

A ballistic model of choice response time

Scott Brown

&

Andrew Heathcote

University of California,

Irvine

University of Newcastle,

Australia

Correspondence should be addressed to:

Scott Brown

Email: scottb@uci.edu

Phone: (949) 824 2051

Department of Cognitive Sciences

University of California, Irvine

Irvine, CA 92697-5100

USA

Abstract Word Count: 151

Body Word Count: 5439 + 49 + 328 = (Body + Acknowledgements + Footnotes)

Reference Count: 48

Figure Count: 7

Table Count: 1

Almost all models of simple and choice response time (RT) employ a stochastic (i.e., variable within trial) accumulation decision process. In order to account for the relationship between correct and error choice RT, it has been found necessary to also include between trial variability in the starting point and/or the rate of accumulation, both in linear (Ratcliff & Rouder, 1998) and nonlinear (Usher & McClelland, 2001) stochastic models. We show that a ballistic (i.e., deterministic within trial) model using a simplified version of Usher and McClelland's nonlinear accumulation process, and assuming only between trial variability in the rate and starting point of accumulation, is not only capable of accounting for the relationship between error and correct RT, but can also model other benchmark behavioural phenomena, such as RT distribution and speed-accuracy trade off. We successfully fit our ballistic model to Ratcliff and Rouder's data, which exhibit many of the benchmark phenomena.

Even for fast and easy decisions, "...a simple summation of sensory and motor transduction delays and conduction times in the nervous system cannot account for the duration and variability of reaction times." (Hanes & Schall, 1996, p.427). The slowness and variability of response time (RT) has been almost universally explained by decision processes involving *stochastic* accumulation of information. Stochastic models assume that the accumulated information varies randomly from moment to moment during the decision process. RT is relatively slow because a criterion amount of information must be accumulated before a response is made, and RT is variable because stochastic accumulation causes variability in the amount of time required to reach the criterion. The existence of within-decision variability in information accumulation is not just a theoretical convenience, it has been imbued with psychological importance as instantiating a process of sequential sampling.

We propose a ballistic accumulation (BA) model in which the accumulation process during a trial is deterministic rather than stochastic. We demonstrate that the BA model can provide a general account of choice RT in both *time-controlled* and *information-controlled* paradigms. Time controlled choice tasks require subjects to make decisions at a range of deadlines after the onset of the stimulus. Interest focuses on the growth of accuracy with time, the *speed-accuracy tradeoff* (SAT) function (see, e.g., **Luce, 1986, pp.237-245**). Information controlled tasks require subjects to make a decision when they feel a criterion amount of information has accumulated, with the criterion set according to task demands, such as an emphasis on accuracy or speed. Interest focuses on RT for both correct and error responses, as well as accuracy.

In order to account for SAT, stochastic models, such as Ratcliff and Rouder’s (1998) diffusion model and Usher and McClelland’s (2001) leaky competitive accumulator model, have also had to incorporate two sources of between trial variability, in input strength and in the starting point of evidence accumulation¹. The BA model incorporates *only* these between-trial sources; it is a simplified deterministic version of the Usher and McClelland model. Dropping the stochastic component reduces model complexity, both analytically and computationally. More importantly, it represents a fundamental change in the psychological interpretation of the model; it can no longer be interpreted as a sequential-sampling process. The fits of the BA model reported below also revealed a fundamental change in model dynamics relative to the fitted version of Usher and McClelland’s stochastic model. The fitted BA model displayed “winner-takes-all” behaviour due to much stronger competition. In the next section we describe the BA model and show how it can account for SAT in time-controlled paradigms. We then fit the model to Ratcliff and Rouder’s data from an information-controlled paradigm. Those data total around 10,000 observations from each of three participants (subjects KR, JF and NH) in a two-alternative forced-choice perceptual categorisation task.

The Ballistic Leaky Competitive Accumulator Model

The BA model associates each possible choice response with a unit, having activation, x_i , which follows the deterministic dynamics specified by Equation 1 (t is time). We restrict our attention to the two-choice case ($i=1,2$), although the model naturally extends to any number of choices, and the analytic results are equally simple for any number of choices.

$$dx_i/dt = I_i - k_i x_i + \alpha_i f(x_i) - \beta_i \sum_{j \neq i} f(x_j) \quad (1)$$

A response is triggered as soon as the activation of either unit reaches or exceeds a response criterion (C), which we assume to be the same for both units. The response made is determined by whichever unit reaches that criterion first, and the latency of the response is the sum of the time taken to reach the criterion and a constant time used to represent the duration of non-decision processes (t_0). In principle non-decision time might have a random component (cf. Ratcliff & Tuerlinckx, 2002), but the extra complexity entailed was not necessary to achieve the fits reported here. Illustrative examples of the accumulation trajectories for the BA model are shown in the right hand panel of Figure 1: they are smooth, deterministic functions, as opposed to the noisy stochastic trajectories of Usher and McClelland's (2001) model, shown in the left panel of Figure 1.

Following most other models of choice RT, we made the simplifying assumption that the input to each unit followed a step function, increasing from zero to a positive value, I_i , simultaneously for each unit (see also Smith, Ratcliff & Wolfgang, 2003). I_i varies randomly and independently from unit-to-unit and trial-to-trial due to additive Gaussian noise: $\mathcal{N}(0, \sigma_1)$. The expected value of the total input, $I_s = I_1 + I_2$, is a constant, with decision difficulty manipulated by varying the input difference, $I_d = I_1 - I_2$. The starting point of the accumulation process is also assumed to vary randomly and independently from unit-to-unit and trial-to-trial according to a uniform distribution on $[x_0 - \Delta x_0, x_0 + \Delta x_0]$. Our assumptions about input and starting point variability are identical to Ratcliff and Rouder (1998; see also Ratcliff & Tuerlinckx, 2002, Usher & McClelland, 2001). Note that when activation is small Equation 1 is approximately linear, with solution $x_i(t) = x_i(0) + I_i t$, corresponding to the deterministic dynamics of the diffusion model.

The parameters α_i , β_i and k_i in Equation 1 are non-negative rates of self-excitation, competition (i.e., inhibition of other units) and passive leakage respectively. Their subscripts allow for possibly different values on different response units, but we make the simplifying assumption that these parameters are equal across units. The transfer function, f , has been used to enforce neurologically plausible constraints on the system. For example, employing a linear threshold transfer function $f(x)=\max\{0,x\}$, ensures that only units with positive activation transmit inhibition and self-excitation (c.f., Usher & McClelland, 2001). Although the linear threshold transfer function is neurally plausible, it complicates the dynamics of the model, introducing three different modes depending on whether activations for both units are negative, positive or have different signs (see Heathcote, 2003). A reviewer commented that this extra complexity may make our model overly flexible, hence we report results only for the simpler model with $f(x)=x$. This version is both analytically simple, and allows us to drop the self-excitation parameter (α), as its effects are indistinguishable from leakage (k). Although not reported here, we have extensively examined the behaviour of the model with $f(x)=\max\{0,x\}$ and found it to be very similar to the simplified model².

When $f(x)\equiv x$, Equation 1 can be solved by elementary matrix algebra, for any number of units (i.e., response alternatives). For two units the solution is (with $x_s=x_1+x_2$, $x_d=x_1-x_2$, $I_s=I_1+I_2$ and $I_d=I_1-I_2$):

$$x_1(t) = \left(\frac{I_s}{2S} + \frac{I_d}{2D} \right) + \left(x_1(0) - \frac{I_s}{2S} \right) e^{-St} - \frac{I_d}{2D} e^{-Dt} \quad D \neq 0$$

$$x_1(t) = x_d(0) + \frac{I_d t}{2} + \frac{I_s}{2S} + \left(x_s(0) - \frac{I_s}{2S} \right) e^{-St} \quad D=0 \quad (2)$$

The solutions for x_2 are identical, except I_d is replaced by $-I_d$, x_1 by x_2 , and x_d by $-x_d$.

