Published in final edited form as:
Neural Comput. 2009 June ; 21(6): 1520–1553. doi:10.1162/neco.2009.03-07-495.

Dynamical analysis of Bayesian inference models for the Eriksen task

Yuan Sophie Liu[1], **Angela Yu**[2], and **Philip Holmes**[3]

[1]Department of Physics, Princeton University, Princeton, NJ 08544, U.S.A.

[2]Center for the Study of Brain, Mind, and Behavior, Princeton University, Princeton, NJ 08544, U.S.A.

[3]Department of Mechanical and Aerospace Engineering and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, U.S.A.

Abstract

The Eriksen task is a classical paradigm that explores the effects of competing sensory inputs on response tendencies, and the nature of selective attention in controlling these processes. In this task, conflicting flanker stimuli interfere with the processing of a central target, especially on short reaction-time trials. This task has been modeled by neural networks and more recently by a normative Bayesian account. Here, we analyze the dynamics of the Bayesian models, which are nonlinear, coupled discrete-time dynamical systems, by considering simplified, approximate systems that are linear and decoupled. Analytical solutions of these allow us to describe how posterior probabilities and psychometric functions depend upon model parameters. We compare our results with numerical simulations of the original models and derive fits to experimental data, showing that agreements are rather good. We also investigate continuum limits of these simplified dynamical systems, and demonstrate that Bayesian updating is closely related to a drift-diffusion process, whose implementation in neural network models has been extensively studied. This provides insight on how neural substrates can implement Bayesian computations.

Keywords

Bayesian inference; decoupling; drift-diffusion model; dynamical system; Eriksen task; linearization

1 Introduction

The psychological [Laming, 1968, Ratcliff, 1978, Ratcliff et al., 1999] and neural bases of decision making [Platt and Glimcher, 2001, Schall, 2001, Gold and Shadlen, 2001] have been widely studied, particularly in constrained situations such as the two-alternative forced-choice (2AFC) task. In 2AFC, subjects are required to discriminate a stimulus and to give one of two permissible responses. The sequential probability ratio test (SPRT) is optimal for 2AFC tasks, whether the objective is to minimize the mean reaction time (RT) for a desired accuracy level [Wald and Wolfowitz, 1948], or to minimize a linear cost function in accuracy and detection delay under the Bayesian formulation [Liu and Blostein, 1992]. The SPRT compares the relative likelihoods of noisy inputs given two possible hypotheses, and reaches a decision when the cumulative evidence for one of them exceeds a fixed threshold. Performance on 2AFC tasks seems broadly consistent with the SPRT [Ratcliff and Smith, 2004], and there is evidence

yuanliu, ajyu@princeton.edu, pholmes@math.princeton.edu.

that competing neural populations sub-serving decision-making may implement a strategy close to the SPRT [Gold and Shadlen, 2002, Schall, 2001, Gold and Shadlen, 2001, Shadlen and Newsome, 2001, Roitman and Shadlen, 2002] and [Schall et al., 2002]. Moreover, the continuum limit of SPRT is an analytically-tractable *drift-diffusion model* (DDM) [Holmes et al., 2005], which yields explicit expressions for error rates and reaction times that can be used to investigate reward-rate maximization in 2AFC [Bogacz et al., 2006], and it has been shown that various neural network models of decision-making [Cohen et al., 1990, Cohen et al., 1992, Usher and McClelland, 2001] can be reduced to variants of the DDM [Bogacz et al., 2006].

The Eriksen flanker task [Eriksen and Eriksen, 1974] is an extension of the classical 2AFC task in which the decision is complicated by potentially-conflicting distractor inputs. Subjects are required to discriminate a central target stimulus (*e.g.* the letter H or S) flanked by distractors. Flankers can either be *compatible* with the central target (*e.g.* HHHHH) or *incompatible* (*e.g.* HHSHH). Subjects display a *compatibility effect*, being typically slower and less accurate on incompatible than compatible trials [Eriksen and Eriksen, 1974]. Furthermore, subjects perform at worse than chance level for short RT's for incompatible trials only. This "dip" in accuracy implies that flanker interference is particularly potent shortly after stimulus presentation. Fig. 1 shows data from two instances of a deadlined Eriksen task. While specific details of reaction time distributions and relationships between accuracy and reaction time differ between the two studies, the basic compatibility effect and the dip in accuracy on incompatible trials are prominent in both.

Since the Eriksen task extends the standard 2AFC task, we suspect that optimal policy in this case is similar to the SPRT. In this vein, [Yu et al., 2007] modeled the computations underlying the Eriksen task as iterative Bayesian updating, with the decision being made (and the trial terminated) when the cumulative posterior for one of the two possible target stimuli exceeds a fixed threshold. It was also proposed that the apparent *suboptimality* in performance can be explained by either an incorrect prior on the relative frequency of compatible and incompatible trials (*compatibility bias model*), or by inherent spatial overlap of visual processing neurons (*spatial uncertainty model*) [Yu et al., 2007]. Here we reduce the Bayesian models to simpler dynamical systems and study them analytically and numerically. While the simpler models closely approximate the original ones in dynamics and perfomance, their analytical tractability yields explicit expressions for the dependence of inferential and psychometric quantities on model parameters. We discuss the relationship between exact Bayesian inference and drift-diffusion processes, emphasising the link that this establishes between Bayesian updating and the neural substrates that may execute it. Our analysis also reveals the formal similarity of computations underlying the compatibility bias and spatial uncertainty models, which were motivated by disparate experimental literature and were formulated differently within the Bayesian framework.

