). Models can differ significantly in their ability to fit various data patterns such that an overly flexible model may be able to produce smaller GOF values than a more restricted model, even when being fit to data generated from the more restricted model. This flexibility may result in differences in GOF values that exceed the theoretical

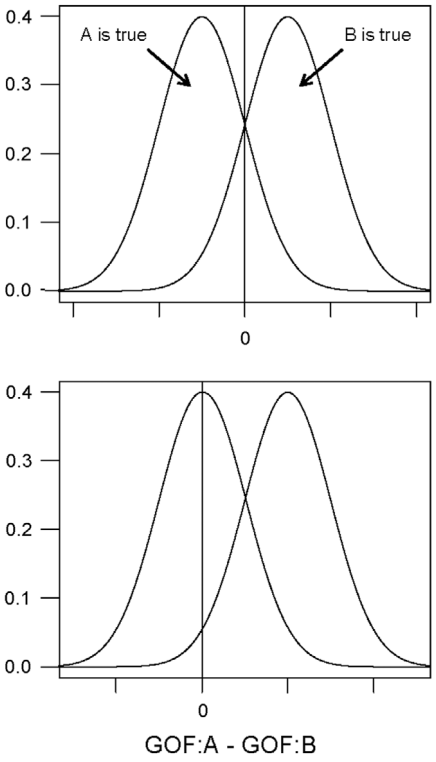

Fig. 2. PBCM differences in GOF densities and decision criteria. The top panel depicts distributions of GOF differences for the situation where each model generally provides smaller GOF values when fit to data generated from that model such that the crossover point between the two distributions is around zero. The bottom panel depicts distributions of GOF differences for the situation where one of the models is able to produce consistently smaller GOF values than the other (e.g., because it is more flexible).

corrections based on the number of model parameters in measures such as BIC or AIC. In this case, the resulting histograms of the differences in GOF from the two models could look like the distributions in the bottom panel of Fig. 2. If model B produced consistently smaller GOF values than model A, both of the GOF distributions would be shifted to the right. Using a criterion of zero would then result in a bias toward choosing the more flexible model. PBCM is a useful procedure for such situations because it provides an appropriate criterion for choosing between the two models. Rather than assuming that the appropriate criterion is at zero (i.e., choosing the model with the smaller GOF), the PBCM determines the appropriate criterion based on the observed distribution of GOF differences between the two models. The distributions in Fig. 2 are intended to be illustrative of a general model-fitting approach and to show the advantage of using PBCM. In the case of the models we consider here, the fixed boundary model is nested in the collapsing boundary model and so the more complex collapsing boundary model should always produce smaller GOF values and the crossover point for the differences in GOF values from the two models will never be at zero (at least for GOF methods such as G^2 or χ^2 that do not include penalty terms based on the number of parameters). In other words, the distributions for the GOF of the fixed boundary model minus the GOF for the collapsing boundary model should look like the distributions in the bottom panel of Fig. 2 and the PBCM will allow us to calculate the optimal criterion for model selection.

PBCM is a useful method for model comparison for several reasons. First, it is flexible in its ability to examine either the entire parameter space of a model or just part of the space. Depending on the question of interest, the PBCM can be set up to sample the entire parameter spaces of the competing models and so provide

[2] A non-parametric version of this method starts with fitting multiple bootstrapped samples of the data with each of the models and then generating predictions from each of those fits. The parametric approach, as described above, was more appropriate given our focus on average parameter values across subjects, but both approaches have merit.

insight into the full range of predictions a model could make, or it can be set up to investigate only the part of the parameter space that is relevant to a particular data set. PBCM is also a fairly intuitive approach to model selection. Even for very complex models, the concept of investigating how well each model can account for data from a competing model is easily understood. Finally, and perhaps most important from a practical standpoint, the method is straightforward to implement as it requires no more code than would already be needed to fit models to data and generate predictions and simulated data.

4. Experiments

The experiments were designed to provide data from simple speeded decision tasks in order to allow the performance of the two versions of the diffusion model to be compared. Our goal was to determine if one of the models provided a better account overall of the accuracy and response time distributions of human subjects performing standard decision-making tasks and, more generally, to compare the overall performance of the two models. The first four experiments used a numerosity discrimination task with a bias manipulation. The fixed-boundary diffusion model has previously been shown to provide a good fit to data from this type of task (Geddes et al., 2010; Leite & Ratcliff, 2011; Ratcliff, 2008, 2014; Ratcliff et al., 2012, 2001, 2010; Ratcliff & Van Dongen, 2009; Ratcliff et al., 1999). The fifth and sixth experiments used a dot motion discrimination task with various difficulty levels and a speed/accuracy manipulation. This task has been used extensively in both the neurophysiological and behavioral literature (e.g. Cain, Barreiro, Shadlen, & Shea-Brown, 2013; Churchland et al., 2008; Ditterich, 2006a; Forstmann et al., 2010, 2008; Gold & Shadlen, 2007; Hanks, Kiani, & Shadlen, 2014; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Morgan & Ward, 1980; Mulder, Boekel, Ratcliff, & Forstmann, 2014; Niwa & Ditterich, 2008; Palmer et al., 2005; Ratcliff & McKoon, 2008; Ratcliff & Starns, 2013; Roitman & Shadlen, 2002; Shadlen & Newsome, 1996; van Ravenzwaaij, Mulder, Tuerlinckx, & Wagenmakers, 2012). We present the fits of the two models to all of the experiments followed by the PBCM results and interpretation.

4.1. Numerosity discrimination: Experiments 1–4

Experiments 1–4 used a two-choice numerosity discrimination task in which subjects had to judge whether the number of asterisks in a 10 × 10 array was greater or less than 50. All four experiments included a bias manipulation in which the relative frequencies of large and small stimuli were varied in different blocks. In some blocks there were an equal number of large and small trials; in others, there were three times more of one type than the other. These experiments differed only in the way in which the number of asterisks in the display was distributed across trials. We originally hypothesized that the range of asterisk values and the relative frequency of extreme asterisk values might induce different response patterns or strategies. However, this did not prove to be the case.

4.1.1. Method

4.1.1.1. Subjects. A total of sixty-three Ohio State University undergraduates participated in these experiments for credit in an introductory psychology course. Twenty-one subjects participated in Experiment 1, 11 in Experiment 2, 22 in Experiment 3, and 9 in Experiment 4.

4.1.1.2. Materials. The stimuli for the numerosity experiments consisted of asterisks presented in a 10×10 array with the number of possible asterisks ranging from 3 to 98 (a smaller range was used in some of the experiments). The positions to be filled with asterisks were selected randomly from 100 positions in the 10×10 array. Experiments 1–4 differed only in the way in which the number of asterisks in the display was distributed across trials. In Experiments 1 and 2, the numbers of asterisks were distributed such that more extreme numbers of asterisks were less common than intermediate numbers. The distribution for Experiment 1 consisted of three uniform distributions and the distribution for Experiment 2 was approximately normal. The means and standard deviations of the distributions in Experiment 1 and Experiment 2 were 50.5 and 13.4, and 50.5 and 20.2, with associated ranges of 21 to 80 and 3 to 98, respectively. In Experiment 3 and 4, the numbers of asterisks were uniformly distributed. Experiment 3 used asterisk values drawn from a uniform distribution ranging from 31 to 70. Experiment 4 used asterisk values drawn from a uniform distribution ranging from 3 to 98. Fig. 3 shows the distributions of asterisk values for each experiment with a criterion marking the cutoff for "large" and "small" values (e.g., for each experiment "large" values were randomly sampled from the region of the distribution to the right of the criterion and "small" values were randomly sampled from the region of the distribution to the left of the criterion).

4.1.1.3. Procedure. Subjects were instructed to decide whether each array contained a large or small number of asterisks (with "large" defined as more than 50 and "small" defined as less than or equal to 50) and were encouraged to make this decision as quickly and accurately as possible. Subjects were also informed that different blocks of trials would contain different numbers of "large" and "small" stimuli and that they would be informed about the nature of each upcoming block (i.e., whether it would have an equal number of "large" and "small" stimuli or more of one type than the other).

The entire experiment took approximately 50 min and consisted of a short block of practice trials followed by 45 blocks of 40 trials each. On each trial, an array of asterisks was displayed on the PC screen and remained on screen until subjects had made a decision. Responses were made using the keyboard with the '/' key used to indicate a "large" number of asterisks and the '*z*' key used to indicate a "small" number of asterisks. If a response was incorrect, the word 'error' was displayed on the screen for 500 ms. If a response was made too quickly (faster than 280 ms), then the message 'TOO FAST' was displayed on the screen for 500 ms. If a response was made too slowly (slower than 2000 ms), then the message 'TOO SLOW' was displayed on the screen for 500 ms.

There were 15 blocks with an equal number of "large" and "small" trials (20 of each), 15 blocks with three times more large trials (30 large and 10 small), and 15 blocks with three times more small trials (30 small and 10 large). Block type was varied randomly after every third block and subjects were informed before each change as to the nature of the upcoming sets of trials (e.g., "For the next set of lists, there will be three times more large trials than small trials"). Each time the block type changed, subjects were required to press a specific key to indicate that they understood which type of stimuli would be more common in the upcoming block (e.g., they had to press the "*L*" key to begin a set of blocks that contained more large trials). Inspection of the data found no evidence of differences in accuracy or response times to large and small stimuli, so the corresponding large and small asterisk values were combined to reduce the number of conditions (e.g., very large and very small asterisk values were combined into a single low-difficulty category with responses recoded as either 'correct' or 'incorrect' rather than 'large' or 'small'). This resulted

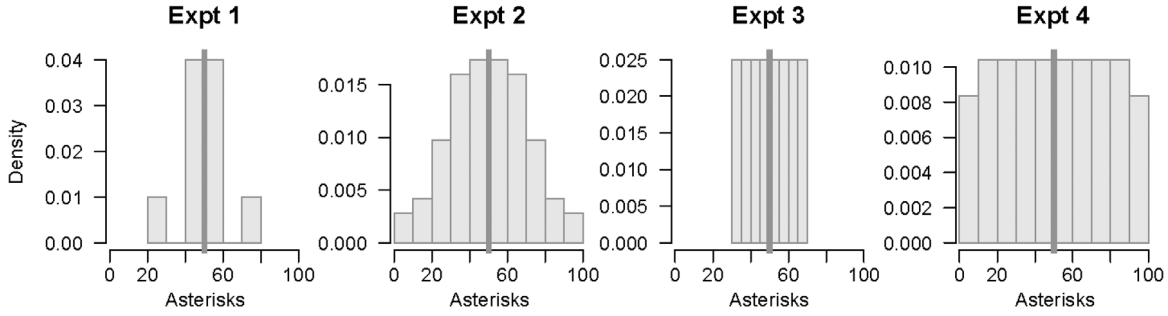

Fig. 3. Stimulus strength distributions for Experiments 1–4.

in three bias conditions: no bias (all responses from blocks with equal numbers of large and small stimuli), bias toward the given response ("large" responses from blocks with more large trials and "small" responses from blocks with more small trials), and bias against the given response ("large" responses from blocks with more small trials and "small" responses from blocks with more large trials). To further reduce the number of conditions, asterisk values in each experiment were grouped resulting in 3 difficulty levels for Experiment 1 and four difficulty levels for Experiments 2–4.

4.1.2. Model fitting

When fitting the models to these data, drift rates were allowed to change with stimulus strength (different numbers of asterisks) and the starting point of the accumulation process was allowed to change with the bias manipulation. In the 'no bias' condition, the starting point was fixed to zero (i.e., equidistant between the 'correct' and 'error' response boundaries). In the bias conditions, the starting point was free to vary with the constraint that the absolute amount of bias in the 'bias against' and 'bias toward' conditions was identical with just the sign changing across conditions. No assumptions were made about the precise relationship between drift rates and difficulty (e.g., no assumed linear relationship between drift rate and number of asterisks, Ratcliff, 2014). This parameterization of the bias manipulation has previously been shown to provide a good account of data from experiments using this type of bias task (Leite & Ratcliff, 2011; Ratcliff, 1985), although some researchers have allowed the drift criterion to change as a function of bias (cf. Hanks et al., 2011) as well as the starting point (Ratcliff, 1985; Ratcliff et al., 1999; van Ravenzwaaij et al., 2012). The drift criterion allows the rates of evidence accumulation for the two stimulus alternatives to be unequal. We also fit versions of the fixed and collapsing boundary models with changes in drift criteria as a function of bias, but found no improvement in GOF from these additional parameters. For the collapsing boundary model, κ was constrained to be between 0 and 1 and $t_{0.5}$ was constrained to be between 0 and 2 (s).