The terms $S=k+\beta$ and $D=k-\beta$ are the total leakage for the sum and difference of activations; these terms govern the BA model's dynamics. We assume that S is positive, so the sum of the units' activations converges to a fixed value (I_S/S). When D is positive the difference between the units' activations also converges to a constant value. When D is negative the difference diverges infinitely, so that a response occurs even for very weak input differences. Like the stochastic component in Usher and McClelland's (2001) model, divergent difference dynamics ensure that a response is made even when the input difference is very small. When D is zero, the activation difference diverges linearly, as is the case for the deterministic dynamics of the diffusion model. In general, Equation 2 cannot be solved explicitly for t , but can be solved implicitly for the decision time (i.e., $x_i(t)=C$) using standard root-finding techniques. We used these methods, coupled with numerical integration, to calculate model predictions for the fits reported below. We validated our solutions using Runge-Kutta direct numerical integration on Equation 1 (see Gard, 1988, p.206).

Speed-Accuracy Tradeoff

Stochastic models account for the SAT because increased accumulation time allows the effects of within-trial variability in information accumulation to be integrated out. Linear accumulation models, such as the diffusion model, with only within-trial variability in evidence accumulation, predict perfect asymptotic accuracy for all decisions. However, less than perfect accuracy is usually observed in practice, even with unlimited decision time.

At least two suggestions have been made to allow stochastic models to account for less than perfect asymptotic accuracy. Usher and McClelland (2001) proposed that

accumulation is “leaky” so that information is lost during accumulation, and hence accuracy is imperfect (although asymptotic accuracy in information-controlled paradigms can still be infinite – see Busemeyer & Townsend, 1992). Ratcliff (1978) added between-trial variability in the input to the diffusion model, thus predicting imperfect asymptotic accuracy, as errors due to input variations are unaffected by accumulation time.

The BA model produces a SAT because extra integration time allows the input to overcome noise in the *starting points*. To illustrate, consider the example BA model accumulation trajectories in the right hand panel of Figure 1. The unit with a smaller input (dashed line) started with larger activation, but with extra integration time, it was overtaken by the unit with a larger input. If the response criterion (horizontal line) were set very low, around 0.6, the model would make the wrong response, because the accumulator corresponding to the wrong response begins with a slight advantage and would reach a low response criterion first. Raising the response criterion (to the value shown) allows sufficient integration time for the accumulator corresponding to the correct response to overcome its initial disadvantage. A consequence of Equation 2 is that extending integration time indefinitely allows all effects of start point variability to be removed. However, even then, asymptotic accuracy in the BA model is still imperfect because of variability in input strength.

Typically, the SAT function, with accuracy measured by d' , **is well approximated by a shifted exponential function of decision time in time-controlled tasks. McElree and Doshier (1989) found that, in recognition memory, the shifted exponential function consistently gave a better fit to empirical SAT functions than the SAT function predicted by the diffusion model. Usher and McClelland (2001) also found that their model gave a**

better fit to perceptual choice SAT functions than the diffusion model. However, Ratcliff (1988) showed that the diffusion model provides a better fit when some responses in the time-controlled paradigm are based on information-controlled decisions. In any case, the shifted exponential provides a good approximation to empirical SAT functions.

The SAT function for the BA model can be derived by first considering start-point variability alone. Let $U_{[x, \Delta]}$ represent the cumulative density function for the uniform distribution on $[x-\Delta, x+\Delta]$, that is:

$$U_{[x, \Delta]}(z) = \begin{cases} 0 & z < x - \Delta \\ (z - x + \Delta)/(2\Delta) & x - \Delta \leq z \leq x + \Delta \\ 1 & z > x + \Delta \end{cases}$$

Then for a fixed input difference I_D :

$$P(x_1(t) > x_2(t)) = U_{[x_0, \Delta x_0]}(-I_D(1 - e^{-Dt})) \quad D \neq 0$$

$$P(x_1(t) > x_2(t)) = U_{[x_0, \Delta x_0]}(I_D t) \quad D = 0$$

Given input noise is independent across units, I_D is distributed $\underline{N}(0, \sqrt{2} \sigma_I)$. We can then find the SAT function by integrating over the distribution of I_D and transforming from probability to d' , using the standard normal density function (ϕ) and cumulative function (Φ):

$$\begin{aligned} d' &= 2 \cdot \Phi^{-1} \left(\int_{\mathbb{R}} U_{[x_0, \Delta x_0]} \left(-\frac{y}{D} (1 - e^{Dt}) \right) \phi \left(\frac{y - I_D}{\sqrt{2} \cdot \sigma_I} \right) dy \right) & D \neq 0 \\ d' &= 2 \cdot \Phi^{-1} \left(\int_{\mathbb{R}} U_{[x_0, \Delta x_0]}(yt) \phi \left(\frac{y - I_D}{\sqrt{2} \cdot \sigma_I} \right) dy \right) & D = 0 \end{aligned} \quad (3)$$

The symbols in Figure 2 illustrate representative SAT functions for the BA model. As required, d' increases gradually with time and approaches a finite asymptote. The solid lines in Figure 2 represent best fitting shifted exponential functions, $d' = A(1 - e^{-Bt})$,

estimated using least-squares nonlinear regression. Note that the BA model's SAT functions are very closely approximated by shifted exponential functions (R^2 values were .986, .990 and .996 for subjects KR, JF and NH respectively). The parameters used to generate BA model predictions in Figure 2 were taken from the fits to Ratcliff and Rouder's (1998) data from an information-controlled paradigm, reported below, demonstrating that the BA model can fit both time- and information-controlled data simultaneously. We also found that approximately shifted exponential SAT functions hold for a wide range of parameter values when $D \leq 0$, and Brown (2002) reports similar findings when $D > 0$.

Fast and Slow Errors

The addition of variability in input strengths fixes another problem for the diffusion model with only Gaussian accumulation noise: it predicts equal correct and error RT distributions. Equal correct and error RTs are occasionally observed but **typically, when response accuracy is emphasized and the decision to be made is relatively difficult, error RTs are longer than correct RTs, a phenomenon we will call "slow errors"**. The addition of between trial variability in input strengths allows the diffusion model to produce slow errors (Ratcliff, 1978). In contrast, Usher and McClelland's (2001) model with only within-trial variability in information accumulation can produce equal correct and error RTs or slow errors, depending on the levels of lateral inhibition and leakage.

When simple decisions are required, and response speed is emphasized, error RTs are typically faster than correct RTs; "fast errors" (e.g., Ratcliff, Van Zandt & McKoon, 1999; see Luce, 1986, p.233 for a review). Fast errors require a third source of variability to be incorporated into the diffusion model, between-trial variability in either the criterion

or start point (these changes are isomorphic when integration is linear, as in the diffusion). Start point variability was originally suggested by Laming (1968; see also Rouder, 1996) as being caused by pre-stimulus accumulation. Usher and McClelland (2001) also incorporated between-trial start point variability into their model in order to account for fast errors, although they did not fit this version of their model to data from an information controlled task, as only slow errors were observed in their data.

A pattern that has proven particularly diagnostic for selecting models of choice RT (e.g., Van Zandt et al., 2000; Ratcliff & Smith, submitted) is a *crossover* effect, in which faster and slower error RTs are observed in easy and hard stimulus discrimination conditions respectively, even when these conditions are randomly intermixed within experimental blocks. Hence, general choice RT models must be able to accommodate crossovers by changing *only* stimulus-related parameters.

Figure 3 illustrates the crossover pattern in Ratcliff and Rouder's (1998) data using quantile-probability (QP) functions. The 'crossover' pattern is most evident in the speed-emphasis data, from subjects KR and JF. QP functions generalise latency probability functions (Audley & Pike, 1965) by plotting quantiles of RT distributions as a function of the probability of a response (see Ratcliff, 2002, for more details). Points on the left of the graph represent the lower probability (error) responses and complementary points on the right of the graph represent the higher probability (correct) responses from the same experimental conditions. The five connected lines in each plot represent equally spaced quantiles below which 1/6, 1/3, 1/2 (the median), 2/3 and 5/6 of the RT distribution lies. The reader may notice some differences between the observed data in Figure 3 and the analogous plots in Ratcliff and Rouder (their Figure 5). These differences are due to

small methodological changes: Ratcliff and Rouder plot the mean for each condition, we plot five quantiles; we do not plot statistics for conditions with fewer than six observations (one per quantile bin); and we grouped the 33 conditions a little more finely than Ratcliff and Rouder. These differences are small, and affect only display (not model estimation). Interestingly, these methodological differences *do* appear to lessen the extent of the fast errors evident in the extremely low-N conditions of Ratcliff and Rouder's graphs.