The paper is organized as follows. After reviewing the Bayesian inference models in Section 2, in Section 3 we derive and analyze the simplified models: uncoupled, linear discrete dynamical systems. From these we derive explicit criteria on parameters that predict the dip in accuracy for incompatible trials, and we compare accuracies and RT distributions generated by the full and simplified models. In Section 4 we show that the DDM is a continuum limit of the simplified models, and from this derive analytical predictions for mean posterior probabilities. We also compute accuracy vs. time curves and reaction time distributions under an approximation that violates the first passage threshold crossing criterion adopted in [Yu et al., 2007], but permits explicit analysis, and we provide direct comparisons between behavioral data and predicitions of the full and approximate compatibility bias models. Section 5 contains a summary and discussion.

2 A Bayesian framework for the Eriksen task

We briefly review the compatibility bias and spatial uncertainty inference models for the Eriksen task proposed by [Yu et al., 2007]. The *generative process* common to both models consists of the prior probability distribution over trial type ($M = 1$ if compatible, $M = 2$ if incompatible), and the stochastic relationship between the trial type M and the stimuli $\mathbf{s}$, and between the stimuli and the noisy inputs into the visual system $\mathbf{x}$. For simplicity, it was assumed that there are three stimuli, $\mathbf{s} \triangleq \{s_1, s_2, s_3\}$, for "left", "center", "right", respectively; and each one of three neural units or populations $\mathbf{x} \triangleq \{x_1, x_2, x_3\}$ responds to one stimulus. Here the pairs of left and right flankers are combined in s_1 and s_3 respectively, and we assume that all three inputs contain independent noise, both among the units/populations, and over time. Using integers $s_i = \pm 1$ to denote S and H, and $M = 1, 2$ to denote compatible and incompatible trials respectively, we may formally describe the process as:

$$\beta \triangleq P(M=1) \in [0, 1] \tag{1}$$

$$P(\mathbf{s}|M=1) = \begin{cases} 0.5 & s_1=s_2=s_3=-1 \quad \text{(HHHHH)} \\ 0.5 & s_1=s_2=s_3=+1 \quad \text{(SSSSS)} \end{cases}, \tag{2}$$

$$P(\mathbf{s}|M=2) = \begin{cases} 0.5 & s_1=s_3=+1, s_2=-1 \quad \text{(SSHSS)} \\ 0.5 & s_1=s_3=-1, s_2=+1 \quad \text{(HHSHH)} \end{cases}, \tag{3}$$

$$p(\mathbf{x}_t|\mathbf{s}) = p(x_1(t)|\mathbf{s})\, p(x_2(t)|\mathbf{s})\, p(x_3(t)|\mathbf{s}), \tag{4}$$

$$p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t|\mathbf{s}) = p(\mathbf{x}_1|\mathbf{s})\, p(\mathbf{x}_2|\mathbf{s}) \ldots p(\mathbf{x}_t|\mathbf{s}). \tag{5}$$

For the compatibility bias model, the prior probability β for compatible trials is assumed to be higher than the "true" value 0.5, and the inputs are taken to be normally distributed as a function of their respective stimuli and independent of neighboring stimuli:

$$p(x_i(t)|\mathbf{s}) = p(x_i(t)|s_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x_i - s_i)^2}{2\sigma^2}\right], \tag{6}$$

i.e., at each step t the $x_i(t)$ are independently drawn from normal distributions with means s_i and standard deviations σ. We denote this procedure below by $x_i(t) \sim \mathcal{N}(s_i, \sigma^2)$.

In the spatial uncertainty model, the correct prior $\beta = 0.5$ is assumed, but the inputs are corrupted by their neighbors according to:

$$p(\mathbf{x}_t|\mathbf{s}) = p(x_1(t)|s_1, s_2)\, p(x_2(t)|s_1, s_2, s_3)\, p(x_3(t)|s_2, s_3), \text{ where} \tag{7}$$

$$\begin{aligned} x_1(t) &\sim \mathcal{N}\left(a_1 s_1 + a_2 s_2, \sigma_1^2 + \sigma_2^2\right), \\ x_2(t) &\sim \mathcal{N}\left(a_1 s_2 + a_2 s_1 + a_2 s_3, \sigma_1^2 + 2\sigma_2^2\right), \\ x_3(t) &\sim \mathcal{N}\left(a_1 s_3 + a_2 s_2, \sigma_1^2 + \sigma_2^2\right), \end{aligned} \tag{8}$$

where a_1, σ_1 denote influence from the primary stimulus, and a_2 and σ_2 that from the flankers.

Define $z_{i,j}^t \triangleq P(s_2=i, M=j|\mathbf{X}_t)$ for the posterior probabilities, and $l_{i,j}^t \triangleq p(\mathbf{x}_t|s_2=i, M=j)$ for the likelihood functions, where $i \in \{-1, +1\}$, $j \in \{1, 2\}$. Based on Bayes' Rule, this yields our inference model: four discrete-time dynamical equations, coupled through normalization:

$$z_{i,j}^t = \frac{l_{i,j}^t z_{i,j}^{t-1}}{\sum_{k,l} l_{k,l}^t, z_{k,l}^{t-1}}, \tag{9}$$

with initial conditions

$$z_{i,j}^0 = \begin{cases} \frac{\beta}{2}, & j=1, \forall i; \\ \frac{(1-\beta)}{2}, & j=2, \forall i. \end{cases} \tag{10}$$

To make a decision based on the accumulating inputs, we compare the cumulative marginal posterior probability,

$$P(s_2=i|\mathbf{X}_t) = z_{i,1}^t + z_{i,2}^t, \tag{11}$$

against a decision threshold q, a policy closely related to the SPRT [Wald, 1947]. As soon as $P(s_2=i|\mathbf{X}_t)$ exceeds q for i=-1 or i=+1, the system chooses the corresponding response (H or S) and terminates observations for the current trial. The computation for the marginal posterior probability over compatibility is analogous: $P(M=j|\mathbf{X}_t) = z_{-1,j}^t + z_{1,j}^t$.