Predictions for both models were obtained using the numerical integral equation methods described by Smith (2000) and used by Smith and Ratcliff (2009). The details of this method are given in Appendix B. We used the G^2 statistic for binned data as our GOF measure. This statistic can be written

$$G^2 = 2\sum_i N_i \sum_j p_{ij} \ln\left(\frac{p_{ij}}{\pi_{ij}}\right)$$

where p_{ij} is the proportion of observations in the jth bin of condition i, π_{ij} is the proportion in the bin predicted by the model, and N_i is the number of observations in condition i. When multinomial sampling assumptions are satisfied, this statistic is equal to twice the difference between the maximum possible log likelihood and the log likelihood predicted by the model. The G^2 and χ^2 statistics approach one another for large sample sizes (Jeffreys, 1961), and, for any given set of data, the G^2 and BIC statistics differ by a constant such that both are minimized by the same set of parameters.

4.1.3. Results

The data were the response proportions and 0.1, 0.3, 0.5, 0.7, and 0.9 response time distribution quantiles for correct and error responses for each difficulty level and bias condition. Both collapsing and fixed boundary versions of the model were fit to the data. All model fits were performed on both individual subject data and averaged data, but we present only the fits to averages data for simplicity. For conditions in which subjects made fewer than 10 error responses the models were fit to just the RT data from the correct responses. For these conditions, p_{ij} and π_{ij} were conditional on a correct response.

Overall, both models were able to provide a qualitatively good fit to the data. The best-fitting parameters for the average data for each model and each experiment are shown in Table 1. The average data and model predictions for each experiment are shown in Fig. 4. The figure shows that the differences in the predictions of the two models are very small. The collapsing boundary model produced RT distributions with smaller 0.9 quantiles, but this difference was very slight (about a 20 ms difference on average across subjects and conditions). G^2 and BIC values for the average data from each experiment and model are shown in Table 4 (note that these are the G^2 and BIC values from a fit of the average data, not the averages of the G^2 and BIC values from the individual fits). For calculating both G^2 and BIC values for the average data, the average number of observations per condition was used (see PBCM section for explanation of BIC_A). The collapsing boundary model produced smaller G^2 values but larger BIC values relative to the fixed boundary model for all of the experiments, as will be discussed in the PBCM section.

The typical bias effects were observed in the data. The column on the left side of Fig. 4 shows the data from blocks with equal numbers of "large" and "small" trials, the middle column shows the data from trials where subjects were biased against a particular response (i.e., "large" responses when there were more "small" stimuli, and "small" responses when there were more "large" stimuli), and the column on the right shows the data from trials where subjects were biased in favor of a particular response (i.e., "large" responses when there were more "large" stimuli and "small" responses when there were more "small" stimuli). For all of these plots, data have been collapsed across 'large' and 'small' numbers of asterisks (i.e., "large" responses to "large" stimuli were grouped with "small" responses to "small" stimuli, and "small" responses to "large" stimuli were grouped with "large" responses to "small" stimuli). Relative to the no-bias condition, when subjects were biased against making a particular response they made that response less often and more slowly (i.e., when there were more "large" stimuli, subjects made fewer "small" responses and made those responses more slowly). Similarly, relative to the no-bias condition, when subjects were biased in favor of making a

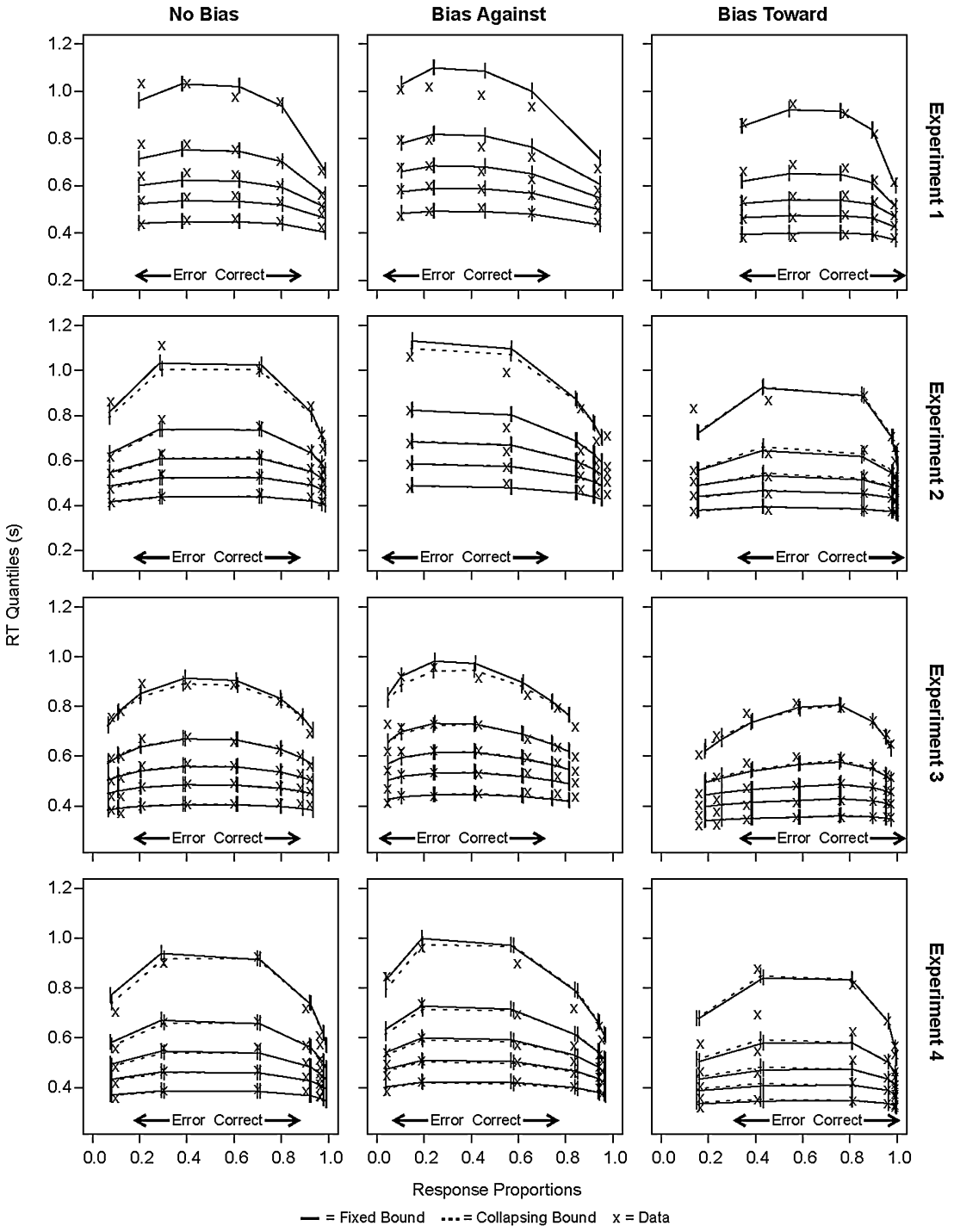

Fig. 4. Average data and model predictions for Experiments 1–4. Response proportions are plotted along the *x*-axis and RT quantiles (0.1, 0.3, 0.5, 0.7, and 0.9) are plotted vertically. Conditions where subjects made an average of fewer than 9 responses are omitted from the figure. The column on the left plots responses from blocks with an equal number of "large" and "small" stimuli. The middle column plots responses when subjects were biased against a particular response (i.e., "large" responses from blocks with more "small" trials and "small" responses from blocks with more "large" trials). The column on the right plots responses when subjects were biased in favor of a particular response (i.e., "large" responses from blocks with more "large" trials and "small" responses from blocks with more "small" trials). The x's are the averaged data from the subjects, and the solid and dashed lines are the predictions from the fixed and collapsing bound models, respectively.

particular response they made that response more often and more quickly (i.e., when there were more "large" stimuli, subjects made more "large" responses and made those responses more quickly). These results were well captured by both models with just a change in starting point. Relative to the no-bias condition, there were shifts in the leading edges of the RT distributions in the two bias conditions. This result is consistent with modeling bias as a change in starting point as opposed to a change in drift criterion.

There were consistent differences between the best-fitting parameters produced by each model. Unsurprisingly, the collapsing boundary models had larger initial boundary separation parameters along with some amount of boundary collapse (governed by the κ and $t_{0.5}$ parameters). However, if we plot the boundary functions associated with these parameters we can see that the boundaries produced by the collapsing bound model closely match the best-fitting fixed boundary values for the time period in which

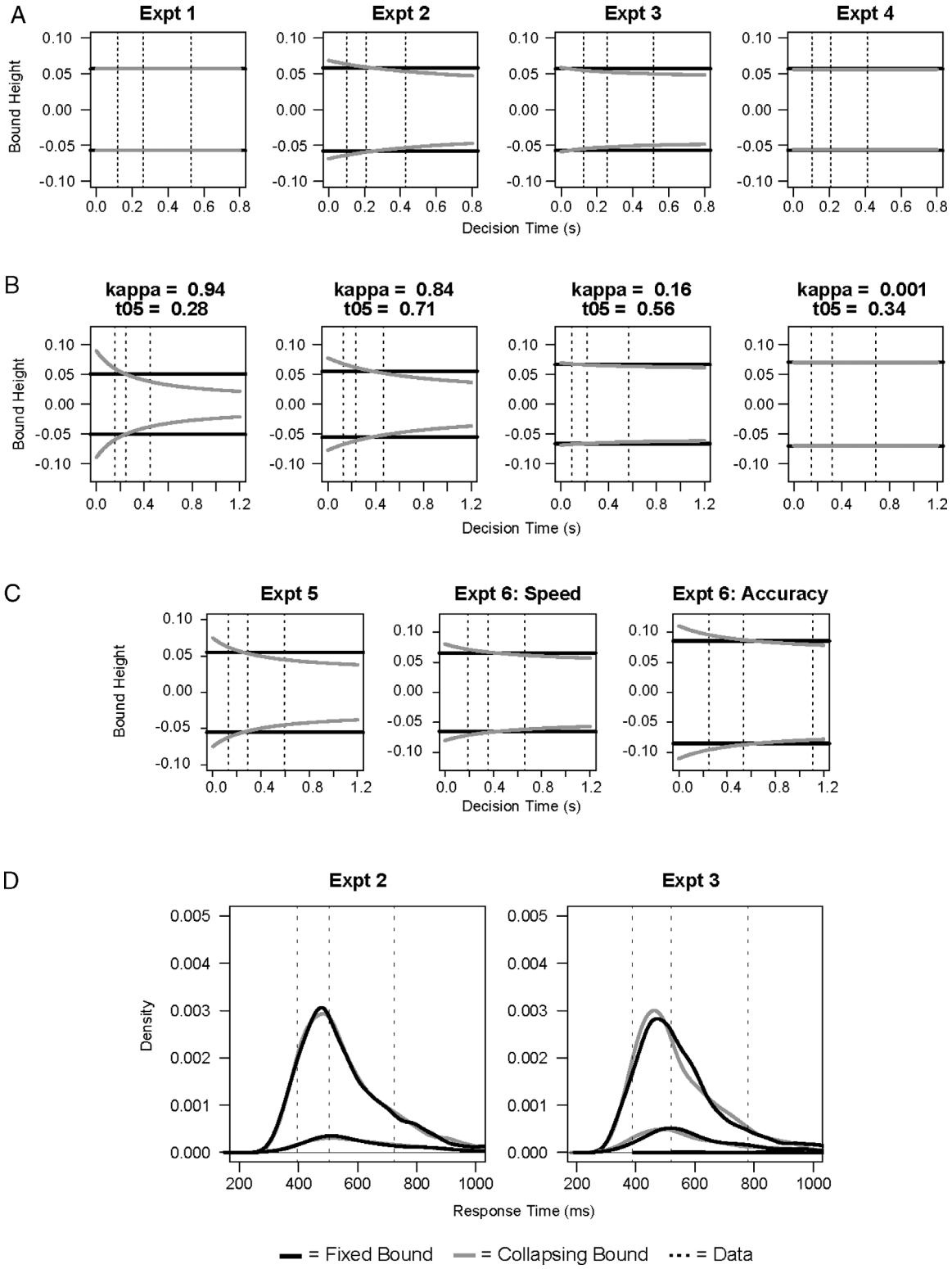

Fig. 5. Boundary values and RT quantiles for all of the experiments and select individuals. Vertical lines depict the average 0.1, 0.5, and 0.9 RT quantiles for each experiment minus the average non-decision time. A: Average boundary values for each model for Experiments 1–4. B: Boundary values for select subjects from Experiments 1–4. C: Average boundary values for each model for Experiment 5 and both conditions of Experiment 6. D: Model-predicted response time distributions from average parameter values for Experiments 2 and 3. Predictions from the fixed bound model are shown in black and predictions from the collapsing bound model are shown in gray. The smaller distributions are the error responses. Vertical lines depict the 0.1, 0.5, and 0.9 RT quantiles from the data.

subjects are making most of their responses. The collapsing boundaries based on the fit of the average data for each experiment are shown in Fig. 5(a) along with the fixed boundary values. The vertical dotted lines represent the 0.1, 0.5, and 0.9 quantile decision times (the response times minus the non-decision component) for each experiment averaged across subjects and conditions: The majority (80%) of the subjects' responses were made in those time periods. Predicted response time distributions (both correct and error responses) from the average model parameters from Experiments 2 and 3 are shown in Fig. 5(d) (note that these are joint

Table 1
Parameter values for fixed (F) and collapsing (C) boundary models for average data from Experiments 1–4.