In order to fit data of the type illustrated in Figure 3, the diffusion model requires between-trial variability in both start points and input strengths, in addition to the within-trial variability in information accumulation. Ratcliff and Rouder (1998) showed that with these extensions the diffusion model could fit the crossover pattern in their data by changing only the mean input, as required. In the next section we show that the BA model can also fit these data patterns without requiring within-trial variability. Brown (2002) reports an extensive set of simulations showing that the BA model can qualitatively accommodate all of the observed relationships between error and correct RT.

Fits to Information-Controlled Data

We fit the BA model to Ratcliff and Rouder's (1998) Experiment 1 data. These data were chosen as they contain data patterns that have proven challenging for choice RT models. A speed versus accuracy emphasis manipulation provided a further test of SAT in the BA model, with the tradeoff determined by task demand rather response deadlines. Three subjects (KR, JF and NH) each provided 7890 usable RTs over ten sessions of a brightness discrimination task. The task stimulus was a patch of grey in the centre of a

grey display. The centre patch had one of 33 levels of brightness, from all black pixels to all white pixels, resulting in conditions of widely different difficulty. The subjects' task was to judge the patch as bright or dark, and feedback was given based on whether the observed brightness level was the result of a sample from (unobserved) overlapping "bright" and "dark" distributions over the 33 levels. Different difficulty conditions (brightness levels) were randomly intermixed within blocks. On different blocks, subjects were instructed to emphasize either speed or accuracy in their responses. We used no censoring other than that specified in the raw output from Ratcliff and Rouder's experimental software, and we aggregated the 33 stimulus conditions into 13 groups for the purpose of display only. Further details of the experimental methods and preliminary data censoring can be found in Ratcliff and Rouder (pp.349-350).

Fits were obtained using quantile maximum probability estimation (QMPE; Heathcote & Brown, in press; Heathcote, Brown & Mewhort, 2002) with the quantiles illustrated in Figure 3. Many experimental conditions were empty of data (e.g., KR never responded "white" to a completely black stimulus in an accuracy-emphasis block). These conditions were handled naturally by QMPE and so were not removed or grouped with other conditions. Numerical integration over input and start point variability used the adaptive Rhomberg technique. Final model evaluations were cross checked using Monte-Carlo integration over both sources of variability, with 1,000,000 replicates for each of the 66 within-subject conditions.

The estimated values of the BA model parameters are shown in Table 1. The estimate of input noise is relatively small, in keeping with the physically identical stimuli in each condition. Task difficulty conditions were modelled by varying only the input

difference, I_d , as was the case for Ratcliff and Rouder's (1998) fits of the diffusion model, with the expected input sum kept constant at $I_S=0.0035$ for all subjects and all conditions. Ratcliff and Rouder found that input estimates for their diffusion fits were an approximately linear function of the probability that any given stimulus was drawn from the "bright" distribution (see their Figure 6). We found that estimates of the input to the BA model behaved similarly, as illustrated in Figure 4³. Hence, in agreement with the diffusion model, the BA model indicates that the subjects based their decisions on approximate probability matching.

We kept the value of the leakage parameter constant at $k=.0014$ across all subjects and all conditions. Apart from input strength values (which serve to parameterise the QQ lines), the fits reported here allowed a total of eight parameters to vary for each participant: four "structural" parameters (lateral inhibition strength, nondecision component time, and variability in start points and in input strengths) and two "location" parameters (accumulation start point and response criterion) each for speed and accuracy emphasis blocks. Eight parameters is slightly more than in the fits of the diffusion model (six freely estimated parameters) but smaller than that for Usher and McClelland's (2001) accumulator model (which requires a parameter for stochastic variability).

Figure 5 shows that the BA model provides an excellent fit to response probability data using the same set of inputs for both speed and accuracy conditions. Figure 3 shows the fits of the BA model to the RT data. RT distribution expands and the leading edge increases slightly as response probability decreases in the accuracy condition. In the speed condition RT distribution spreads most for middle probability values and the leading edge is relatively invariant. The BA model captures these trends in all cases (cf.

Ratcliff, Thapar & McKoon, 2003, Figures 4 and 5 for similar findings). A reviewer requested that we use a χ^2 statistic to globally quantify model fit, yielding: KR $\chi^2(df=513)=1180$; JF $\chi^2(df=546)=1778$ and NH $\chi^2(df=491)=980$. Although these values ranged from 1.8 to 2.9 times the .05 significance level, it is important to note that the χ^2 test is biased to reject a model when bins are not equi-probable (Rayner, 1985) or expected frequencies are small (Kijewski, Swensson & Judy, 1989). We aggregated neighbouring bin based on quantiles in order to maintain a minimum expected frequency of just one observation (cf. Ratcliff & Murdock, 1976), but bin frequencies still varied by almost two orders of magnitude, and more than 20% of bins had an expected frequency of less than five. Aggregation and the use of quantile boundaries also violate the χ^2 test's assumption of data independent boundaries, but given the widely varying ranges of RT data in speed and accuracy conditions fixed boundaries were not viable. Although Ratcliff and Rouder (1998) do not report χ^2 values for their fits, Ratcliff (2002) reports χ^2 values (Table 1) for fits of the diffusion model to data from three subjects in a very similar paradigm, but using only six brightness levels. Diffusion model χ^2 values ranged from 1.8 to 6.0 times the .05 significance level. Hence, the fits of the BA model are comparable to or better than the fits of the diffusion model.

Figure 3 allows the reader to estimate the gross shape of the observed and expected RT distributions by inferring from quantile values. A more accessible illustration of the BA model's account of RT distribution is shown in Figure 6, which plots twelve data histograms and the corresponding model predictions. These histograms represent three subjects (rows) always using data from brightness level 13 in the speed emphasis blocks (two left columns) and brightness level 16 in the accuracy condition (two right columns).

The histograms also show both correct responses and model fits (1st and 3rd columns) and error responses and model fits (2nd and 4th columns). It is clear that the BA model successfully models these distributions, even in the absence of within-trial variability. For comparison, analogous histograms for the diffusion model are shown in Ratcliff and Rouder (1998, Figure 3), although they do not show any histograms for fits to error data.

Only two parameters were allowed to vary between speed and accuracy conditions, the response criterion (C) and the mean starting point (x_0). This shows that the BA model passes a test of selective influence. That is, an experimental manipulation designed to influence only subjects' response caution was successfully modelled by changes only in parameters that could feasibly represent subjects' caution. Note that changing both response criterion and starting point in the BA model is analogous to Ratcliff and Rouder (1998), who allowed the response criterion to change, and assumed that the start point changed as a function of the response criterion. We found that independent adjustment of both start points and response criteria were necessary to fit Ratcliff and Rouder's data. Hence, the reduced complexity in the BA model inherent in assuming ballistic rather than stochastic accumulation comes at the cost of increased complexity in BA model's account of information-controlled speed accuracy tradeoff.

The BA model parameter estimates from data in the speed condition reveal that subjects adopted a simple and consistent strategy: the mean starting point was lowered below the baseline value (zero), and the response criterion was lowered to just above the upper boundary of the starting point distribution. The lower mean starting point and criterion produce faster accumulation, as the leakage works to return negative units' activations back to baseline (zero), and a node with negative activation excites other

nodes via inhibitory connections⁴. By placing the criterion near the upper boundary of starting activation speed is maximised, but sufficient time is still allowed for the input to overcome start-point differences on most trials. In the accuracy emphasis condition the mean starting point was much higher, and the criterion was set well above the upper bound of the start-point variability. The placement of the criterion allows substantial time for start-point noise to be overcome by the input, despite the stronger competition engendered by higher activation values.