Examples of accuracies and RTs thus predicted are shown in Fig. 4, below. For these and other calculations, unless otherwise noted, we adopt the parameter values used in [Yu et al., 2007]: σ = 9 for the compatibility bias model and σ_1 = 7, σ_2 = 5, a_1 = 0.7, a_2 = 0.3 for the spatial uncertainty model, and q = 0.9 for both.

3 Linearization and parametric dependence

It was shown in [Yu et al., 2007] that certain choices of parameters allow both the compatibility bias and spatial uncertainty models to capture key properties of the behavioral data in Fig. 1 (see Fig. 4 below). Here we derive general constraints on the parameters in each model that allow them to reproduce the behavioral data: σ for the compatibility bias model, a_1, a_2, σ_1, and σ_2 for the spatial uncertainty model; and n, the number of distractors. While we cannot analyze the complex relationship between accuracy and reaction time directly, we wish to at least constrain parameters so that the mean posterior probability for s_2=1 (the correct answer) dips below that for s_2=-1 after one or a few timesteps of observations. Although the relative probability of a correct response at time t depends not just on the mean but also on higher-order moments, such an analysis would illuminate the magnitude and range of the effective parameters. Unfortunately, even this partial analysis is difficult for the original Bayesian

model, as $P(s_2|\mathbf{X}_t)$ involves the summation of two exponential functions of the inputs, as in Eq. (11), and there is no obvious way to derive the expectation of $P(s_2|\mathbf{X}_t)$ as an explicit function of the parameters that specify the generation of the inputs **x**.

Due to such computational intractability, we instead work with a linearized approximation to the exact posterior update rule of Eq. (9). We will motivate and describe the approximations for the two Bayesian models, and demonstrate via simulations that the parametric constraints derived from this approximate scheme provide useful bounds for the original Bayesian models.

3.1 The compatibility bias model

Given our assumption of independent, normally-distributed inputs (Eqs. (4) and (6)), we have

$$p(\mathbf{x}_t|\mathbf{s}) = \left\{\frac{\exp\left[\frac{-(x_1-s_1)^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}}\right\}\left\{\frac{\exp\left[\frac{-(x_2-s_2)^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}}\right\}\left\{\frac{\exp\left[\frac{-(x_3-s_3)^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}}\right\}, \tag{12}$$

where each s_i can take on the value ±1. We now derive an approximation to Eq. (12) that is linear in the $x_i(t)$'s. Defining

$$\Theta_k \triangleq \frac{\exp\left[\frac{-(x_k-1)^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}} + \frac{\exp\left[\frac{-(x_k+1)^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}},$$

the likelihood function for s_2=1 and M=1 (*i.e.* s_1=s_2=s_3=1) can be approximated as follows:

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{s}) &= \Theta_1\Theta_2\Theta_3\left\{\frac{\exp\left[\frac{-(x_1-1)^2}{2\sigma^2}\right]}{\Theta_1\sqrt{2\pi\sigma^2}}\frac{\exp\left[\frac{-(x_2-1)^2}{2\sigma^2}\right]}{\Theta_2\sqrt{2\pi\sigma^2}}\frac{\exp\left[\frac{-(x_3-1)^2}{2\sigma^2}\right]}{\Theta_3\sqrt{2\pi\sigma^2}}\right\} \\ &= \Theta_1\Theta_2\Theta_3\left[\frac{1}{1+\exp\left(-\frac{2x_1}{\sigma^2}\right)}\frac{1}{1+\exp\left(-\frac{2x_2}{\sigma^2}\right)}\frac{1}{1+\exp\left(-\frac{2x_3}{\sigma^2}\right)}\right] \\ &= \frac{\Theta_1\Theta_2\Theta_3}{8}\left[1+\frac{x_1}{\sigma^2}\right]\left[1+\frac{x_2}{\sigma^2}\right]\left[1+\frac{x_3}{\sigma^2}\right]+O\left(x_k^2/\sigma^4\right) \\ &= \frac{\Theta_1\Theta_2\Theta_3}{8}\left[1+\frac{x_1+x_2+x_3}{\sigma^2}\right]+O\left(x_k^2/\sigma^4\right) \\ &\approx \frac{\Theta_1\Theta_2\Theta_3}{8}\left[1+\frac{x_1+x_2+x_3}{\sigma^2}\right]. \end{aligned} \tag{13}$$

The first step uses the fact that quadratic and constant terms cancel in the ratios, the next two rely on Taylor series expansion of the exponential terms and the binomial series approximation:

$$\left[1+\exp\left(-\frac{2x_k}{\sigma^2}\right)\right]^{-1} \approx \left[2\left(1-\frac{x_k}{\sigma^2}\right)\right]^{-1} \approx \frac{1}{2}\left[1+\frac{x_k}{\sigma^2}\right],$$

and the approximation is justified by the fact that $x_k(t) \in [-1-2\sigma, 1+2\sigma]$ with > 99% probability, if we can assume that $\sigma \gg 1$. This latter assumption is reasonable since we are modeling the time-scale at which on average many time steps of inputs are needed to perform the discrimination.