Expt	v_1	v_2	v_3	v_4	η	T_{er}	s_t
1F	0.38	0.13	0.05	–	0.04	0.30	0.18
1C	0.38	0.13	0.05	–	0.04	0.30	0.18
2F	0.40	0.33	0.24	0.09	0.05	0.30	0.17
2C	0.40	0.32	0.23	0.08	0.01	0.30	0.15
3F	0.27	0.22	0.14	0.05	0.06	0.27	0.18
3C	0.25	0.20	0.13	0.04	0.00	0.26	0.16
4F	0.42	0.35	0.23	0.08	0.05	0.25	0.14
4C	0.40	0.33	0.22	0.08	0.003	0.25	0.13

	a	z	s_z	κ	
1F	0.11	0.02	0.003	–	–
1C	0.11	0.02	0.003	0.0006	0.43
2F	0.12	0.02	0.03	–	–
2C	0.14	0.02	0.05	0.59	0.70
3F	0.10	0.02	0.02	–	–
3C	0.12	0.02	0.01	0.25	0.32
4F	0.11	0.02	0.001	–	–
4C	0.11	0.02	0.002	0.002	0.35

T_{er} is the mean nondecision time, s_t is the range of the between-trial variability in nondecision time, a is the boundary separation, s_z is the range of the between-trial variability of the starting point, z is the starting point for the bias conditions (i.e., the starting point in the 'no bias' condition is fixed to zero, and in the two bias conditions it is either positive or negative z), $v_1 - v_4$ are the mean drift rates, η is the between-trial variability in drift rate, κ is the percentage of boundary collapse, and $t_{0.5}$ is the rate of boundary collapse.

distributions; the probability masses in the correct and error distributions sum to one). The predictions of the collapsing boundary model are plotted in gray, the predictions of the fixed boundary model are plotted in black, and the 0.1, 0.5, and 0.9 quantile response times, averaged across subjects and conditions, are plotted as vertical dashed lines. The predicted distributions from the two models are highly similar, even in the tails. The collapsing and fixed boundary functions are also very similar for individual subjects. Boundary values from each of the two models for four extreme subjects from across the four experiments are shown in Fig. 5(b). These subjects were chosen based on their boundary collapse parameters. The best fitting κ values for the first two subjects were quite large (indicating a large decrease in boundary over time) and the values for the second two subjects were quite small (indicating a very small decrease in the boundary over time). The vertical dotted lines depict the 0.1, 0.5, and 0.9 average quantile decision times for each subject. While there is a larger difference between the fixed and collapsing boundaries for subjects with larger values of κ, the difference in the boundaries is still relatively small in the time region in which subjects are making most of their responses. In general, both versions of the diffusion model are able to provide good fits to the behavioral data and the predicted distributions from the collapsing boundary model do not qualitatively differ from the predicted distributions from the standard fixed boundary model.

4.2. Motion discrimination: Experiments 5 and 6

Experiments 5 and 6 used a motion discrimination task in which a cloud of randomly moving dots was displayed on the screen, some proportion of which moved in the same direction. The average proportion of dots that move in the same direction is called the *coherence*. Subjects had to judge whether the coherently moving dots were moving to the right or to the left. Data from Experiments 5 and 6 were previously reported in Ratcliff and McKoon (2008), as Experiments 1 and 2, respectively); the data from Experiment 5 were also previously reported in Ratcliff and Smith (2010) and used in model fitting by Hawkins et al. (2015).

4.2.1. Method

On each trial, a sequence of frames was displayed on a PC screen at a rate of 16.7 ms per frame. On each frame, five single-pixel dots were displayed in a circular aperture 5.4° in diameter centered on the screen. On the first three frames, the dots were located in random positions. On the fourth and each subsequent frame, a proportion of the dots moved coherently either left or right from their position three frames ago. On each of the frames, the dots that were not chosen to move coherently appeared in random locations.

Coherence was defined as the proportion of dots that moved in the same direction across frames. For example, if the direction of coherent motion was left and the probability was 0.05, then the probability that a dot in each frame would move left would be 0.05. Subjects were asked to respond as quickly and accurately as possible, pressing the forward slash key if the coherent motion was toward the right and the *Z* key if the motion was toward the left. Subjects were given error feedback throughout the task. For both experiments, response times and accuracy to leftward and rightward motion stimuli were fairly similar, so the data were collapsed across direction of motion.

Experiment 5 had six levels of coherence (0.05, 0.10, 0.15, 0.25, 0.35, or 0.50). When fitting the models to these data, only drift rates were allowed to change across conditions. Experiment 6 had four levels of coherence (0.05, 0.10, 0.15, or 0.35) and a speed/accuracy manipulation, in which subjects were encouraged to respond more quickly on half of the blocks of trials and more accurately on the other half. When fitting the models to these data, the drift rates were allowed to change across the different coherence levels and the boundary separation (a) was allowed to change across the speed/accuracy levels. Some researchers have also found changes in non-decision time and drift rate across speed/accuracy levels (Rae, Heathcote, Donkin, Averell, & Brown, 2014; Ratcliff, 2006; Starns et al., 2012), but these changes tend to be found only with extreme speed stress and large changes in accuracy across conditions. Rae et al. found that both the diffusion model and the LBA underestimated the difference in error rates between speed and accuracy conditions when boundary separation was the only parameter that varied between them. In contrast, we found only small differences in error rates between these two conditions. These were reasonably well fit with only a change in boundary separation by both the fixed and collapsing boundary models.

4.2.2. Results

The data were the response proportions and the 0.1, 0.3, 0.5, 0.7, and 0.9 response time distribution quantiles for correct and error responses for each experimental condition. Model fits were again performed using Smith's (2000) method for generating exact predictions and both collapsing and fixed boundary versions of the model were fit to the data. All model fits were performed on both individual and averaged data, but just the fits of the average data are presented for simplicity. Since these data have all been modeled previously, our analysis will focus primarily on the results as they pertain to the question of fixed or collapsing boundaries.

Overall, both models were able to provide a qualitatively good fit to the data sets. The best-fitting parameters for the average data for each model and each experiment are shown in Tables 2 and 3, the average data and model predictions for each experiment are shown in Fig. 6, and G^2 and BIC values for the average data from each experiment and model are shown in Table 4. Again, the differences between the predictions of the two models were quite small. The collapsing boundary model produced RT distributions with smaller 0.9 quantiles, but this difference was again slight (about a 10 ms difference in Experiment 5 and about a 20 ms difference in Experiment 6 on average). The fixed and collapsing boundaries for these two experiments are shown in Fig. 5(c). As before, the best-fitting collapsing and fixed boundaries were quite similar to each other over the range of times in which subjects made most of their responses.

Table 2
Parameter values for fixed and collapsing boundary models for average data from Experiment 5.

Expt	v_1	v_2	v_3	v_4	v_5	v_6	η
5F	0.04	0.08	0.12	0.20	0.25	0.33	0.10
5C	0.04	0.07	0.11	0.19	0.24	0.32	0.09

	a	z	s_z	T_{er}	s_t	κ	$t_{0.5}$
5F	0.11	0.00	0.005	0.30	0.19	–	–
5C	0.15	0.00	0.07	0.30	0.17	0.64	0.36

T_{er} is the mean nondecision time, s_t is the range of the between-trial variability in nondecision time, a is the boundary separation, s_z is the range of the between-trial variability of the starting point, z is the starting point, v_1–v_6 are the mean drift rates, η is the between-trial variability in drift rate, κ is the percentage of boundary collapse, and $t_{0.5}$ is the rate of boundary collapse.

Table 3
Parameter values for fixed and collapsing boundary models for average data from Experiment 6.

Expt	v_1	v_2	v_3	v_4	η	T_{er}	s_t
6F	0.04	0.08	0.10	0.20	0.07	0.29	0.19
6C	0.05	0.09	0.12	0.22	0.10	0.24	0.18

	a_1	a_2	z	s_z	κ	$t_{0.5}$
6F	0.11	0.15	0.00	0.002	–	–
6C	0.15	0.21	−0.01	0.001	0.52	0.61

T_{er} is the mean nondecision time, s_t is the range of the between-trial variability in nondecision time, a_1–a_2 are the two boundary separation values, s_z is the range of the between-trial variability of the starting point, z is the starting point, v_1–v_4 are the mean drift rates, η is the between-trial variability in drift rate, κ is the percentage of boundary collapse, and $t_{0.5}$ is the rate of boundary collapse.

5. PBCM results for the experiments

While both versions of the diffusion model were able to account for the data in all of these experiments, our goal is determine whether the addition of collapsing decision boundaries represents a substantial improvement over the standard fixed boundary diffusion model. The parametric bootstrap cross-fitting method (PBCM) was used to determine which model provided a better account for each set of data. For each experiment, the best-fitting parameter values from each model from fits of the average data were used to generate 40 sets of simulated data each containing the same number of observations as in the behavioral data. This is a relatively small number of simulated data sets, but our initial bootstraps with only 20 sets of data produced quite similar results. Both the fixed and collapsing boundary models were fit to each of these simulated data sets, resulting in four sets of GOF values for each experiment (fits of the collapsing and fixed boundary models to collapsing and fixed boundary simulated data). We then took the difference in GOF for the fixed and collapsing boundary model fits for each simulated data set; the resulting distributions for each experiment are plotted in Fig. 7. The criterion for each pair of GOF distributions is determined based on classification accuracy. That is, the criterion for each pair of distributions is the point where a maximum number of the simulated data sets were correctly classified in terms of which model they had been generated from. The observed GOF for the fixed boundary model minus the observed GOF for the collapsing boundary model for each experiment is plotted as a triangle. If this observed GOF is less than the criterion, then the fixed boundary model is preferred for that data set and vice versa.

The GOF values for the average data, the PBCM criteria, and the results are presented in Table 4. Based on the results of the PBCM analyses, the fixed boundary model was preferred for four of the six experiments and the collapsing boundary model was preferred for the average data from Experiments 2 and 4 (although for Experiment 4 the estimated amount of collapse was negligible). Note that for some of the experiments, namely 1 and 4, the two PBCM distributions are almost completely overlapping. This occurs when the two models are nearly indistinguishable from each other in terms of their predictions. For these two experiments, the best-fitting parameters for the collapsing boundary model produce a boundary with a negligible amount of collapse (e.g., less than 1%). Thus the data-informed PBCM approach is attempting to distinguish between a model with a fixed boundary and a model with a boundary that collapses only a miniscule amount. Unsurprisingly then, these two models are nearly indistinguishable and the PBCM criterion has low classification accuracy in these instances. This is a general shortcoming of the data-informed PBCM when comparing nested models. However, in such instances the inference problem is straightforward because the two models lead to the same conclusion, namely, that the amount of boundary collapse is negligible.

Table 4
GOF values and PBCM results for each experiment.

Expt	Fix: G^2, BIC, G^2_A, BIC_A	Col: G^2, BIC, G^2_A, BIC_A	Fix-Col: G^2	PBCM criterion	PBCM result
1	39.4, 106.7, 253.3, 337.4	39.3, 121.6, 252.7, 355.5	0.1	1.0	Fixed*
2	39.5, 114.4, 211.3, 303.0	36.8, 126.7, 196.9, 306.9	2.7	2.6	Collapsing
3	34.2, 108.9, 343.7, 441.5	34.0, 123.7, 341.7, 459.0	0.2	1.9	Fixed
4	42.8, 117.7, 244.8, 337.2	41.3, 131.1, 236.2, 347.0	1.5	0.8	Collapsing*
5	17.6, 100.6, 46.8, 141.6	14.8, 111.6, 39.4, 149.9	2.8	3.1	Fixed
6	32.4, 107.5, 273.5, 372.0	32.3, 121.1, 272.6, 389.1	0.1	1.5	Fixed

G^2 and BIC values for the average parameter values and average data for both models for each experiment are shown in the first two columns. The first two values were calculated using the average number of observations and the second two values were calculated using the average number of observations multiplied by a dispersion factor. To determine whether a given data set is better characterized as coming from a fixed or a collapsing boundary model, the difference in GOF (third column) is compared to the PBCM criterion (fourth column) for that data set. If the difference is less than the criterion, then the fixed boundary model is preferred. If the difference is greater than the criterion, then the collapsing boundary model is preferred.