Discussion

We have shown that a ballistic accumulation model is able to account for accuracy in time-controlled tasks and accuracy and RT distribution for both correct and error responses in information-controlled tasks. Few ballistic choice RT models have been proposed previously, and none with the same ability to account for data as the BA model. For example, in Luce's (1986) landmark book on RT research, only two ballistic models are discussed (Grice, 1968, 1972, and McClelland, 1979). Both models employ accumulation to a criterion and assume that RT variability is due to trial-to-trial fluctuations, in the criterion for Grice's model and in the input strength for McClelland's. The success of the BA model is due to the inclusion of between-trial variability in input strength and in the starting point of accumulation, as well as the use of a nonlinear accumulation process. As far as we are aware this is the first demonstration that a ballistic accumulation process can provide a general account of choice RT behaviour at the same level as the leading stochastic accumulation models (e.g., Bussemeyer & Townsend, 1992, 1993; Ratcliff & Rouder, 1998; Usher & McClelland, 2001).

Our ballistic model incorporates Usher and McClelland's (2001) neurally plausible dynamics, activation dependent passive leakage, self-excitation and competition.

However, without stochastic accumulation the dynamics differed substantially from Usher and McClelland's, requiring higher levels of competition to fit data. Lateral inhibition means that the BA model displays "winner-takes-all" behaviour, with one unit eventually suppressing the other. Winner-takes-all dynamics have been used extensively in deterministic neural network models of a range of cognitive phenomena (e.g., Grossberg, 1980, 1987; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Our results indicate that such models may be able to be modified to account for choice RT data without needing to incorporate stochastic accumulation, with its attendant computational cost and reduction in analytic tractability.

Massaro (1989) criticised a more recent deterministic competitive neural network model, the TRACE model of speech perception (McClelland & Elman, 1986), because it could not account for the quantitative form of the effect of context on perceptual choice. Typically context information has an additive effect on the linear relationship between input difference and the inverse cumulative normal transform of choice probability ($\Phi^{-1}(p)$). McClelland (1991, 1993) showed that either between-trial variability in inputs or within-trial variability in evidence accumulation allowed leaky competitive networks to produce the correct pattern of context effects when context provided a biasing input to the network. Figure 7 shows a plot of $\Phi^{-1}(p)$ versus input strength (i.e., brightness condition) for three different "biases" in the BA model with parameters corresponding the fit for Ratcliff and Rouder's (1998) subject KR in the accuracy condition. Context was manipulated through "biases" operationalised by associating the input values previously estimated for brightness level i (1..33) with brightness level $i-b$, where b is a bias value. As required, the BA model produces an approximately linear effect of context.

When competition is high, the BA model is also not subject to Ashby's (1982) criticism of McClelland's (1979) ballistic choice model, that the decision criterion is not exceeded, and so no response is made, for an appreciable proportion of trials. With reasonable parameter choices, this is not a problem for the fits of the BA model reported below. For two of the three subjects, the parameter governing the growth of the difference in activation between the units (D) is negative, so that even minute input differences are nonlinearly and infinitely amplified, ensuring a decision is made in finite time⁵. For the third subject (NH), the estimate of D is very close to zero (0.0001) so that the nonlinear amplification of input differences is, for all practical purposes, unbounded. For example, for the worst-case scenario – equal input strengths – the mean asymptotic activation for subject NH is 5.9, far above the maximum decision criterion of 0.8.

Models such as Usher and McClelland's (2001) and our BA model are possibly interpretable as models for neural firing rates. For the BA model, a monotonic transformation would be required to map negative activation levels to below-baseline firing rates (e.g., firing rate could be an exponential function of "activation"). The stochastic behaviour of spike trains in single neurones is often raised as an objection to the neural plausibility of deterministic accumulation models. Luce (1986) puts the case directly: "...given what is known physiologically, it is...unrealistic to suppose that sensory information is deterministic." (p.149). This view rests on equating sensory information with the behaviour of single neurones, whereas Shadlen and Newsome (1998) suggest that it is represented as rate codes in ensembles, or groups of ensembles, of 50-100 neurones. Smooth deterministic dynamics may provide an accurate approximation to such ensemble measures, or averages across ensembles⁶. It seems at

least plausible that the brain, with an apparent abundance of neurones, might make use of redundancy in order minimize the effects of accumulation noise, and so maximise fast and accurate decisions (see Anderson, 1991, for similar arguments in favour of dynamic deterministic RT models).

Ballistic models of averaged neural firing rates with only trial-to-trial variability have been proposed (Hanes & Schall, 1996; Reddi & Carpenter, 2000). However, Ratcliff (2001) points out that these models have no, or inadequate, mechanisms for producing errors, and so are incomplete. Recently, Ratcliff, Cherian and Segraves (2003) showed that in rhesus monkeys the diffusion model can simultaneously fit two-choice behavioural data and the difference between the averaged firing rates for neurones corresponding to each choice, including a delay in the onset of the difference for slow relative to fast responses. The latter phenomena requires a stochastic component as otherwise the linear dynamics of the diffusion model predict that the accumulation of evidence begins at the same time for slow and fast responses.

Unlike the ballistic models criticised by Ratcliff (2001), the BA model can account for errors, and nonlinear accumulation implies that it can in principle account for the delay. Delays in the emergence of a difference occur in the BA model when

$I_1 - I_2 \approx Dx_d(0)$ and competition is substantial, as was the case for the fits of the BA model reported here. Clearly, however, simultaneous fits to behavioural and neural data are required to determine whether it can fulfil this promise. As acknowledged by Ratcliff et al. (2003) the diffusion model's account of the neural data is also incomplete, as it models only the difference in firing rates, not the firing rates of neurones representing each choice. Competitive dynamics such as in the BA model, or Usher and McClelland's

(2001) model, provide one way of modelling the average firing rates of neurones representing each choice (see Wang, 2002, for more on the neural plausibility, or otherwise, of the diffusion model).

Throughout this article we have presented a strong position favouring purely ballistic accumulation. Our aim was to demonstrate that, contrary to modelling which has almost exclusively assumed stochastic accumulation, ballistic accumulation with between-trial variability can provide a viable approximation to choice behaviour as long as the accumulation process is sufficiently nonlinear. Our fits demonstrate that many of the roles fulfilled by stochastic variability can also be fulfilled by strong competition. Although a combination of competition and stochastic variability may be required for a full account of the behavioural and neural data, our results indicate that stochastic variability is not required by the existing behavioural data alone.

References

- Anderson, J. A. (1991). Why, having so many neurons, do we have so few thoughts? In W. E. Hockley & S. Lewandowsky (Eds.), Relating theory and data: Essays on human memory in honor of Bennet B. Murdock, Lawrence Erlbaum, Hillsdale, New Jersey.
- Ashby (1982). Deriving exact predictions for the cascade model. Psychological Review, **89**, 599-607.
- Audley, R.J. & Pike, A.R. (1965) Some alternative stochastic models of choice. British Journal of Mathematical and Statistical Psychology, **18**(2), 207-225
- Brown, S. (2002) Quantitative approaches to skill acquisition in choice RT. Unpublished doctoral thesis, University of Newcastle, Australia.
(<http://www.newcastle.edu.au/school/behav-sci/ncl/publications.html>)
- Busemeyer, J.R. & Townsend, J.T. (1992) Fundamental derivations from decision field theory. Mathematical Social Sciences, **23**, 255-282
- Busemeyer, J.R. & Townsend, J.T. (1993) Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review, **100** 432-459
- Carpenter, R. H. S. & Reddi, B. A. J. (2001). Reply, Nature Neuroscience, **3**, 827-830.
- Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing, **37**, 54-115.

- Gard, T.C. (1988) Introduction to stochastic differential equations. New York: Marcel Dekker Inc.
- Grice, G. R. (1968). Stimulus intensity and response evocation. Psychological Review, 75, 359-373.
- Grice, G. R. (1972) Application of a variable criterion model to auditory reaction time as a function of the type of catch trial. Perception and Psychophysics, 12, 103-107.
- Grossberg, S. (1980). How does a brain build a cognitive code? Psychological Review, 87, 1-51.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. Cognitive Science, 11, 23-63.
- Hanes, D. P. & Schall, J. D. (1996). Neural control of voluntary movement initiation. Science, 274, 427-430.
- Heathcote, A. (1998). Neuromorphic models of response time, Australian Journal of Psychology, 50, 157-164.
- Heathcote, A. (2003). BAM! The nonlinear dynamics of choice in a ballistic accumulation model. Paper presented at the 11th Australian Mathematical Psychology Conference, Sydney, Australia.
- Heathcote, A., & Brown, S. (2000). The law of practice and localist neural network models. In Page, M., Connectionist modelling in psychology: A localist Manifesto, Behavioral and Brain Sciences, 23, 479-480.