Generalizing the approximation (13) to the other three cases and using the four resulting expressions in Eq. (9), we obtain the following approximate update rules:

$$z_{i,j}^t \approx \frac{1}{D_t}\begin{cases}\left(1+\frac{x_1(t)+x_2(t)+x_3(t)}{\sigma^2}\right)z_{+1,1}^{t-1} & s_2=+1, M=1,\\ \left(1-\frac{x_1(t)+x_2(t)+x_3(t)}{\sigma^2}\right)z_{-1,1}^{t-1} & s_2=-1, M=1,\\ \left(1-\frac{x_1(t)-x_2(t)+x_3(t)}{\sigma^2}\right)z_{+1,2}^{t-1} & s_2=+1, M=2,\\ \left(1+\frac{x_1(t)-x_2(t)+x_3(t)}{\sigma^2}\right)z_{-1,2}^{t-1} & s_2=-1, M=2,\end{cases} \tag{14}$$

in which the denominator

$$D_t=\left(1+\frac{x_1(t)+x_2(t)+x_3(t)}{\sigma^2}\right)z_{+1,1}^{t-1}+\ldots+\left(1+\frac{x_1(t)-x_2(t)+x_3(t)}{\sigma^2}\right)z_{-1,2}^{t-1} \tag{15}$$

is the sum of all four numerators and normalizes the posterior distribution, and the common factors $\Theta_1\Theta_2\Theta_3/8$ in the numerators and denominator of Eq. (14) have canceled. Initial conditions are as in Eq. (10). Since this simplified system is still nonlinearly coupled through the denominator D_t, we work with the joint probability $v_{i,j}^t \triangleq p(s_2=i, M=j, \mathbf{X}_t)$ instead. The two are related as follows:

$$\begin{aligned} z_{i,j}^t &= P(s_2=i, M=j|X_t)=\frac{P(s_2=i,M=j|\mathbf{X}_{t-1})p(\mathbf{x}_t|s_2=i,M=j)}{p(\mathbf{x}_t|\mathbf{X}_{t-1})}\\ &=\frac{P(s_2=i,M=j)\prod_{t'=1}^{t}p(\mathbf{x}_{t'}|s_2=i,M=j)}{p(\mathbf{X}_t)}=\frac{v_{i,j}^t}{\sum_{k,l}v_{k,l}^t}. \end{aligned} \tag{16}$$

The joint probability $v_{i,j}^t$ obeys the uncoupled update rule:

$$v_{i,j}^t=l_{i,j}^t v_{i,j}^{t-1} \approx \left(1+\frac{\pm x_1(t)\pm x_2(t)\pm x_3(t)}{\sigma^2}\right)v_{i,j}^{t-1}, \tag{17}$$

where the sign preceding each x_i depends on i and j as in Eq. (14). As is apparent in Eq. (16), z_{ij}^t can be obtained by normalizing v_{ij}^t on timestep t, but v_{ij}^t cannot be used directly in the perceptual decision process, since a fixed threshold in the posterior probability space has no equivalent fixed value in the joint posterior space. However, v_{ij}^t *is* sufficient for deriving bounds on the generative parameters that on average make the posterior probability for s_2=1 dip below that for s_2=-1 after *one* time step, when the inputs are generated from the incompatible stimulus array: **s**= (-1, 1, -1) (the analysis for **s**=(1, -1, 1) is analogous). Specifically, since $P(s_2, M|\mathbf{X}_t)=p(s_2, M, \mathbf{X}_t)/p(\mathbf{X}t)$, the condition $\left\langle z_{1,1}^1+z_{1,2}^1\right\rangle<\left\langle z_{-1,1}^1+z_{-1,2}^1\right\rangle$ is equivalent to $\left\langle v_{1,1}^1\right\rangle+\left\langle v_{1,2}^1\right\rangle<\left\langle v_{-1,1}^1\right\rangle+\left\langle v_{-1,2}^1\right\rangle$. We therefore require

$$\beta\left(1-\frac{1}{\sigma^2}\right)+(1-\beta)\left(1+\frac{3}{\sigma^2}\right)<\beta\left(1+\frac{1}{\sigma^2}\right)+(1-\beta)\left(1-\frac{3}{\sigma^2}\right)\Rightarrow\beta>\frac{3}{4}, \tag{18}$$

since the mean values of x_1, x_2, and x_3 are -1, 1,and -1, respectively, and the compatible terms are weighted by the compatibility prior bias β (and the incompatible ones weighted by 1-β).

Hence $\beta > 3/4$ is the necessary and sufficient condition for the average posterior probabilty for s_2=1 to dip below that for s_2=-1 after one observation, when the true stimuli are the incompatible

array (-1, 1, -1). More generally, we can show that the constraint is $\beta > (n + 1)/(2n)$, where n is the total number of flankers. This makes intuitive sense, for it suggests that the dip is more prominent or more likely to happen when the subject's prior compatibility bias is stronger and/or the number of flankers is larger. Indeed, there is behavioral data suggesting that flanker interference effects are stronger when there is as a lower frequency of incompatible trials [Gratton et al., 1992].

These analytical constraints only guarantee a dip in the posterior probability. As shown in Figure 2 (left), for a particular set of model parameters, the mean accuracy in compatible trials terminating within 20 timesteps steadily decreases as a function of β, and it drops below .5, indicating the presence of a dip, for all values of $\beta > 0.82$: somewhat higher than $\beta = 0.75$, the lower bound of inequality (18) that results in a dip in posterior probability. A major factor underlying the discrepancy between the two constraints is that we only considered the mean of the posterior probability and not the full distribution. The mean accuracy depends not only on the mean posterior value, but also on higher moments. If the distribution were symmetric about its mean, then the dip in the mean posterior would directly translate into a dip in accuracy, but as we will show in Section 4, the distribution of the posterior trajectories is strongly skewed, and the interaction of that skewness with the decision threshold also plays a role in determining the presence of the dip in accuracy versus reaction time.