* Indicate experiments where the GOF distributions were largely overlapping such that the PBCM criterion has low classification accuracy.

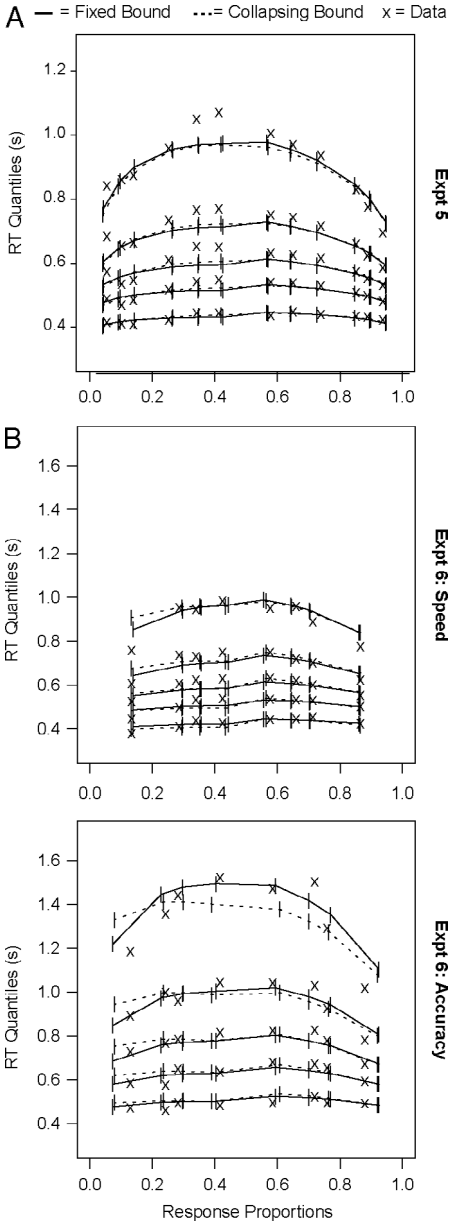

Fig. 6. Average data and model predictions for Experiments 5 (A) and 6 (B). Response proportions are plotted along the *x*-axis and RT quantiles (0.1, 0.3, 0.5, 0.7, and 0.9) are plotted vertically. Conditions where subjects made an average of fewer than 9 responses are omitted from the figure. The *x*'s are the averaged data from the subjects, and the solid and dashed lines are the predictions from the fixed and collapsing bound models, respectively.

It is important to note that the collapsing boundary model always produced smaller G^2 values than the fixed bound model, even when the models were fit to data generated from a fixed boundary model. This was expected, as the collapsing boundary model contained the fixed boundary model as a special case. However, the fixed boundary model always produced smaller BIC values than the collapsing boundary model. BIC values can be calculated from G^2 values by adding a penalty term based on the number of free parameters in the model (M) and the number of observations in the data (N):

$$\mathrm{BIC} = G^2 + M\ln(N).$$

The PBCM analyses can be performed with either G^2 or BIC values and identical results will be obtained. When converting the G^2 values to BIC values, the values for each model will change by a constant (given by the equation above) such that the difference between the GOF values from the two models will also change by a constant.[3] The PBCM is thus agnostic about differences between the models in terms of number of parameters or observations (at least in methods such as AIC or BIC where these aspects of the model are represented as simple penalty terms) and is only sensitive to model complexity in terms of the ability of each model to fit various patterns of data. If we had used the standard procedure of simply choosing the model with a smaller BIC value (i.e., assuming a criterion of zero for distributions of GOF differences), then the fixed boundary model would have been preferred for all of these empirical data sets. Moreover, the standard procedure of choosing the model with the smaller BIC value would also have misclassified the great majority of the simulated data sets. For the number of observations in these experiments, the PBCM distributions for BIC values would be shifted to the left by about 14 units compared to the distributions for G^2 values such that the majority of all of the BIC difference distributions would be below zero. Rather than simply choosing the model with the best GOF (whether penalized or not), the PBCM approach reinterprets an observed difference in GOF between two models in light of the ability of each model to account for data generated from the other model.

Our use of the PBCM method circumvents a problem that would otherwise arise when performing inference using penalized model statistics like the BIC on quantile averaged data. The sample size, N, that we used to calculate G^2 and the associated BIC was the number of trials per subject. This choice is consistent with our interpretation of the quantile averaged data as characterizing the performance of an "average subject", but it can lead an inappropriate choice of a penalty term in statistics like the AIC and BIC, because the use of these statistics for binned data presupposes multinomial likelihood functions. The expected value of G^2 for a well-fitting model then equals its residual degrees of freedom (data degrees of freedom minus the number of estimated parameters). Quantile averaged data are typically underdispersed relative to the multinomial distribution, leading to differences in G^2 between competing models that are smaller than would be obtained with multinomial sampling and, consequently, penalty terms that are too large and that overly penalize more complex models.

There appears to be no entirely satisfactory solution to this problem. One approach, described by Burnham and Anderson (2002), is to use the fit of a chi-square type statistic to estimate a dispersion factor and then to correct the GOF statistic by this amount prior to applying a penalty. This approach is open to the objection that it presupposes that the set of candidate models contains a "true" model (in which the difference between the model and data is due to sampling variability alone and thus can provide an unbiased estimate of the dispersion factor), and also that the same data are used to estimate the dispersion factor and to compare the fit of the competing models.

Instead of using the same data to estimate dispersion and model fit, another approach is to use simulated data to estimate the underdispersion associated with quantile averaging and then to use the estimated dispersion factor to correct the GOF statistic before applying the penalty. We performed this analysis for each of the experiments and found no change in the results. For each experiment we used the mean parameter values from the collapsing boundary model to simulate N data sets with n trials each (where N is equal to the number of subjects and n is equal

[3]
$$\begin{aligned}\mathrm{BIC}_F - \mathrm{BIC}_C &= (G_F^2 + M_F\ln(N)) - (G_C^2 + M_C\ln(N))\\ &= (G_F^2 - G_C^2) + \ln(N)(M_F - M_C).\end{aligned}$$

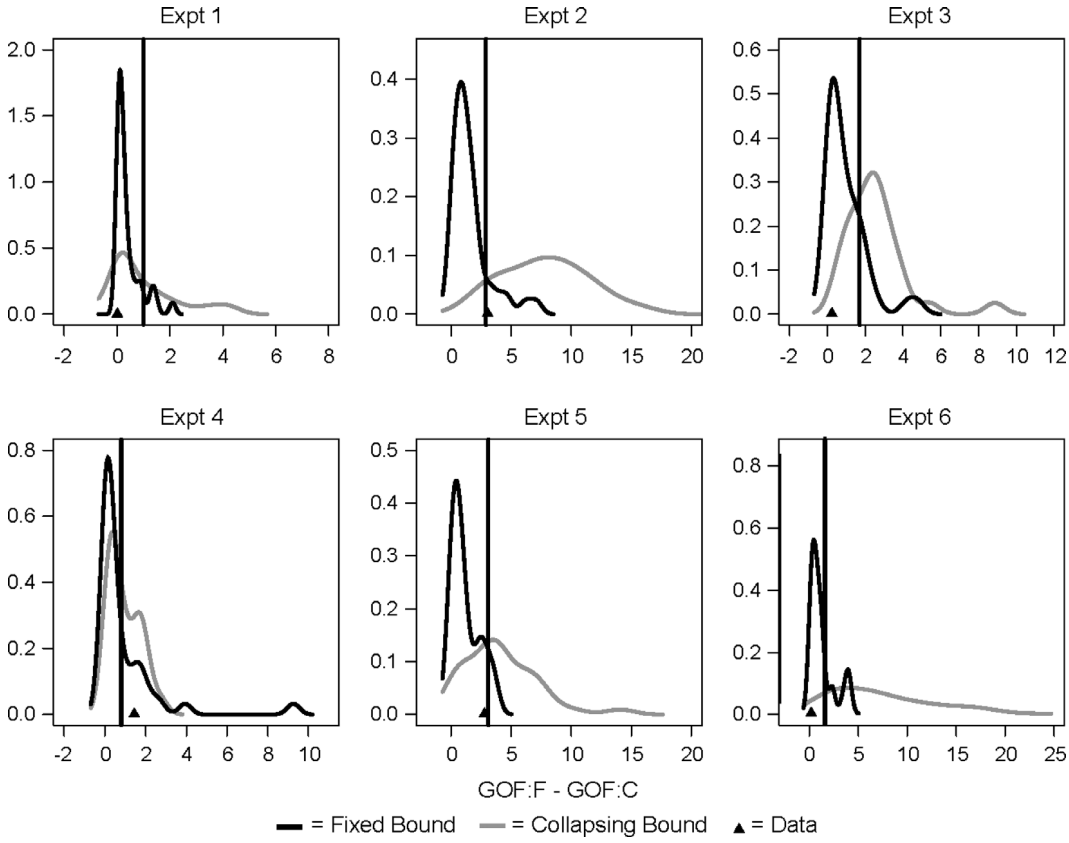

Fig. 7. Distributions of the differences in GOF for each experiment based on the parametric bootstrap cross-fitting method. The black distribution in each plot is based on data generated from the fixed bound model and the gray distribution in each plot is based on data generated from the collapsing bound model. The criteria for the distributions are marked with vertical lines and the observed GOF difference for each experiment is marked with a triangle.

to the mean number of trials in that experiment). The response proportions and RT quantiles from these simulated data sets were then averaged and a G^2 value was calculated for this averaged data set using the generating parameters. This process was repeated 100 times for each experiment to give a distribution of G^2 values and the mean of this distribution was used to calculate a dispersion factor that was used to adjust the number of observations used to calculate both the G^2 value and the BIC penalty term. That is, the number of observations was multiplied by a dispersion factor equal to the degrees of freedom in the model fit (i.e., the degrees of freedom in the data minus the number of parameters in the model) divided by the mean of the simulated G^2 distribution. These adjusted G^2 and BIC values are shown in Table 4 and show the same pattern of results as the un-adjusted GOF values.

One of the advantages of the PBCM method is that it avoids the problem of having to determine the correct penalty term for underdispersed or overdispersed data. The PBCM method identifies a decision criterion that maximizes the separation between competing models for whatever statistic is used to characterize model fit. Changing (or omitting) the penalty term shifts the distributions of differences in fit statistics along the GOF axis and shifts the decision criterion by the same amount. As a result, whether a particular data set is judged as more likely to have been produced by one or other model will be unaffected by whether the penalty term was too large, too small, or was omitted entirely.

It is also important to note that these analyses are data-driven; the numerical GOF criterion for each experiment should only be applied to the specific data set being analyzed and not be taken as a criterion for choosing between the two models for other data sets. To choose between these two models for a new data set, the PBCM analysis would need to be repeated using parameter values and samples sizes appropriate for those data. The goal of our data-driven approach is to determine which of two specific forms of the model provides a better fit to a specific set of observed data. However, a closer examination of the results of the cross-fitting of the models can yield valuable insights about the general performance of the two models and the amount of mimicry between them, as we discuss in the following section.

6. Model selection considerations: mimicry and recovery

The collapsing and fixed boundary versions of the model produced very similar results for all of the experiments (see Figs. 4 and 6). This was not particularly surprising given that we are comparing instances of the models that were chosen based on their fit to the observed data. Nevertheless, we were still surprised to not see greater differences between the models in the shapes of the predicted response time distributions. When we plot the predicted quantiles of the fixed boundary model against the predicted quantiles of the collapsing boundary model, the plots are approximately linear. Fig. 8 plots all of the predicted quantile values from each model's fit to the subjects from Experiment 5 along with a regression line. The fact that the relationship between the models' quantiles is approximately linear indicates that the models are producing similarly shaped RT distributions, at least in this region of the parameter space. In other words, the collapsing boundary model is not producing RT distributions that are more symmetric than the distributions produced by the fixed boundary model. This is because, as shown in Fig. 5, there is little difference between the two boundaries across the main part of the RT distribution.