- Heathcote, A. & Brown, S. (2002). SEEXC: A model of response time in skill acquisition, Noetica, Cognitive Science Conference 2002 Papers, <http://www2.psy.uq.edu.au/CogPsych/Noetica>.
- Heathcote, A., Brown, S. & Mewhort, D.J.K. (2000) Repealing the power law: The case for an exponential law of practice. Psychonomic Bulletin and Review, 7, 185-207.
- Heathcote, A., Brown, S. & Mewhort, D.J.K. (2002). Quantile Maximum Likelihood Estimation of Response Time Distributions. Psychonomic Bulletin and Review, 9, 394-401.
- Kijewski, M.F., Swensson, R.G. & Judy, P.F. (1989) Analysis of rating data from multiple-alternative tasks. Journal of Mathematical Psychology, 33, 428-451
- Laming, D.R. (1968) Information theory of choice-reaction times. New York: Academic Press.
- Luce, R. D. (1986). Response Times. New York: Oxford University Press.
- Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. Cognitive Psychology, 21, 398-421.
- Mazurek, M.E. & Shadlen, M.N. (2002) Limits to the temporal fidelity of cortical spike rate signals. Nature Neuroscience, 5(5) 463-470
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review, 86, 287-330.
- McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception, Cognitive Psychology, 23, 1-44.

McClelland, J. L. (1993). Towards a theory of information processing in graded, random, interactive networks. In D. E. Meyer & S. Kornblum (Eds.), Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neurosciences (pp. 655-688). Cambridge, MA: MIT Press.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.

McElree, B. & Doshier, B.A. (1989). Serial position and set size in short-term memory: the time course of recognition. Journal of Experimental Psychology, 118, 346-373.

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

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. Psychological Review, 95, 238-255.

Ratcliff, R. (2001). Putting noise into neurophysiological models of simple decision making. Nature Neuroscience, 4, 336.

Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a brightness discrimination task. Psychonomic Bulletin and Review, 9, 278-291.

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

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. Psychological Review, 83, 190-214.

Ratcliff, R. & Rouder, J. N. (1998). Modeling response times for two-choice decisions. Psychological Science, 9, 347-356.

- Ratcliff, R. & Smith, P.L. (submitted). A comparison of sequential sampling models for two-choice reaction time.
- Ratcliff, R., Thapar, A. & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. Perception & Psychophysics, 65, 523-525.
- Ratcliff, R. & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. Psychonomic Bulletin & Review, 9, 438-481.
- Ratcliff, R., Van Zandt, T. & McKoon, G. (1999) Comparing connectionist and diffusion models of reaction time. Psychological Review, 106(2) 261-300.
- Rayner, J.C.W. (1985) Bias of the pearson chi-squared test. Journal of Statistical Computation and Simulation, 21, 329-331.**
- Reddi, B. A. J. & Carpenter, R. H. S. (2000). The influence of urgency on decision time, Nature Neuroscience, 3, 827-830.
- Roe, R., Busemeyer, J. R., & Townsend, J. T. (2001) Multi-alternative decision field theory: A dynamic connectionist model of decision-making. Psychological Review, 108, 370-392.
- Rouder, J. N., (1996). Premature Sampling in Random Walks. The Journal of Mathematical Psychology, 40, 287-296.
- Shadlen, M. N. & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation and information coding. Journal of Neuroscience, 18, 3870-3896.

- Smith, P.L., Ratcliff, R. & Wolfgang, B.J. (2003) Attention and the representation of visual stimuli in psychophysical decisions. *Paper presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver, Canada.*
- Wang, X.-J. (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron, **36** 955-968
- Usher, M. & McClelland, J. L. (2001) On the time course of perceptual choice: the leaky competing accumulator model. Psychological Review, **108**, 550- 592.

Footnotes

¹Brown (2002) examined leaky competitive accumulator models with all possible single, pairwise and triple combinations of four different sources of variability: within-trial accumulation noise, and between trial variability in start points, response criteria or input strengths. He concluded that the combination we employ in the BA model led to the simplest feasible model.

² Brown (2002) and Heathcote and Brown (2000, 2002) consider a similar model in which the self-excitation parameter is kept separate from leakage, and used to explain the effects of practice on RT (Heathcote, Brown & Mewhort, 2000).

³Note Ratcliff and Rouder's (1998) Figure 6 has six lines as in their initial set of fits they allowed different input values for speed and accuracy conditions. They then constrained the input values to be equal for both speed and accuracy in order to demonstrate selective influence. Our Figure 4 only has three lines as we used the same input values for both speed and accuracy conditions in all fits, in line with the selective influence constraint.

⁴ See Roe, Busemeyer and Townsend (2001) for a similar use of "negative inhibition", used to explain the attraction effect in consumer preference. Note also that we have obtained very good fits (not reported here) using the linear-threshold version of our model, in which units with negative activation do *not* transmit any effects to other units.

⁵An exception occurs when the input difference is exactly balanced by an opposite start-point difference, which occurs when $I_1 - I_2 = Dx_d(0)$, where $x_d(t) = x_1(t) - x_2(t)$. In this case, the activation difference remains constant, and activations approach $I_s/2S + x_d(0)/2$ and $I_s/2S - x_d(0)/2$ for units one and two respectively. As long as long as the criterion is less than the larger of the two values, activation will eventually exceed the decision criterion. If the input difference exactly equals zero and the starting points for both units are equal a choice will not be made as both units cross the criterion at the same time. However, in the continuous case this event has a probability measure of zero. Even in our finite precision fits the probability of this event was smaller than 10^{-18} (we used 64 bit real values everywhere).

⁶Recent work suggests that weak correlations between the firing rates of adjacent neurones may limit the smoothing effects of averaging, particularly when the signal fluctuates at a high frequency (Mazurek & Shadlen, 2002).

Acknowledgements

Thanks to Jeff Rouder for supplying the data from Ratcliff and Rouder (1998). We acknowledge support from an Australian Research Council Discovery Project Grant to S. Andrews and A. Heathcote. Thanks to Jerry Busemeyer and Jeff Rouder for providing helpful feedback on earlier drafts, and to two anonymous reviewers.

Table

Table 1. Parameter estimates for the BA model fits to Ratcliff and Rouder's (1998, Experiment 1) data. Parameters are non-decision time (t_0), and competition (β), the trial-to-trial standard deviation in inputs (σ_I), the half width of rectangular start point noise (Δx_0), the response criterion (C) and mean activation starting point (x_0). We used a constant expected input sum ($I_S=.0035$) and rate of passive leakage ($k=.0014$).

		Subject		
		KR	JF	NH
t_0		220	240	231
β		.0018	.004	.0013
σ_I		.0008	.0015	.00074
Constant Across All Conditions	Δx_0	.25	.35	.24
Speed Emphasis	C	.17	.07	.25
	x_0	-.1	-.3	-.09
Accuracy Emphasis	C	0.98	1.13	.80
	x_0	.65	.5	.37

Figure Captions

Figure 1. Sample paths from Usher and McClelland's (2001) model (left panel) and from the BA model (right panel). The abscissa represents accumulation time, and the ordinate represents activation.

Figure 2. Symbols represent 21 values of the BA model SAT functions (d' vs. accumulation time), using parameters estimated from Ratcliff and Rouder's (1998) three subjects (JF, KR and NH) in their Experiment 1. Lines show best-fitting shifted exponential functions.

Figure 3. QP plots from Ratcliff and Rouder's (1998) Experiment 1. Dotted lines show observed RT distribution quantiles in milliseconds with numbers 1..5 representing the quantiles 1/6, 1/3, 1/2, 2/3 and 5/6 plotted against response probability. BA model fits are shown as solid lines. Points on the right of the plots represent both very common "dark" responses (e.g., to black stimuli) and very common "bright" responses (e.g., to white stimuli). Both data and theoretical predictions are grouped over neighbouring p -values, as in Ratcliff and Rouder. Thus, the expected RT quantiles are not smooth as they include different numbers of "bright" and "dark" responses, which are not exactly equivalent. Quantile estimates for data conditions with fewer than six observations were omitted.