A second reason for the discrepancy is that the theoretical bounds are for the dip to occur in the posterior after one time step, whereas in the numerical simlations, due to the infrequency of responses at very short RT's, all trials that terminate within the first 20 timesteps were used to estimate accuracy. If the temporal extent of the dip in the *posterior distribution* is very small (which is likely in boundary cases), then conditional accuracy may not fall below 0.5 when averaged over 20 timesteps. The numerically-obtained constraints are therefore likely to be more conservative than the analytical approximations.

3.2 The spatial uncertainty model

Derivation of iterated maps for the spatial uncertainty model are more tedious than those of (14) due to the extra "cross-talk" links in the generative model, but they follow from similar reasoning. Defining $h^t_{k,i,j} \triangleq p(x_k(t)|s_k{=}i, M{=}j)$, forming the triple product and dividing through by

$$\Theta' = \prod_{k=1}^{3} \left[h^t_{k,1,1} + h^t_{k,-1,1} + h^t_{k,1,2} + h^t_{k,-1,2} \right], \tag{19}$$

we obtain the approximate update rule:

$$z^t_{i,j} = \frac{1}{D'_{t-1}} \times \begin{cases} [1 - A_1 + (A_2 + A_3)(x_1 + x_3) + (A_4 + A_5)x_2]\, z^t_{+1,1}, \\ [1 - A_1 - (A_2 + A_3)(x_1 + x_3) - (A_4 + A_5)x_2]\, z^t_{-1,1}, \\ [1 + A_1 - (A_2 - A_3)(x_1 + x_3) + (A_4 - A_5)x_2]\, z^t_{+1,2}, \\ [1 + A_1 + (A_2 - A_3)(x_1 + x_3) - (A_4 - A_5)x_2]\, z^t_{-1,2}, \end{cases} \tag{20}$$

where D'_t is again the sum of the numerators and the parameters A_i are

$$A_1=2a_1a_2\left(\frac{1}{\sigma_1^2+\sigma_2^2}+\frac{1}{\sigma_1^2+2\sigma_2^2}\right), A_2=\frac{a_1}{\sigma_1^2+\sigma_2^2},$$
$$A_3=\frac{a_2}{\sigma_1^2+\sigma_2^2}, A_4=\frac{a_1}{\sigma_1^2+2\sigma_2^2}, A_5=\frac{2a_2}{\sigma_1^2+2\sigma_2^2}. \quad (21)$$

Since the prior distribution is uniform, the initial conditions for (20) are

$$z_{i,j}^0=\frac{1}{4}, \text{for} \quad i=\pm 1 \quad \text{and} \quad j=1 \quad \text{or} \quad 2. \quad (22)$$

Again, the constraint

$$\langle P(s_2=1, M=1|\mathbf{X}_1)\rangle + \langle P(s_2=1, M=2|\mathbf{X}_1)\rangle < \langle P(s_2=-1, M=1|\mathbf{X}_1)\rangle + \langle P(s_2=-1, M=2|\mathbf{X}_1)\rangle . \quad (23)$$

is satisfied if $A_4(a_1 - 2a_2) < 2A_3(a_1 - a_2)$, or equivalently, if the ratio of means, a_1/a_2, lies in the interval

$$\left[\frac{2r+3-\sqrt{2r^2+6r+5}}{1+r}, \frac{2r+3+\sqrt{2r^2+6r+5}}{1+r}\right], \quad (24)$$

where $r \triangleq \sigma_1^2/\sigma_2^2$ is the ratio of the variances. Intuitively, if the ratio a_1/a_2 is too large, the flankers have negligible effects; if it is too small, the inputs lose their spatial selectivity altogether. More generally, if there are n flankers, the range is

$$\left[\frac{\left(\frac{n}{2}+1\right)r+n+1-\sqrt{\left(\frac{n^2}{4}+1\right)r^2+(n^2+2)r+n^2+1}}{1+r}, \frac{\left(\frac{n}{2}+1\right)r+n+1+\sqrt{\left(\frac{n^2}{4}+1\right)r^2+(n^2+2)r+n^2+1}}{1+r}\right].$$

We now compare these constraints with numerical simulations of the full inference model for the specific noise parameters (σ_1=7, σ_2=5). We simulated the full model using a range of values of a_1 and a_2 (with their sum held at 1), and obtained accuracy of all responses falling within the first 20 timesteps as a function of a_1/a_2. As can be seen in Figure 2 (right), the accuracy in this short-RT bin is less than .5 when a_1/a_2 falls within (0.70, 3.55), a somewhat more stringent condition than the analytically derived (approximate) interval (0.67, 3.98).

3.3 Evaluating the cost of linearization

Direct simulations of the linear approximation can be compared with those of the original inference model. Figure 3 shows the results for the compatibility bias model for a particular setting of parameters (σ=9), comparing the full inference model with the simplified iteration of (17). The same sequence of noisy observations $x_i(t)$ was used for both processes and in computing the value of $P(s_2 = 1|\mathbf{X}_t)$ for the latter at each timestep t, normalization was applied only at that step. The agreement is remarkably good, validating our linear approximations to the products of probabilites (5-4) developed in Section 3. The quality of the linear approximation for the spatial uncertainty model is similarly good (details not shown).