Since the PBCM involves fitting models to data generated from known parameter values, we can investigate how well

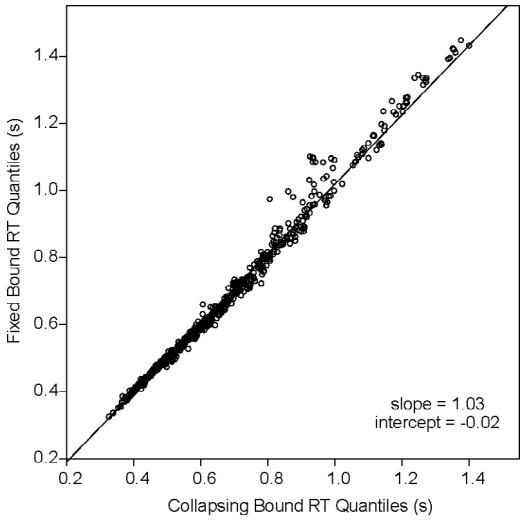

Fig. 8. Predicted RT quantiles from each model plotted against each other along with the best-fitting linear regression line.

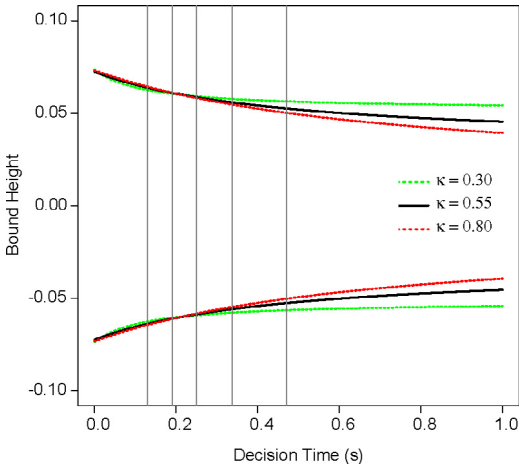

Fig. 10. Best-fitting boundaries with kappa parameter fixed to extreme values. The solid lines depict the boundaries used to generate the data ($\kappa = 0.55$, $t_{0.5} = 0.48$, $a = 0.15$), the green dashed lines depict the best fitting boundaries when κ was fixed to a smaller value ($\kappa = 0.30$, $t_{0.5} = 0.13$, $a = 0.15$), and the red dashed lines depict the best fitting boundaries when kappa was fixed to a larger value ($\kappa = 0.80$, $t_{0.5} = 0.72$, $a = 0.15$). The gray vertical lines depict the 0.1, 0.3, 0.5, 0.7, and 0.9 RT quantiles minus the non-decision time.

these original parameters are recovered. Ideally, our model fitting routine should recover parameter values that are similar to the original parameter values. Fig. 9 shows histograms of the recovered values of κ and $t_{0.5}$ (the boundary collapse parameters) for Experiment 5. The black lines indicate the true values of these parameters (the values used to generate simulated data). On average, the recovered parameter estimates overestimate the value of κ and underestimate the value of $t_{0.5}$. In other words, the recovered model parameters correspond to a boundary that collapses to a smaller percentage of the starting point and does so more quickly than the boundary that was used to generate the data. To further illustrate this effect, we generated data from a model with collapsing boundaries with a moderate amount of collapse ($\kappa = 0.55$) and then fit this data with varying levels of parameter constraint. We constrained κ to be either 0.25 greater or less than the actual value (fixed to 0.3 or 0.8) and then allowed only the other boundary parameters (a and $t_{0.5}$) to vary when fitting the model. All other parameters were fixed to their 'true' values (i.e., the values used to generate the data, which were the mean parameter values from Experiment 5). The resulting best-fitting boundaries are shown in Fig. 10, along with the original boundary and the RT quantiles. Within the time period in which decision processes reach the boundaries, there is very little difference between the three sets of boundaries.

The degree to which these boundary parameters are able to trade-off with each other to produce nearly identical boundaries despite having very different parameter values indicates that there is an identifiability issue with the collapsing boundary model as well as a parameter recovery issue. If very different values for these parameters are able to produce nearly identical boundaries, then inevitably model-fitting routines will have difficulty recovering the 'true' values of these parameters. If model fits are unable to accurately recover parameter values that characterize the boundary collapse then it is difficult to draw conclusions based on these parameter values. If these boundary parameters are not recovered reliably from data, then they cannot be meaningfully compared across subjects or linked to experimental manipulations or individual differences. Furthermore, we found that the estimated boundary values correlated significantly with the estimates of other model parameters, such as drift rate, non-decision time, and the initial boundary separation (see Appendix for more details).

Previous research has demonstrated that the parameters of the fixed-boundary diffusion model have psychological validity and meaningful interpretations: Manipulations of task difficulty primarily affect drift rates (Ratcliff, 2014; Ratcliff & McKoon, 2008), manipulations of response speed primarily affect boundary separation (Ratcliff & McKoon, 2008), and manipulations of response modality primarily affect non-decision time (Gomez, Ratcliff, & Childers, 2015). The presence of correlations between these psychologically meaningful parameters and boundary collapse parameters (which are not thought to be related to individual

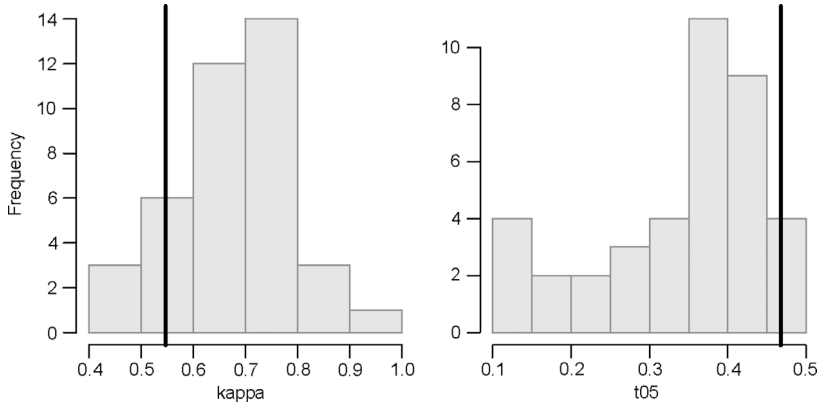

Fig. 9. Collapsing bound parameter recovery. Original generating values are plotted in black and recovered values are plotted in gray.

differences), means it is more difficult to provide unambiguous interpretations of the estimated parameters and how they relate to individual differences and task manipulations. Consequently, the addition of collapsing boundaries to the standard diffusion model makes the parameter estimates less reliable and makes it more difficult to relate the parameters to underlying psychological processes.

7. Between-trial variability parameters

Unlike other collapsing boundary models (Ditterich, 2006b; Hanks et al., 2011), the models we fit included between-trial variability in non-decision time, drift rate, and starting point. In the fixed boundary version of the diffusion model, these parameters are necessary to produce fast and slow errors (Ratcliff, 1981; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Ratcliff et al., 1999; Wagenmakers et al., 2008). However, proponents of the collapsing boundary diffusion model have claimed that the model is capable of producing slow errors without these between-trial sources of variability (Ditterich, 2006b; Palmer et al., 2005; Shadlen & Kiani, 2013). We demonstrate that the collapsing boundary model without between-trial variability is unable to produce either the pattern of slow error RTs observed in some of our data sets or the slow errors produced by a fixed boundary model with between-trial variability parameters.

We generated simulated data from a fixed boundary model with non-zero between-trial variability parameters using a random-walk approximation of the diffusion model, parameter values based on the average parameter values from Experiment 5 plus small amounts of variability, and 300 observations per condition. These simulated data sets were then fit with both fixed and collapsing boundary versions of the diffusion model with between-trial variability parameters fixed to zero. When fitting these data sets, neither a fixed-boundary model nor a collapsing-boundary model was able to account for the error RT distributions. Representative fits of the simulated data with slow error response times are presented in Fig. 11(A). While the model with collapsing boundaries was able to produce slightly slower error RTs than the model with fixed boundaries, neither model was able to capture the full error RT distribution. Specifically, both models produced error RT distributions that were more symmetric than the simulated data, leading them to under-predict the 0.7 and 0.9 quantiles of these distributions. To demonstrate this more quantitatively with empirical data, we performed two additional model comparisons using PBCM and data from a subject from Experiment 5. This data set was chosen because this particular subject made error responses more slowly than correct responses. Since the goal of these comparisons is to assess the models' ability to account for a particular pattern of responses, using a single set of data with that pattern is sufficient for demonstration purposes. First, we compared the fixed boundary model (including all between-trial variability parameters) with a version of the collapsing boundary model which included variability in starting-point and non-decision time, but not in drift rate. Slow errors in the fixed-boundary model are primarily explained by between-trial variability in drift rate so this comparison is intended to contrast the two potential mechanisms for producing slow errors.

The fits of these models and the results of the PBCM analyses are shown in Fig. 11(B). The fixed boundary model is better able to capture the slow errors observed in the data, and the observed difference in GOF between the two models (plotted as a triangle) is below the PBCM criterion indicating that the fixed boundary model is preferred. It may be the case that the collapsing boundary model would be preferred for other data sets with less pronounced error slowing, but our results show that it is unable to account for the full range of error RTs that are found experimentally. Second, we compared a version of the collapsing boundary model with between-trial variability with a version of the collapsing boundary model without between-trial variability. The fits of these models and the results of the PBCM analyses are shown in Fig. 11(C). The collapsing boundary model with variability is better able to capture the slow errors in the data, and the observed difference in GOF between the two models is above the PBCM criterion indicating that the model with variability is preferred over the model without variability. These comparisons demonstrate that these between-trial variability parameters are needed to fit slow errors and that the addition of collapsing boundaries to the diffusion model does not eliminate the need for this source of variability. While the collapsing boundary model is able to produce slower error responses without including between-trial variability (as shown in Fig. 11(A)), the slowdown it is able to produce is quite small compared to that produced by a fixed boundary model with variability (as shown in Fig. 11(B)). For this data set, the slowdown that the collapsing boundary model is able to produce without variability parameters is also considerably smaller than the slowdown observed in the data. For these analyses, our main focus has been on comparing diffusion models with fixed or collapsing boundaries, rather than on the role of variability in these models. Although it is possible that the inclusion of variability parameters may also affect conclusions about boundaries, this should be investigated with data more suited to discriminating between model parameterizations (in this case, data with faster or slower errors).

We also examined how the presence or absence of between-trial variability affected model mimicry. We simulated data from a collapsing boundary model with either a large and fast or small and slow amount of collapse (κ values of 0.3 and 0.9 and $t_{0.5}$ values of 0.75 and 0.15), 4000 observations per condition, and either typical values of between-trial variability (as found in fits to data) or between-trial variability parameters set close to zero (see Table A.1 for exact values). We then fit a fixed boundary model with between-trial variability to all four simulated data sets. The resulting fits are plotted in Fig. 12. The triangles represent the simulated results from the collapsing boundary models, and the lines represent the fits of the fixed boundary models. The fixed boundary model is able to provide a good fit to the collapsing boundary data for all of the simulations except the one with a large amount of boundary collapse and no between-trial variability. Even in this case, the fixed boundary model only misses the 0.9 quantiles of the RT distributions.

In value-based decision making, Milosavljevic et al. (2010) found that the addition of collapsing boundaries to a simple diffusion model (i.e., one without between-trial variability parameters) provided an improvement in model fit while the addition of collapsing boundaries to a full diffusion model (i.e., one with between-trial variability in drift) did not. This result is consistent with our cross-fitting results in that the greatest difference between the two versions of the model is observed when the between-trial variability parameters were fixed to zero. However, Milosavljevic et al.'s result was based on a simple comparison of BIC values for each model, so it is possible that the addition of collapsing boundaries to a simple diffusion model would not be considered an improvement when using a more appropriate model selection method. Moreover, the simple diffusion model considered by Milosavljevic et al. is not generally used in applications because it is unable to fit error response times that are faster or slower than correct response times (Laming, 1968; Ratcliff, 1981; Ratcliff & Tuerlinckx, 2002). It is therefore unsurprising that a simple collapsing boundary model performs better than a restricted version of the fixed boundary model that is known to not perform well.