Figure 4. Thick line represents the probability (p) that a stimulus from a given brightness (x-axis) condition was drawn from the "high" distribution. Data points represent the best linear fit of the estimated BA model input values (I) from each subject to those probabilities (estimated linear transformations were: JF $p=0.069+212I$; KR $p=0.015+252I$; and NH $p=-0.025+276I$).

Figure 5. Ratcliff and Rouder's (1998) Experiment 1 observed probabilities for "dark" responses plotted against response probabilities predicted by the BA model. Solid $y=x$ line shows a perfect fit, error bars are based on binomial approximations. Graphs on the left are for speed-emphasis blocks; those on the right are for accuracy emphasis blocks.

Figure 6. Observed histograms and expected densities for brightness level 13 from speed emphasis blocks and brightness level 16 for accuracy blocks. Each row represents a different subject's data (top to bottom: KR, NH, JF). The left-hand two columns represent speed-emphasis data (correct responses on left, errors on right). The right-hand two columns represent accuracy-emphasis data (correct on left, error data on right).

Figure 7. Context effects in the BA model. The inverse cumulative normal transform of choice probability, $\Phi^{-1}(p)$, is plotted against input strength (i.e., brightness condition) for three different bias values in the BA model. The parameters used were those estimated from the data of Ratcliff and Rouder's (1998) subject KR in the accuracy condition. Expected probabilities are shown over a wide range - from a probability of responding "bright" of only 0.025% up to 99.975% - and are evidently linear and the effect of context additive across most of this range.