We can also simulate perceptual discrimination based on the linearized evidence accumulation process, using the first passage criterion for threshold crossing appropriate for free response conditions. As in [Yu et al., 2007], we adopt the decision threshold q = 0.9 for both the

compatibility bias and the spatial uncertainty model. The time span, taken here as 200 steps, is divided into ten bins and sample paths for the full model (9) and the approximate decoupled system (17) and its analogue for spatial uncertainty are computed. The decoupled results are then normalized by dividing by the sum $\sum_{i,j} v_{i,j}^t$ at each t in the current bin (normalization is not applied for steps 1 through t - 1). The same (unit) step size is used in all cases. Responses are logged when the first of the probabilities $P(s_2 = +1|\mathbf{X}_t) = P(s_2 = +1, M = 1|\mathbf{X}_t) + P(s_2 = +1, M = 2|\mathbf{X}_t)$ or $P(s_2=-1|\mathbf{X}_t) = P(s_2=-1, M = 1|\mathbf{X}_t) + P(s_2=-1, M = 2|\mathbf{X}_t)$ crosses q. After collecting sufficiently many paths (2000 in this case), response time histograms are formed and the fraction of correct responses in each bin summed to yield accuracy vs. time curves.

Figure 4 illustrates the results of such simulations for the compatibility and spatial uncertainty models. Accuracy vs. reaction time, and empirical distributions of reaction time are shown for both the full and approximate models. The approximate systems reproduce the characteristic dip in accuracy for fast incompatible trials for both models, and the accuracy curves and reaction time distributions predicted by the approximate theory agree well with those of the full inference models. Note that the use of the first passage criterion for response produces reaction time distributions that agree with the exact model in details of their shapes: a rise at short reactions times to a peak, followed by a long tail. The distributions for incompatible trials are also flatter and shifted rightward compared to those for compatible trials, as in the data of Figure 1.

4 A continuum limit

The key difficulty in working with the discrete dynamical systems (14) and (20) lies in the nonlinear coupling of the posteriors $z_{i,j}^t$ through the denominators D_t and D_t'. It can be proved that individual sample paths generated with the same noise inputs are identical whether computed by iteration of Eqs. (14) and (20) or by the analogous uncoupled systems Eq. (17), with posteriors normalised only at the last time step; cf. Eq. (16). (In computing the values for the approximate model (17) at each step t for Figure 3, normalization was applied only at that step, but not at steps 1 through t - 1, while the full iteration (9) is normalised at every step.) However, it *does not follow* that we may average over many realizations of the unnormalized process, and then normalize: as discussed further in Section 4.3, since these operations do not commute. Nonetheless, we can decouple the dynamics by replacing the normalization constant D_t at each time step with its expectation $\langle D_t \rangle$, which does not depend on the inputs, and replacing that in turn by a constant. We then take continuum limits of the resulting decoupled linear systems to form stochastic differential equations (SDEs), allowing us to use simple analytical results to compute properties of interest. As described further in Section 5, these SDEs may in turn be related to neurally-based models of evidence accumulation.

4.1 Approximating the denominators

We first examine the denominator $\langle D_t \rangle$ for the compatibility bias model:

$$\begin{aligned}\langle D_t \rangle = &\left\langle 1+\tfrac{x_1(t)+x_2(t)+x_3(t)}{\sigma^2}\right\rangle \left\langle z_{+1,1}^{t-1}\right\rangle + \left\langle 1-\tfrac{x_1(t)+x_2(t)+x_3(t)}{\sigma^2}\right\rangle \left\langle z_{-1,1}^{t-1}\right\rangle \\ &+\left\langle 1-\tfrac{x_1(t)-x_2(t)+x_3(t)}{\sigma^2}\right\rangle \left\langle z_{+1,2}^{t-1}\right\rangle + \left\langle 1+\tfrac{x_1(t)-x_2(t)+x_3(t)}{\sigma^2}\right\rangle \left\langle z_{-1,2}^{t-1}\right\rangle,\end{aligned}$$

where the approximation comes from assuming that the input-dependent terms (functions of $x_k(t)$) are independent from the z_{ij} terms, which depend on the previous inputs $\mathbf{x}_k(1)$, ...$\mathbf{x}_k(t)$. Although the inputs are conditionally independent (cf. Eq. (5)), they are *marginally* dependent. That is, if previous inputs favored a particular setting of s_2 and M, then the current one also tends to do the same. For analytical simplicity, we ignore this statistical dependence. Note that

in the limit as $t \to \infty$, one of the $z^t_{i,j}$'s (corresponding to the actual stimulus setting) converges to 1 (and the others to 0), and that no matter which $z^t_{i,j}$ it is,

$$\langle D_t \rangle \to 1+\frac{3}{\sigma^2}. \tag{25}$$

More generally, we expect $\langle D_t \rangle$ to increase from 1 (D_0 is just the sum of the priors) to $1+\frac{3\mu}{\sigma^2}$, where μ denotes the mean value of the x_j's. Figure 5 shows exactly this for both compatible and incompatible stimuli for a particular setting of the model parameters, where s_2=1 and averaged over 10^5 trials. Convergence is slower for incompatible stimuli, since the compatibility prior takes time to update from its initial value $P(M) = 0.9$.