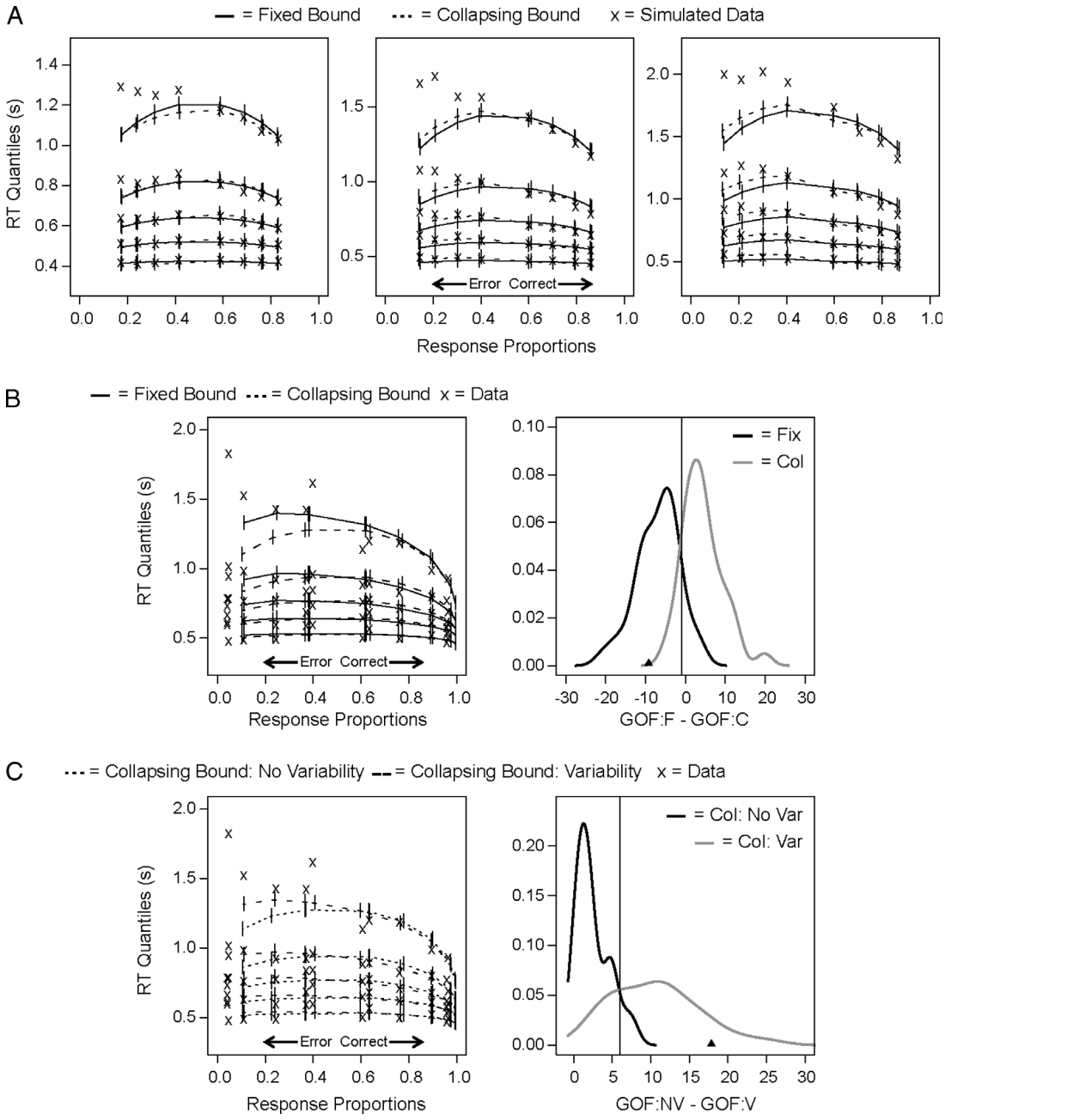

Fig. 11. A: Fits of the fixed and collapsing bound models without between-trial variability parameters to simulated data sets with slow errors. Response proportions are plotted along the *x*-axis and RT quantiles (0.1, 0.3, 0.5, 0.7, and 0.9) are plotted vertically. B: PBCM comparison of a fixed bound model with between-trial variability in drift rate, starting point, and non-decision time and a collapsing bound model with no between-trial variability in drift rate. The figure on the left shows the fits of each model to data, and the figure on the right shows the difference in GOF distributions for the two models when fit to data generated from the fixed bound model (black line) or data generated from the collapsing bound model (gray line) along with the observed GOF difference from fits to the data (triangle). C: PBCM comparison of a collapsing bound model with no between-trial variability parameters and a collapsing bound model with between-trial variability in drift rate, starting point, and non-decision time. The figure on the left shows the fits of each model to data, and the figure on the right shows the difference in GOF distributions for the two models when fit to data generated from the model without variability (black line) or data generated from the model with variability (gray line) along with the observed GOF difference from fits to the data (triangle).

8. Discussion

Our results highlight the importance of using appropriate model selection methods that address model complexity and flexibility. Rather than simply considering how well a given model fits a set of data, the PBCM approach also takes into account the relative flexibility of alternative models by examining how well each model is able to account for data generated by another model. Contrary to claims made elsewhere in the literature, application of the PBCM model selection method showed that allowing the decision boundaries of a diffusion model to decrease over time does not result in a substantial improvement over a model with fixed decision boundaries. As mentioned previously, the model comparison method utilized in this work is data-specific and only provides information about which model provides a better fit to data with these particular parameters and sample sizes. Subject to this constraint, this approach has revealed issues of substantial mimicry between the two models and poorer parameter recovery

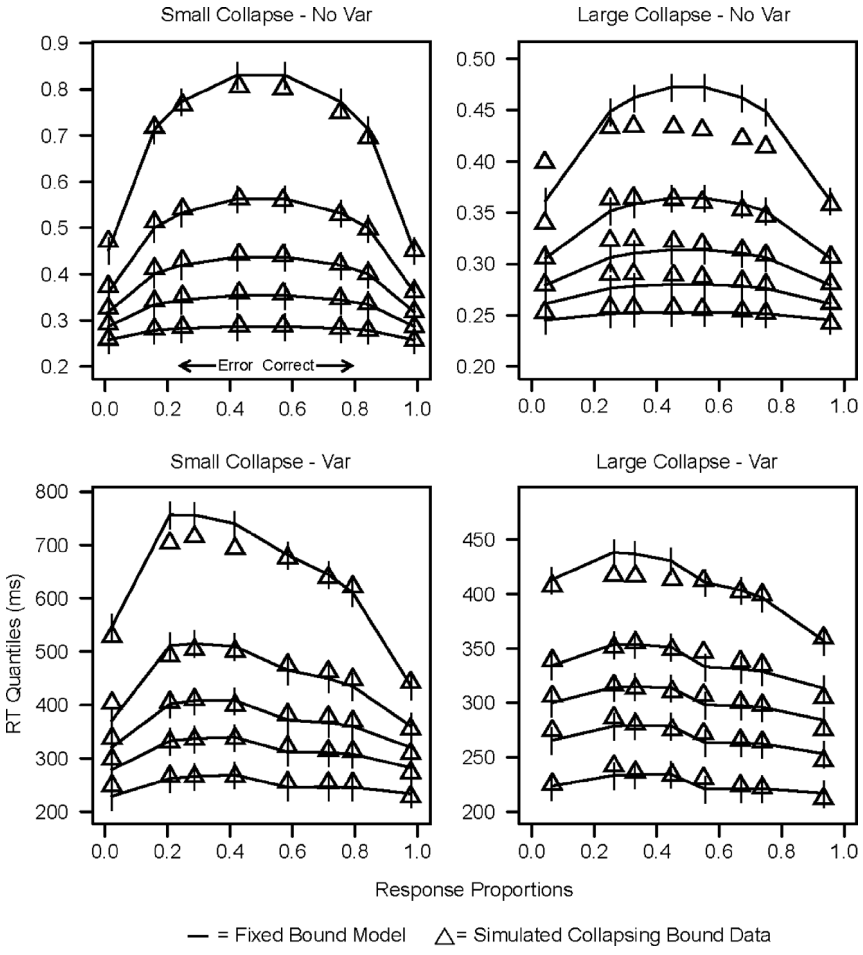

Fig. 12. Fits of the fixed bound model (lines) to data generated from the collapsing bound model (triangles). Data were generated from a collapsing bound model with either a large or small amount of collapse ($\kappa = 0.3$ or 0.9) and with between-trial variability parameters either fixed to zero or fixed to typical values (based on fits to data).

for the collapsing boundary model. It seems reasonable to expect that these same issues would exist for these two models for other data sets.

While the collapsing boundary model did not provide a better fit to the data in most of our experiments, the model may be useful in characterizing other tasks or populations. Researchers using expanded judgment or deferred decision-making tasks have found that subjects are willing to make a decision based on smaller amounts of evidence as time progresses within a trial (Busemeyer & Rapoport, 1988; Rapoport & Burkheimer, 1971; Sanders & Ter Linden, 1967). However, these tasks are very different from the fast perceptual judgments in our experiments. In these types of tasks, new information is presented to the subjects at a relatively slow rate (e.g., a new piece of information every 2 s) and the tasks may also impose an explicit cost on information sampling. In contrast, the response times in all of our experiments were made in well under two seconds and the information necessary to make a decision was presented all at once and maintained until subjects had made a response. Given the differences between these tasks, there is no reason to assume that decision thresholds and processing would behave in a similar manner in for both tasks.

Other support for the collapsing boundary version of the diffusion model has come from the neurophysiological literature on awake behaving monkeys (Churchland et al., 2008; Hanks et al., 2011). The results of single-cell recording studies of monkeys performing saccade-to-target decision tasks have been interpreted as evidence for an 'urgency signal.' Consistent with this interpretation, the RT distributions of monkeys may be more symmetric than typical RT distributions produced by human subjects. However, numerous other researchers have continued to use models which do not include collapsing boundaries or an urgency signal (Boucher, Palmeri, Logan, & Schall, 2007; Costello, Zhu, Salinas, & Stanford, 2013; Ding & Gold, 2010, 2012; Purcell et al., 2010; Purcell, Schall, Logan, & Palmeri, 2012; Ramakrishnan & Murthy, 2013; Ramakrishnan et al., 2012; Salinas, Shankar, Costello, Zhu, & Stanford, 2010; Salinas & Stanford, 2013; Shankar et al., 2011; Stanford, Shankar, Massoglia, Costello, & Salinas, 2010) and other monkey studies have yielded more typical, positively-skewed RT distributions (Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007), as well as those that were fit with ex-Gaussian distributions, (Camalier et al., 2007; Heitz & Schall, 2012, 2013; Middlebrooks & Schall, 2014). Additionally, while a standard diffusion process does not map directly onto neural firing rates, patterns of neural firing rates have been modeled using racing diffusion processes without any kind of urgency signal (Ratcliff et al., 2007). While monkeys may be more likely than human subjects to produce data that are well-described by a model with collapsing boundaries (Hawkins et al., 2015), there seems to be considerable variability in monkey behavior, possibly as a result of differences in training.

Our results complement and partially replicate those of Hawkins et al. (2015) who found that a collapsing boundary version of the diffusion model did not provide an improvement in fit over the standard fixed boundary model for most human subjects based on BIC. However, our methods differed from those of Hawkins et al. in that we used exact rather than simulated predictions, a different functional form for the boundary (one that has been used in the literature), and PBCM model selection

rather than BIC (Hawkins et al. also did some model recovery analyses, although they did not use the resulting GOF difference distributions for model selection).

Hawkins et al. fit several versions of a Weibull boundary model allowing various boundary parameters to vary freely or fixing them to particular values. The results presented in their figures correspond to a version of the model with boundaries that are forced to collapse relatively late within each trial (as opposed to an earlier or more gradual collapse). When this model was fit to the Ratcliff and McKoon (2008) data (our Experiment 5), their Weibull model produced best-fitting collapsing boundaries with little to no collapse (see Fig. 5 in Hawkins et al.). Model recovery analyses for their Weibull model then revealed that the Weibull model and fixed-boundary model could not be discriminated using these data (i.e., the GOF difference distributions were almost completely overlapping). This illustrates one potential issue with the PBCM approach which we also observed for two of our data sets (Experiments 1 and 4): when comparing models that produce very similar patterns of data, the GOF difference distributions produced will be highly overlapping, indicating a high degree of mimicry between the models and low discriminability between the two. In other words, the data-informed PBCM can only distinguish between models that are functionally different (i.e., generate different patterns of data). Given a pair of models with high mimicry and low discriminability, it is entirely appropriate to rely on other model selection tools. In contrast, when we fit the data from Experiment 5 with a hyperbolic ratio boundary model, the best-fitting collapsing boundaries had a moderate amount of collapse, although the boundaries still varied only slightly from fixed boundaries in the regions where subjects made most of their responses (see Fig. 5). For this experiment, the model predictions were sufficiently different that the PBCM approach was able to discriminate between the two models and slightly favored the fixed boundary model.