Figure 1

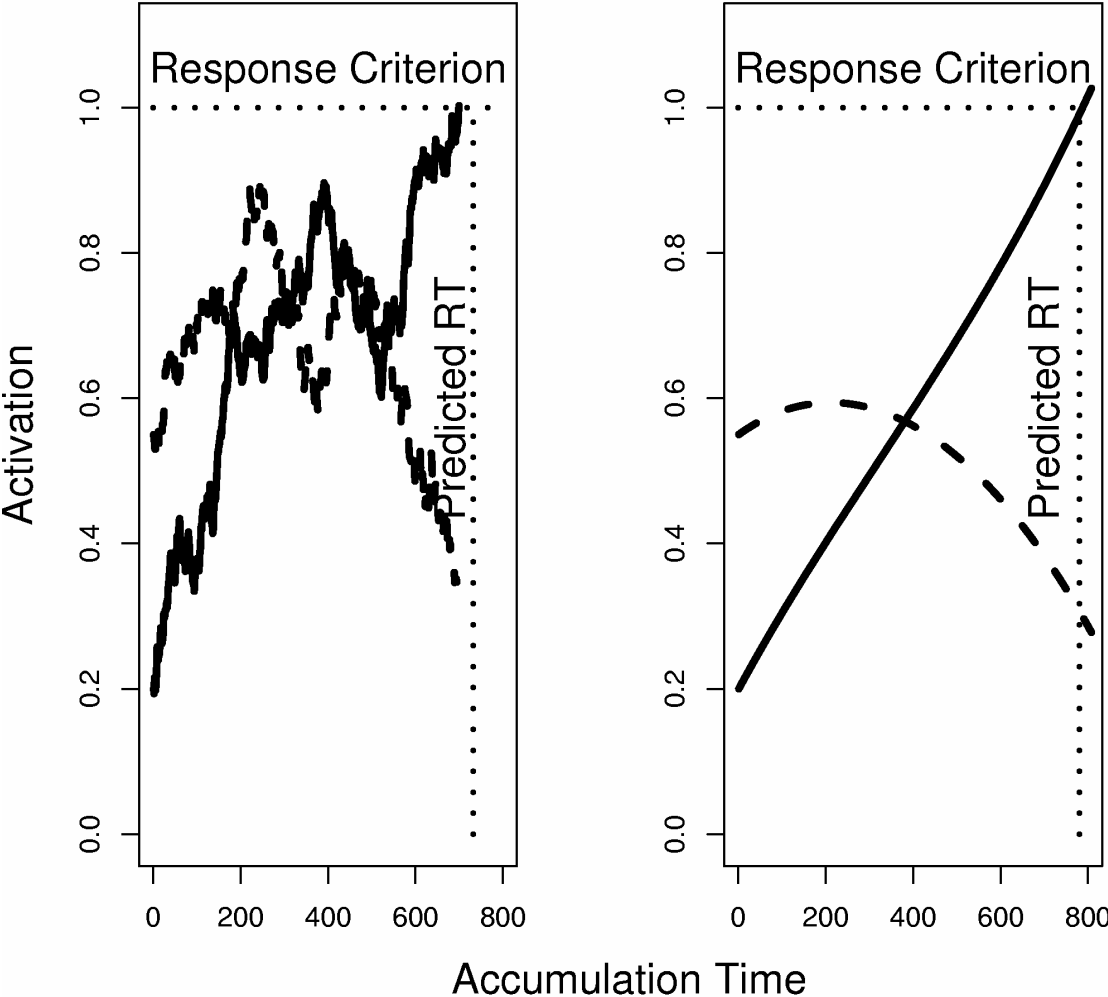


Figure 2

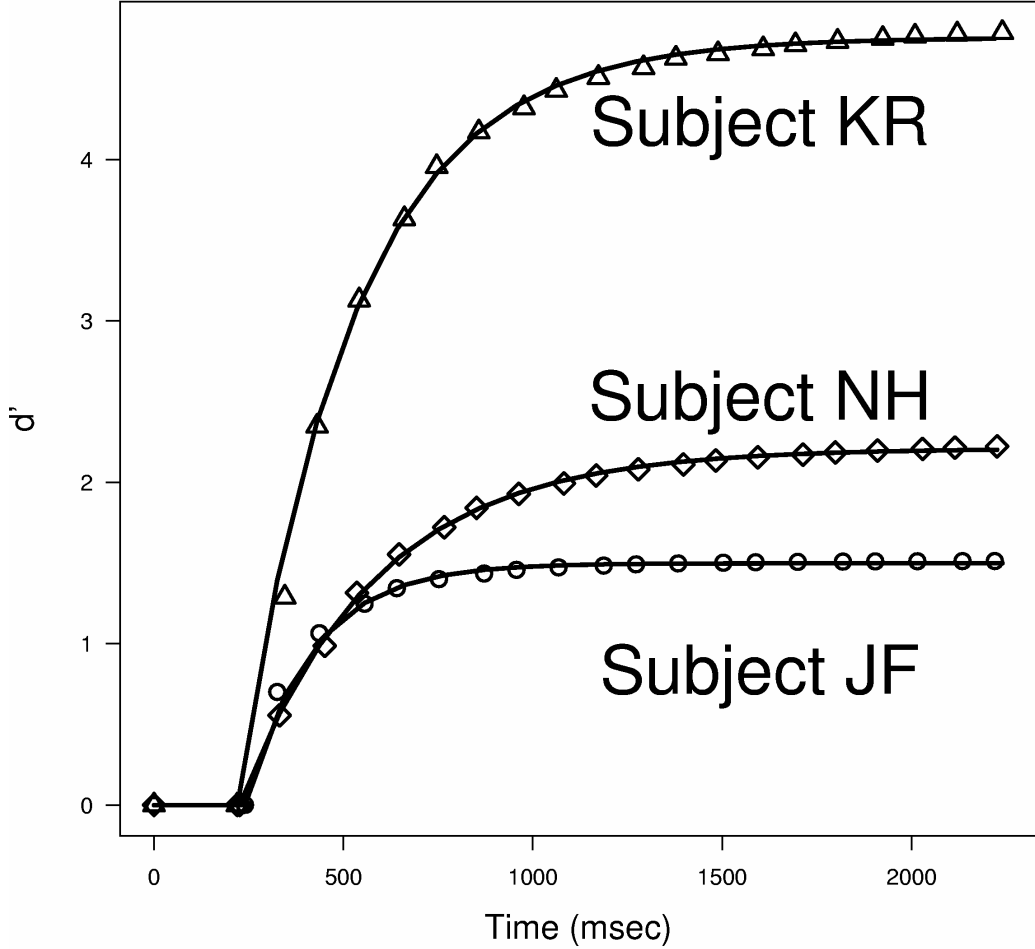


Figure 3, panel A

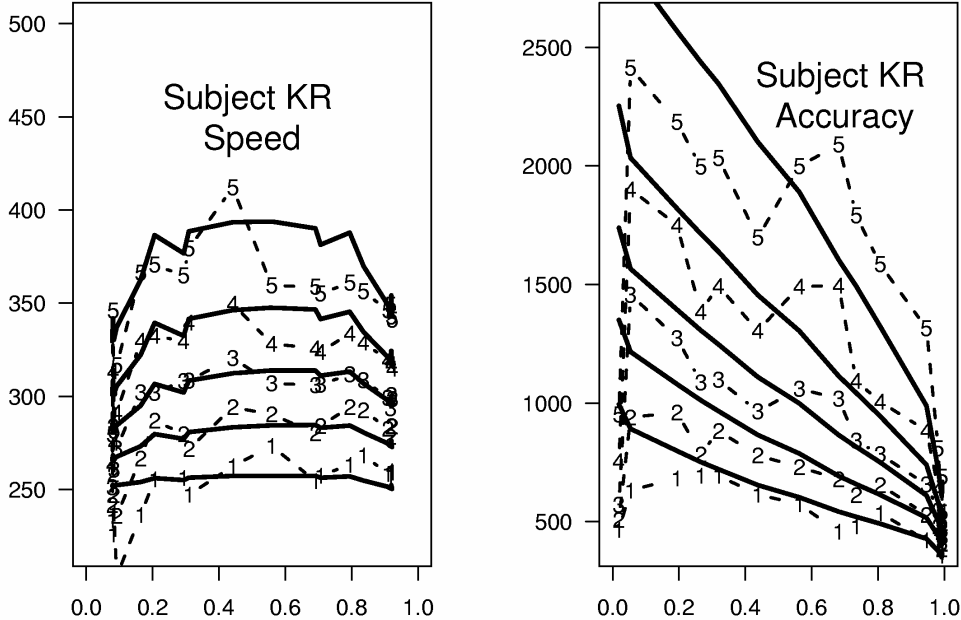


Figure 3, Panel B

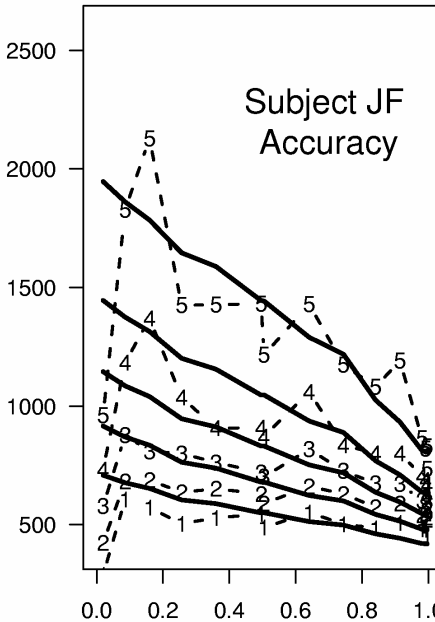
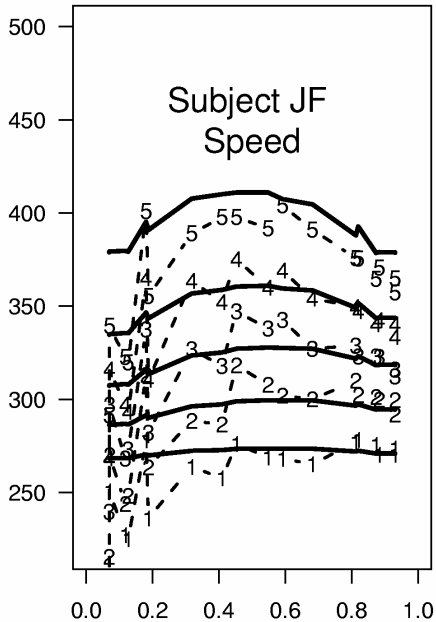


Figure 3, Panel C

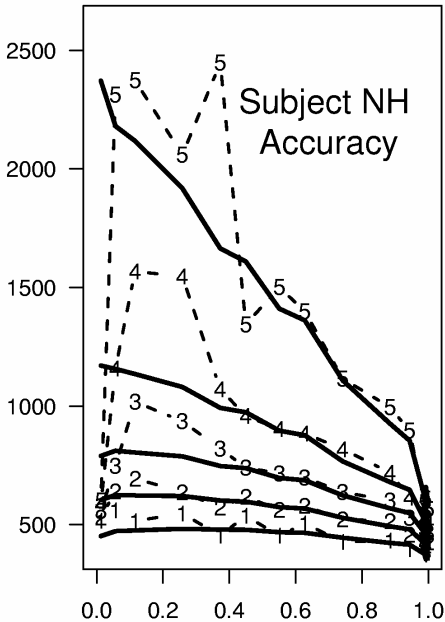
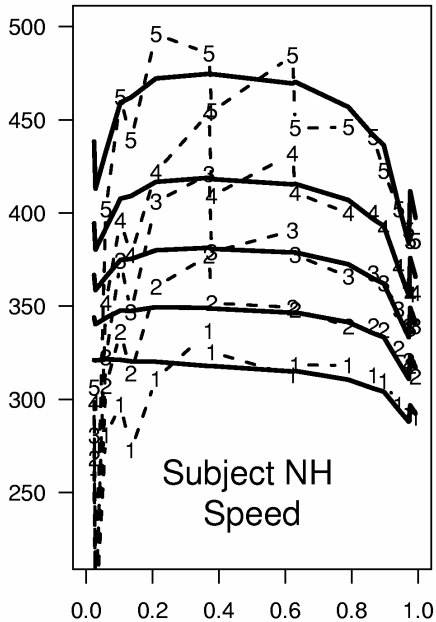