Based on these arguments, and in spite of the fact that D_t can exhibit large variance on individual trials, we assume $D_t \approx \langle D_t \rangle \approx 1$, and approximate the dynamics of Eq. (14) by the following linear, decoupled system:

$$\begin{aligned} z^t_{+1,1} &= \left(1+\tfrac{x_1+x_2+x_3}{\sigma^2}\right) z^{t-1}_{+1,1}, \\ z^t_{-1,1} &= \left(1+\tfrac{x_1-x_2+x_3}{\sigma^2}\right) z^{t-1}_{-1,1}, \\ z^t_{+1,2} &= \left(1-\tfrac{x_1-x_2+x_3}{\sigma^2}\right) z^{t-1}_{+1,2}, \\ z^t_{-1,2} &= \left(1-\tfrac{x_1+x_2+x_3}{\sigma^2}\right) z^{t-1}_{-1,2}, \end{aligned} \tag{26}$$

with initial conditions

$$z^0_{\pm 1,1} = \frac{1}{2}\beta, \quad z^0_{\pm 1,2} = \frac{1}{2}(1-\beta). \tag{27}$$

Similiar reasoning can be used to derive a linear, decoupled approximation for Eq. (20) for the spatial uncertainty model. The approximate dynamics for both models can be written as an iterated linear mapping in the following form

$$z^t_{i,j} = \left(a_{i,j} + b_{i,j}\eta(t)\right) z^{t-1}_{i,j}, \quad i=1,\ldots,4, \tag{28}$$

where the random variables $\eta(t)$ are drawn from a standard normal distribution, and $a_{i,j}$ and $b_{i,j}$ are constant parameters whose values depend on the model, the probability being computed, and the compatibility condition of the given trial.

For the compatibility bias model, from the details presented in §3.1 if the current stimulus array $\mathbf{s}(t)$ is compatible and $s_2 = 1$ we have

$$a_{i,j} = \begin{cases} 1+\frac{3}{\sigma^2}, & i=+1,\ j=1, \\ 1-\frac{3}{\sigma^2}, & i=-1,\ j=1, \\ 1-\frac{1}{\sigma^2}, & i=+1,\ j=2, \\ 1+\frac{1}{\sigma^2}, & i=-1,\ j=2, \end{cases} \quad \text{and} \quad b_{i,j} = \frac{\sqrt{3}}{\sigma} \quad \forall i, j; \tag{29}$$

and if $\mathbf{s}(t)$ is incompatible and $s_2 = 1$ we have

$$a_{i,j}=\begin{cases} 1-\frac{1}{\sigma^2}, & i=+1,\ j=1,\\ 1+\frac{1}{\sigma^2}, & i=-1,\ j=1,\\ 1+\frac{3}{\sigma^2}, & i=+1,\ j=2,\\ 1-\frac{3}{\sigma^2}, & i=-1,\ j=2,\end{cases} \quad \text{and} \quad b_{i,j}=\frac{\sqrt{3}}{\sigma} \quad \forall i,j. \tag{30}$$

For s_2 = -1 all the signs in $a_{i,j}$ above are reversed.

For the spatial uncertainty model with compatible stimulus array and s_2 = 1, the calculations of §3.2 imply:

$$a_{i,j}=\begin{cases} 1-A_1+2(a_1+a_2)(A_2+A_3)+(a_1+2a_2)(A_4+A_5), & i=+1,\ j=1,\\ 1-A_1-2(a_1+a_2)(A_2+A_3)-(a_1+2a_2)(A_4+A_5), & i=-1,\ j=1,\\ 1+A_1-2(a_1+a_2)(A_2-A_3)+(a_1+2a_2)(A_4-A_5), & i=+1,\ j=2,\\ 1+A_1+2(a_1+a_2)(A_2-A_3)-(a_1+2a_2)(A_4-A_5), & i=-1,\ j=2,\end{cases} \tag{31}$$

and for an incompatible stimulus array and s_2 = 1:

$$a_{i,j}=\begin{cases} 1-A_1-2(a_1-a_2)(A_2+A_3)+(a_1-2a_2)(A_4+A_5), & i=+1,\ j=1,\\ 1-A_1+2(a_1-a_2)(A_2+A_3)-(a_1-2a_2)(A_4+A_5), & i=-1,\ j=1,\\ 1+A_1+2(a_1-a_2)(A_2-A_3)+(a_1-2a_2)(A_4-A_5), & i=+1,\ j=2,\\ 1+A_1-2(a_1-a_2)(A_2-A_3)-(a_1-2a_2)(A_4-A_5), & i=-1,\ j=2.\end{cases} \tag{32}$$

In both cases the standard deviation of the noise is given by

$$b_{i,j}=\begin{cases} \sqrt{2\left(\sigma_1^2+\sigma_2^2\right)(A_2+A_3)^2+\left(\sigma_1^2+2\sigma_2^2\right)(A_4+A_5)^2}, & i=\pm 1,\ j=1,\\ \sqrt{2\left(\sigma_1^2+\sigma_2^2\right)(A_2-A_3)^2+\left(\sigma_1^2+2\sigma_2^2\right)(A_4-A_5)^2}, & i=\pm 1,\ j=2.\end{cases} \tag{33}$$

Figure 6 illustrates normal distributions from which these multiplicative terms in (28) are drawn.