The PBCM approach addresses both model mimicry and model selection. The amount of overlap between the two distributions of differences in GOF demonstrates the amount of mimicry between the two models (i.e., their ability to produce the same patterns of data). When there is not complete overlap between these two distributions, they can also be used to determine an optimal criterion for recovering the generating model. Methods like BIC and AIC will not always correctly recover the generating model (e.g. Donkin, Tran, & Nosofsky, 2014; Pitt & Myung, 2002) and can over- or under-penalize models depending on their complexity, because the penalty term is dependent on the number of parameters in the model and the functional form of these additional parameters can vary.

9. Conclusions

Overall, the collapsing boundary version of the diffusion model did not provide an improvement on the standard fixed boundary model in our experiments or those considered by Hawkins et al. (2015). When an appropriate model selection method was used, the fixed boundary model was chosen as the preferred model for the majority of the data. Moreover, the additional parameters in the model with collapsing boundaries were not well recovered and were correlated with other model parameters. While there may be tasks or populations for which the collapsing boundary model is well-suited, this extension of the diffusion model is unnecessary to characterize the performance of human subjects in standard decision-making situations.

Acknowledgments

This article was supported by National Institute on Aging grant R01-AG041176 to Roger Ratcliff and Australian Research Council Discovery Grant DP140102970 to Philip L. Smith.

Table A.1
Parameter values for simulated data from Fig. 12 in the main text.

v_1	v_2	v_3	v_4	η
0.025	0.10	0.15	0.40	0.0001
0.025	0.10	0.15	0.40	0.0001
0.025	0.10	0.15	0.40	0.1
0.025	0.10	0.15	0.40	0.1

a	z	s_z	T_{er}	s_t	κ	$t_{0.5}$
0.12	0.06	0.0001	0.2	0.01	0.3	0.75
0.12	0.06	0.0001	0.2	0.01	0.9	0.15
0.12	0.06	0.01	0.2	0.12	0.3	0.75
0.12	0.06	0.01	0.2	0.12	0.9	0.15

Appendix A. Parameter recovery and simulations

When fitting a model to noisy data it is typical for there to be some correlation between model parameters. In the fixed boundary version of the diffusion model, Ratcliff and Tuerlinckx (2002) demonstrated that there were positive correlations between recovered parameter values because response times tend to affect several parameters in similar ways (i.e., to accommodate longer response times, the model will likely produce a larger boundary separation, a larger non-decision time, and a larger drift rate). However, it is not known if the collapsing boundary version of the model will exhibit similar correlations between recovered parameter values. Since the PBCM approach involved fitting multiple simulated data sets generated from a single set of model parameters, we can check for correlations between the parameter values recovered from these fits. Here we investigate whether recovered boundary parameters from the PBCM fits from Experiment 5 correlate with other key model parameters. Fig. A.1 plots the recovered values for κ and $t_{0.5}$ against some of the other recovered parameters for these simulated data sets (one of the drift rates, boundary separation, and non-decision time). There were significant correlations between the recovered value of κ and each of the other three parameters, and there were significant correlations between $t_{0.5}$ and both drift rate and boundary separation. These correlations between parameters related to boundary collapse and parameters that should not be related to boundary collapse (such as drift rate) undermine the usefulness of the model in the sense that parameter interpretation as a whole becomes less straightforward. See Table A.1.

Appendix B. Response time distributions and response probabilities for a diffusion process with time-varying boundaries

Predicted response time distributions and response probabilities for the model with time-varying boundaries were obtained using the integral equation methods described in Smith (2000). These methods were originally developed in the theoretical neurobiology literature to study the firing time distributions of integrate-and-fire neurons. The method described by Smith was based on a method proposed by Buonocore, Giorno, Nobile, and Ricciardi (1990) and incorporates refinements suggested by Gutiérrez Jáimez, Román Román, and Torres Ruiz (1995). A summary of the method can be found in the Appendix of Ratcliff and Smith (2004), who used it to obtain first-passage-time distributions for the Ornstein–Uhlenbeck process, although they did not consider the time-varying boundary case. The method is based on a general renewal equation representation of the first-passage-time density of a diffusion process, derived by Fortet (1943).

Fortet's representation is based on the fact that a sample path of a diffusion process at a point on or an infinitesimal distance outside the boundary, $a_i(t), i = 1, 2$, at time t must have made a first such boundary crossing, either at time t or at some time τ prior to t. The transition density of the process at time t can therefore be decomposed into a product of two terms: One is the first passage

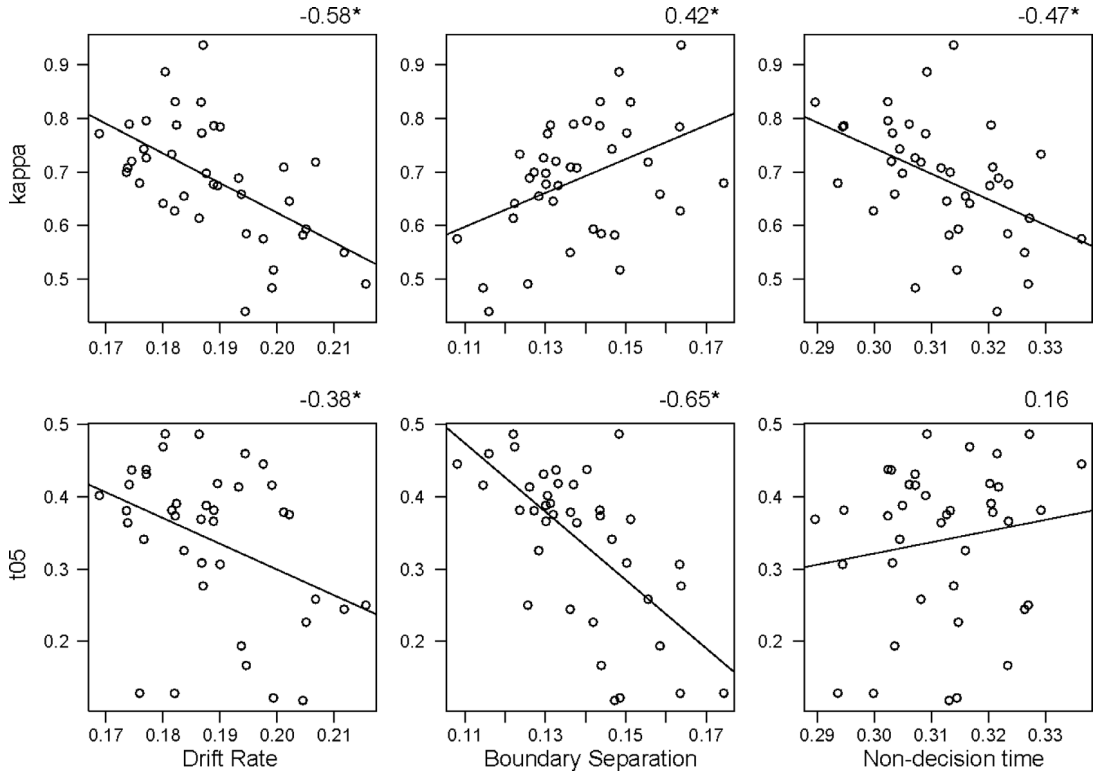

Fig. A.1. Correlations between recovered boundary collapse parameters (κ and $t_{0.5}$) and other recovered model parameters (drift rate, boundary separation, and non-decision time) from fits of simulated data generated from the parameters from Experiment 5. Correlation coefficients are presented above each figure ($* = p < 0.05$).

time density of the process through the boundary $a_i(\tau)$ at time τ; the other is the free transition density of the process from the point $a_i(\tau)$ to the point $a_i(t)$ in the subsequent interval $[\tau, t]$. For a two-boundary process, the first boundary crossing can occur either at the same boundary as the one where the process is found at time t or at the other boundary. Following the first boundary crossing, the process can make an arbitrary number of additional boundary crossings of either the upper or lower boundary. The transition density at the point $a_i(t)$ at time t is obtained integrating the product of the first passage time density and the free transition density over all values of time τ, $\tau < t$, and summing over the two boundaries. Two such equations are obtained: one for processes at the upper boundary and one for processes at the lower boundary. Details of this decomposition may be found in Smith (2000) and in the Appendix of Ratcliff and Smith (2004).

The integral equation method of Buonocore et al. (1990) is based on a transformation of the Fortet equations, which renders them suitable for numerical computation. We define $g_A[a_1(t), t|z, 0]$ and $g_B[a_2(t), t|z, 0]$ to be, respectively, the first-passage-time probability densities for a process starting at z at time 0 through either the upper boundary, $a_1(t)$, or the lower boundary, $a_2(t)$, at time t, where $a_2(0) < z < a_1(0)$. Our notation assumes that if the first boundary crossed is the upper boundary, then response R_A is made; if the first boundary crossed is the lower boundary, then response R_B is made. Buonocore et al. obtained the following integral equation representations of the first passage time densities:

$$g_A[a_1(t), t|z, 0] = -2\Psi[a_1(t), t|z, 0] + 2\int_0^t g_A[a_1(\tau), \tau|z, 0]\Psi[a_1(t), t|a_1(\tau), \tau]d\tau + 2\int_0^t g_B[a_2(\tau), \tau|z, 0]\Psi[a_1(t), t|a_2(\tau), \tau]d\tau$$

and

$$g_B[a_2(t), t|z, 0] = 2\Psi[a_2(t), t|z, 0] - 2\int_0^t g_A[a_1(\tau), \tau|z, 0]\Psi[a_2(t), t|a_1(\tau), \tau]d\tau - 2\int_0^t g_B[a_2(\tau), \tau|z, 0]\Psi[a_2(t), t|a_2(\tau), \tau]d\tau.$$

In these equations, $\Psi[a_i(t), t|a_j(\tau), \tau]$ is a kernel function that depends jointly on the boundaries, their derivatives, and the free transition density of the process, denoted $f(x, t|y, \tau)$. The free transition density gives the probability that a process starting at position y at time τ will be found at position x at time t, ignoring the effects of any boundaries. For a Wiener, or Brownian motion, diffusion process with time-varying boundaries and constant drift, the kernel function is

$$\Psi[a_i(t), t|a_j(\tau), \tau] = \frac{f[a_i(t), t|a_j(\tau), \tau]}{2}\left\{a_i'(t) - \frac{a_i(t) - a_j(\tau)}{t - \tau}\right\},$$

where

$$f[a_i(t), t|a_j(\tau), \tau] = \frac{1}{\sqrt{2\pi s^2(t - \tau)}}\exp\left\{-\frac{[a_i(t) - a_j(\tau) - v(t - \tau)]^2}{2s^2(t - \tau)}\right\}$$

is the transition density of a Wiener process with drift rate v and diffusion coefficient s^2, and $a_i'(t)$ is the derivative with respective to time of boundary i.

The preceding integral equations can be discretized and evaluated simultaneously and recursively to yield approximations to the first-passage-time densities by trapezoidal integration

$$g_A[a_1(k\Delta), k\Delta|z, 0] = -2\Psi[a_1(k\Delta), k\Delta|z, 0] + 2\Delta\sum_{j=1}^{k-1} g_A[a_1(j\Delta), j\Delta|z, 0]\Psi[a_1(k\Delta), k\Delta|a_1(j\Delta), j\Delta]$$

$$+ 2\Delta \sum_{j=1}^{k-1} g_B[a_2(j\Delta), j\Delta|z, 0]\Psi[a_1(k\Delta), k\Delta|a_2(j\Delta), j\Delta]$$

and

$$g_B[a_2(k\Delta), k\Delta|z, 0] = 2\Psi[a_2(k\Delta), k\Delta|z, 0] - 2\Delta \sum_{j=1}^{k-1} g_A[a_1(j\Delta), j\Delta|z, 0]\Psi[a_2(k\Delta), k\Delta|a_1(j\Delta), j\Delta] - 2\Delta \sum_{j=1}^{k-1} g_B[a_2(j\Delta), j\Delta|z, 0]\Psi[a_2(k\Delta), k\Delta|a_2(j\Delta), j\Delta],$$

for $k = 2, 3, \ldots$, where Δ is the time step of the discrete approximation. For $k = 1$, the equations reduce to

$$g_A[a_1(\Delta), \Delta|z, 0] = -2\Psi[a_1(\Delta), \Delta|z, 0]$$

and

$$g_B[a_2(\Delta), \Delta|z, 0] = 2\Psi[a_2(\Delta), \Delta|z, 0].$$

Buonocore et al. (1990) proved that the discrete approximations converge to the true first passage time densities as Δ becomes small.