Figure 4

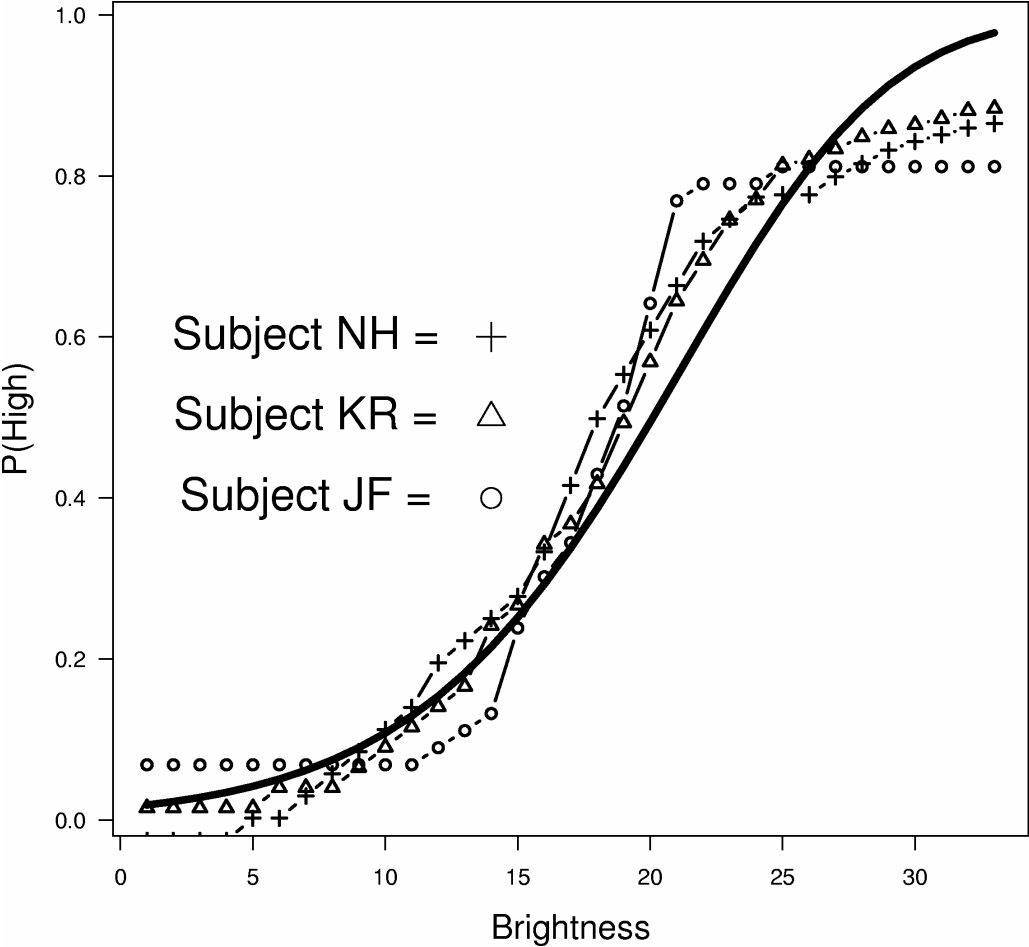


Figure 5

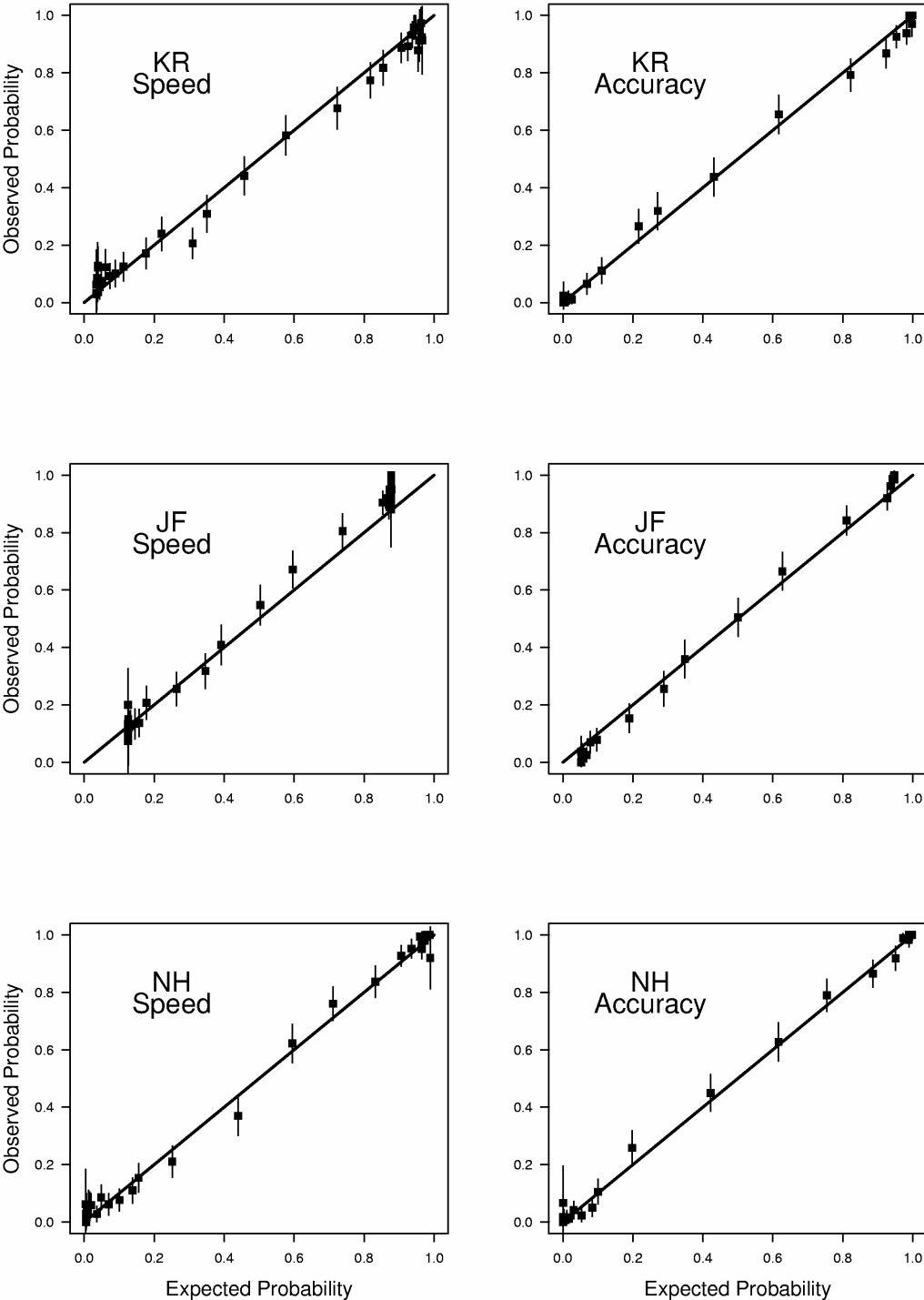


Figure 6

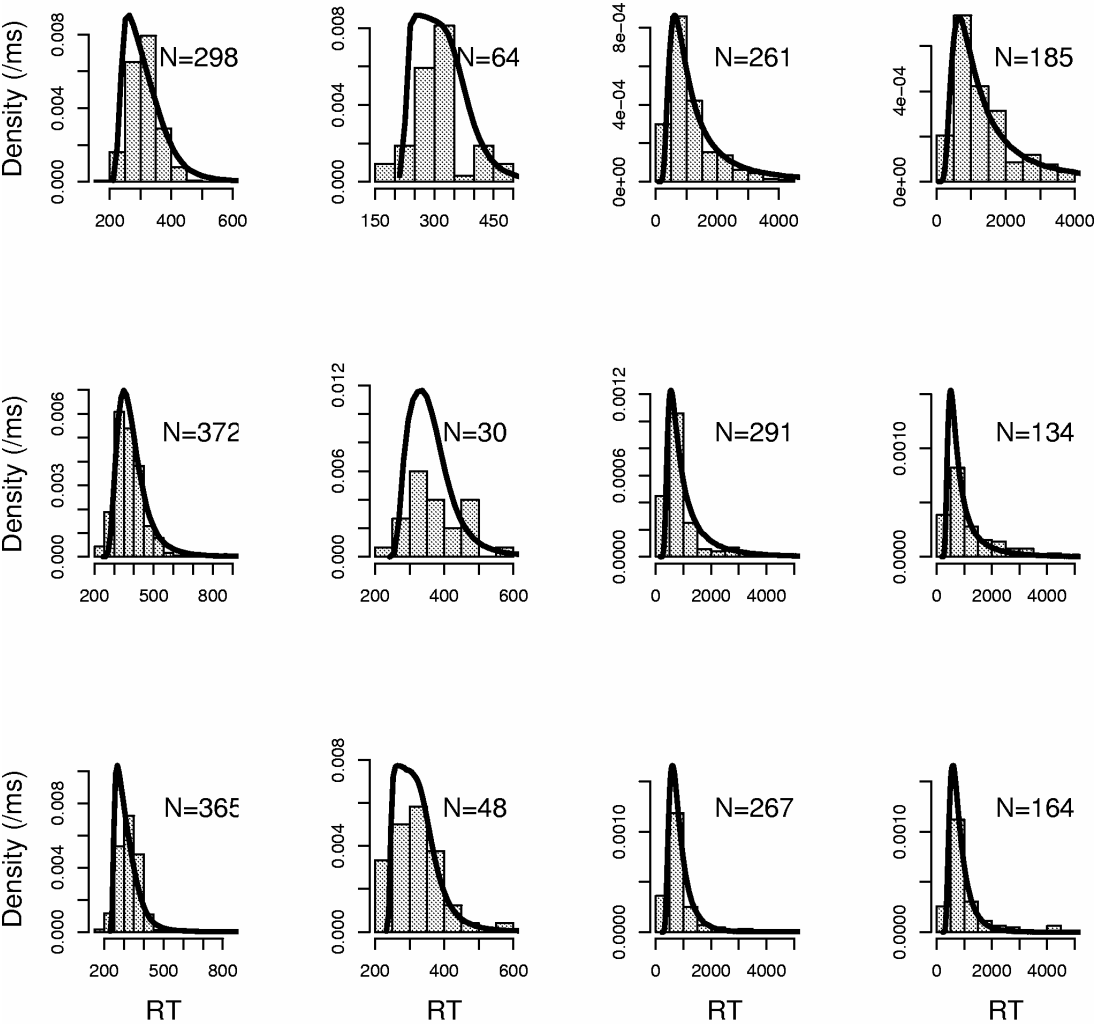


Figure 7