The equations $g_A[a_1(t), t|z, 0]$ and $g_B[a_2(t), t|z, 0]$ are the joint densities of responses R_A and R_B made at time t. The associated response probabilities, $P(R_A)$ and $P(R_B)$, are obtained by numerically integrating the joint densities from 0 to $t_{\max}$, the maximum time index. In applications, the value of $t_{\max}$ must be chosen to be large enough that the response probabilities sum to 1.0.

To obtain the model predictions used in this article, the boundaries $a_i(t)$ were set to the hyperbolic ratio functions given in the text. The predictions were obtained using an implementation of the equations written in *C* and called from Matlab, using 300 time steps on the range [0, 2] s. The predictions for models with across-trial distributions of drift and starting point were computed as probability mixtures of first passage time distributions, as they are in the standard model. These mixtures were obtained by numerically integrating a 19-point approximation to a normal distribution of drift and an 11-point approximation to a uniform distribution of starting points. The mixture distributions were then convolved with a discrete approximation to a uniform distribution of nondecision times, using the same time step as was used to compute the component first passage time densities. (Mixing and convolution are commutative, so the convolution operation only had to be performed once, after computing the mixture, rather than individually, for each of its components.) The resulting density functions were numerically integrated to obtain predicted cumulative distributions for correct responses and errors. The cumulative distributions were then numerically inverted to obtain the predicted probability masses between the quantiles of the empirical distribution functions used in the fit statistic. As background to the work reported in Ratcliff and Smith (2004), they compared the performance of the integral equation method, discretized in a similar way as here, to that of the infinite-series solution used in the standard diffusion model, and found the predictions and the parameters recovered from fitting were essentially indistinguishable.

In other, related applications of the integral equations method, Smith and Ratcliff (2009) used it to obtain predictions for their integrated system model, which assumes that the decision process is driven by a time varying visual short-term memory trace that is formed under the control of spatial attention. This leads to a model with time-varying drift rates and diffusion coefficients. Smith et al. (2014) used the method to model decisions about stimuli embedded in dynamic noise. This allowed them to derive predictions for models with time-varying drift rates and time-varying response inhibition.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Caski (Eds.), *Proceeding of the second international symposium on information theory*. Budapest: Akademiai Kiado.

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

Bode, S., Sewell, D. K., Lilburn, S., Forte, J. D., Smith, P. L., & Stahl, J. (2012). Predicting perceptual decision biases from early brain activity. *The Journal of Neuroscience*, *32*(36), 12488–12498.

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*(2), 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.

Buonocore, A., Giorno, V., Nobile, A. G., & Ricciardi, L. (1990). On the two-boundary first-crossing-time problem for diffusion processes. *Journal of Applied Probability*, *27*, 102–114.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, *32*, 91–134.

Cain, N., Barreiro, A. K., Shadlen, M., & Shea-Brown, E. (2013). Neural integrators for decision making: a favorable tradeoff between robustness and sensitivity. *Journal of Neurophysiology*, *109*, 2542–2559.

Camalier, C. R., Gotler, A., Murthy, A., Thompson, K. G., Logan, G. D., Palmeri, T. J., et al. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research*, *47*(16), 2187–2211.

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.

Costello, M. G., Zhu, D., Salinas, E., & Stanford, T. R. (2013). Perceptual modulation of motor — but not visual — responses in the frontal eye field during an urgent-decision task. *Journal of Neuroscience*, *33*, 16394–16408.

Deneve, S. (2012). Making decisions with unknown sensory reliability. *Frontiers in Neuroscience*, *6*, 1–13.

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

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

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.

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, *16*(6), 1129–1135.

Donkin, C., Tran, S. C., & Nosofsky, R. (2014). Landscaping analyses of the ROC predictions of discrete-slots and signal-detection models of visual working memory. *Attention, Perception, and Psychophysics*, *76*(7), 2103–2116.

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.

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

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

Fortet, R. (1943). Les fonctions aléatoires du type de Markoff associées à certaines équations linéaires aux derivées partielles du type parabolique. *Journal de Mathématiques Pures et Appliquées*,.

Frazier, P., & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines. *Advances in Neural Information Processing Systems*, *20*, 465–472.

Geddes, J., Ratcliff, R., Allerhand, M., Childers, R., Wright, R. J., Frier, B. M., et al. (2010). Modeling the effects of hypoglycemia on a two-choice task in adult humans. *Neuropsychology*, *24*, 652–660.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

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

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*, 389–413.

Gutiérrez Jáimez, R., Román Román, P., & Torres Ruiz, F. (1995). A note on the Volterra integral equation for the first-passage-time probability. *Journal of Applied Probability*, *32*, 635–648.

Hanks, T. D., Kiani, R., & Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife*, *3*, http://dx.doi.org/10.7554/eLife.02260.
Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *Journal of Neuroscience*, *31*, 6339–6352.
Hawkins, G. E., Forstmann, B. U., Wagenmakers, E. J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience*, *35*(6), 2476–2484.
Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed–accuracy tradeoff. *Neuron*, *76*(3), 616–628.
Heitz, R. P., & Schall, J. D. (2013). Neural chronometry and coherency across speed-accuracy demands reveal lack of homomorphism between computational and neural mechanisms of evidence accumulation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *368*(1628), 20130071.
Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, England: Oxford University Press.
Karsilar, H., Simen, P., Papadakis, S., & Balci, F. (2014). Speed-accuracy tradeoff under response deadlines. *Frontiers in Decision Neuroscience, Research Topic on Speed-Accuracy Tradeoff*, *8*, 248.
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). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of The United States of America*, *108*(34), 13852–13857.
Kryukov, V. I. (1976). Wald's identity and random walk models for neuron firing. *Advances in Applied Probability*, *8*(2), 257–277.
Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York, NY: Wiley.
Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making*, *6*, 651–687.
Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, *13*, 1257–1269.
Middlebrooks, P. G., & Schall, J. D. (2014). Response inhibition during perceptual decision making in humans and macaques. *Attention, Perception and Psychophysics*, *76*(2), 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.
Moran, R. (2015). Optimal decision making in heterogeneous and biased environments. *Psychonomic Bulletin & Review*, *22*(1), 38–53.
Morgan, M. J., & Ward, R. (1980). Conditions for motion flow in dynamic visual noise. *Vision Research*, *20*(5), 431–435.
Mulder, M. J., Boekel, W., Ratcliff, R., & Forstmann, B. U. (2014). Cortico-subthalamic connection predicts sensitivity to reward in perceptual decision-making. *Brain Structure and Function*, *219*, 1239–1249.
Mulder, M. J., Bos, D., Weusten, J. M. H., van Belle, J., van Dijk, S. C., Simen, P., et al. (2010). Basic impairments in regulating the speed-accuracy tradeoff predict symptoms of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *68*, 1114–1119.
Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*(1), 190–204.
Niwa, M., & Ditterich, J. (2008). Perceptual decisions between multiple directions of visual motion. *The Journal of Neuroscience*, *28*(17), 4435–4445.
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.
Parzen, E., Tanabe, K., & Kitagawa, G. (Eds.) (1998). *Selected papers of Hirotugu Akaike*. New York: Springer.
Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*(10), 421–425.
Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*(4), 1113–1143.
Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From salience to saccades: Multiple-alternative gated stochastic accumulator model of visual search. *The Journal of Neuroscience*, *32*(10), 3433–3446.
Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1226–1243.
Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge: Blackwells.
Ramakrishnan, A., & Murthy, A. (2013). Brain mechanisms controlling decision making and motor planning. In V. C. Pammi, & N. Srinivasan (Eds.), *Decision making: Neural and behavioural approaches* (pp. 321–345). Amsterdam, Netherlands: Elsevier.
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.
Rao, R. P. N. (2010). Decision making under uncertainty: a neural model based on partially observable Markov decision processes. *Frontiers in Computational Neuroscience*, *4*, 146.
Rapoport, A., & Burkheimer, G. J. (1971). Models for deferred decision making. *Journal of Mathematical Psychology*, *8*, 508–538.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, *88*, 552–572.
Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review*, *92*, 212–225.
Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, *95*, 238–255.
Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*, 195–237.
Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging*, *23*, 900–916.
Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*, *120*, 281–292.
Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 870–888.
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., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182.
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., Love, J., Thompson, C. A., & Opfer, J. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*, *83*, 367–381.
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 Sciences*, *106*, 6539–6544.
Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.
Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.
Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decisionmaking. *Journal of Experimental Psychology: General*, *139*, 70–94.
Ratcliff, R., & Starns, J. J. (2013). Modeling response times, choices, and confidence judgments in decision making: recognition memory and motion discrimination. *Psychological Review*, *120*, 697–719.
Ratcliff, R., & Strayer, D. (2014). Modeling simple driving tasks with a one-boundary diffusion model. *Psychonomic Bulletin and Review*, *21*, 577–589.
Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Perception and Psychophysics*, *19*, 278–289.
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., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception and Psychophysics*, *65*, 523–535.
Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*, 408–424.
Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*, 127–157.
Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*, 464–487.
Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches tovdealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438–481.
Ratcliff, R., & Van Dongen, H. P. A. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin & Review*, *16*, 742–751.
Ratcliff, R., & Van Dongen, H. P. A. (2011). A diffusion model for one-choice reaction time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*, 11285–11290.
Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*, 261–300.
Roitman, J. D., & Shadlen, M. N. (2002). Responses of neurons in the lateral interparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*, 9475–9489.
Salinas, E., Shankar, S., Costello, M. G., Zhu, D., & Stanford, T. R. (2010). Waiting is the hardest part: Comparison of two computational strategies for performing a compelled-response task. *Frontiers in Computational Neuroscience*, *4*, 153.
Salinas, E., & Stanford, T. R. (2013). The countermanding task revisited: Fast stimulus detection is a key determinant of psychophysical performance. *The Journal of Neuroscience*, *33*(13), 5668–5685.
Sanders, A. F., & Ter Linden, W. (1967). Decisions during paced arrival of probabilistic information. *Acta Psychologica*, 27170–27177.
Schall, J. D. (2003). Neural correlates of decision processes: Neural and mental chronometry. *Current Opinion In Neurobiology*, *13*(2), 182–186.
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 Sciences*, *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.
Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences*, *93*, 628–633.
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.
Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*, 567–591.
Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*(3), 408–463.
Smith, P. L., & McKenzie, C. L. (2011). Diffusive information accumulation by minimal recurrent neural models of decision making. *Neural Computation*, *23*(8), 2000–2031.
Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*(2), 283–317.
Smith, P. L., Ratcliff, R., & Sewell, D. K. (2014). Modeling perceptual discrimination in dynamic noise: Time-changed diffusion and release from inhibition. *Journal Of Mathematical Psychology*, 5995–6113.
Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, *44*(12), 1297–1320.
Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 101–117.
Stanford, T. R., Shankar, S., Massoglia, D. P., Costello, M. G., & Salinas, E. (2010). Perceptual decision making in less than 30 milliseconds. *Nature Neuroscience*, *13*(3), 379–385.
Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variability and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, *64*, 1–34.
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.
Thura, D., & Cisek, P. (2014). Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron*, *81*(6), 1401–1416.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.
van Ravenzwaaij, D., Mulder, M. J., Tuerlinckx, F., & Wagenmakers, E. (2012). Do the dynamics of prior information depend on task context? An analysis of optimal performance and an empirical test. *Frontiers In Psychology*, *3*, 1–15.
Viviani, P. (1979a). A diffusion model for discrimination of temporal numerosity. *Journal of Mathematical Psychology*, *19*(2), 108–136.
Viviani, P. (1979b). Choice reaction times for temporal numerosity. *Journal Of Experimental Psychology: Human Perception And Performance*, *5*(1), 157–167.
Viviani, P., & Terzuolo, C. A. (1972). On the modeling of the performance of the human brain in a two-choice task involving decoding and memorization of simple visual patterns. *Kybernetik*, *10*, 121–137.
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.
Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*, 140–159.
Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, *19*, 326–339.
Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*, 955–968.
White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010a). Anxiety enhances threat processing without competition among multiple inputs: A diffusion model analysis. *Emotion*, *10*, 662–677.
White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010b). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, *54*, 39–52.
Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, *9*, 60–62.
Wong, K., & Wang, X. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience*, *26*(4), 1314–1328.
Zeguers, M. H. T., Snellings, P., Tijms, J., Weeda, W. D., Tamboer, P., Bexkens, A., et al. (2011). Specifying theories of developmental dyslexia: A diffusion model analysis of word recognition. *Developmental Science*, *14*, 1340–1354.