4.2 Taking the continuum limit

We now take continuum limits of the discrete dynamical systems derived above that will allow us compute properties of interest analytically. First consider the following finite-difference limit of the iterated mapping (28):

$$\frac{d\left(z_{i,j}^t\right)}{dt}=\lim_{\delta t\to 0}\frac{z_{i,j}^t-z_{i,j}^{t-\delta t}}{\delta t}=\lim_{\delta t\to 0}\left[\left(a_{i,j}-1\right)+b_{i,j}\eta(t)\right]z_{i,j}^{t-\delta t}, \tag{34}$$

where the $z_{i,j}^t$ represent the four posteriors $P(s_2, M|\mathbf{X}_t)$. For finite but small $\delta t = 1/k$, this represents a finer-grained discretization in which k steps are taken for every one step of (28), the deterministic increments being of order δt and the random ones of order $\sqrt{\delta t}$ [Higham, 2001]. Taking the limit $\delta t \to 0$ in Eq. (34), letting $y_{i,j} = \log(z_{i,j})$, and appealing to the Ito formula [Oksendal, 2002, Section 4.1], we obtain independent, uncoupled SDEs for $y_{i,j}(t)$:

$$dy_{i,j} = \left[\left(a_{i,j} - 1 \right) - \frac{b_{i,j}^2}{2} \right] dt + b_{i,j} dW \stackrel{\text{def}}{=} A_{i,j} dt + B_{i,j} dW, \tag{35}$$

with constant coefficients $A_{i,j} = \left(a_{i,j} - 1 \right) - \frac{b_{i,j}^2}{2}$ and $B_{i,j} = b_{i,j}$, whose values are specified in §4.1. Since each $z_{i,j}(t)$ represents a posterior probability, it should take values in the interval [0, 1], so we shall be interested in sample paths $y_{i,j}(t)$ that start at $y_{i,j}(0) < 0$ and satisfy $-\infty < y_{i,j}(t) \leq 0$.

4.3 Analytical approximations for the mean posteriors

The SDE (35) describes a drift-diffusion process with constant signal and noise level, which has been extensively studied (e.g. [Gardiner, 1985,Oksendal, 2002]). In particular, for solutions (sample paths) started at $y(0) = \mu_0$ and $t = 0$ the probability density function of y at time t is the following Gaussian distribution:

$$p(y,t) = \frac{1}{\sqrt{2\pi\sigma(t)^2}} \exp\left[-\frac{(y - \mu(t))^2}{2\sigma(t)^2} \right], \tag{36}$$

where

$$\mu(t) = At + \mu_0 \quad \text{and} \quad \sigma(t) = B\sqrt{t}. \tag{37}$$

(Here and below we drop the subscripts $\{i, j\}$ in y and z in the understanding that the appropriate coefficients will be used in the final formulae.) We now transform back into z-space, using $y = \log(z)$ and $dy = \frac{dz}{z}$ to obtain the density:

$$p(z,t) = \frac{1}{z\sqrt{2\pi\sigma(t)^2}} \exp\left[-\frac{(\log(z) - \mu(t))^2}{2\sigma(t)^2} \right]. \tag{38}$$

The inverse transformation $z = \exp(y)$ takes the Gaussian distribution over y into a function skewed towards $z = 1$, as illustrated in Figure 7.

The Gaussian distribution over y takes positive values on $y > 0$ for all $t > 0$. This presents a problem, since $z = \exp(y) > 1$ for $y > 0$, contrary to z's designation as a probability measure. Therefore, when computing expected values of $P(s_2, M|\mathbf{X}_t)$, which requires integration of the quantity $z\,p(z, t)$, we replace all values of $z > 1$ by $z = 1$ (or values of $y > 0$ by $y = 0$ in the equivalent integral over y). However, to retain analytical tractability, we continue to assume a Gaussian distribution over y at time t when generating the distribution at time $t+1$ - that is, we only replace the inappropriate values of y (or x) in the integral, not in the underlying drift-diffusion process. The expected (mean) value of z is therefore approximated as

$$\langle z(t)\rangle \approx \int_0^1 z p(z,t)\,dz = \int_0^1 \frac{1}{z\sqrt{2\pi\sigma(t)^2}} \exp\left[-\frac{(\log(z)-\mu(t))^2}{2\sigma(t)^2}\right] z\,dz + \int_1^\infty \frac{1}{z\sqrt{2\pi\sigma(t)^2}} \exp\left[-\frac{(\log(z)-\mu(t))^2}{2\sigma(t)^2}\right] dz, \tag{39}$$

which may be evaluated as explained in Appendix A to yield

$$\frac{\exp\left[\mu(t)+\frac{\sigma(t)^2}{2}\right]}{2}\left[1-\operatorname{erf}\left(\frac{\mu(t)+\sigma(t)^2}{\sqrt{2\sigma(t)^2}}\right)\right]+\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{\mu(t)}{\sqrt{2\sigma(t)^2}}\right)\right]. \tag{40}$$

Substituting values appropriate for the compatibility bias model from Eqs. (29-30) for the parameters $a_{i,j}$ and $b_{i,j}$, and hence for $A_{i,j}$, $B_{i,j}$, and via Eqs. (37), for $\mu(t)$ and $\sigma(t)$, we obtain estimates for the four mean posterior probabilities at time t:

$$\langle P(s_2, M|\mathbf{X}_t)\rangle \approx \frac{1}{2D(t)}\times \left\{\exp\left[\mu(t)+\frac{\sigma(t)^2}{2}\right]\left[1-\operatorname{erf}\left(\frac{\mu(t)+\sigma(t)^2}{\sqrt{2\sigma(t)^2}}\right)\right]+\left[1+\operatorname{erf}\left(\frac{\mu(t)}{\sqrt{2\sigma(t)^2}}\right)\right]\right\}. \tag{41}$$

where $D(t)$ is the sum of all four probabilities that normalizes the expressions, and for compatible stimuli the functions $\mu(t)$ and $\sigma(t)$ are:

$$\mu(t)=\begin{cases} +\frac{3t}{2\sigma^2}+\log\left(\frac{\beta}{2}\right), & s_2=+1, M=1,\\ -\frac{9t}{2\sigma^2}+\log\left(\frac{\beta}{2}\right), & s_2=-1, M=1,\\ -\frac{5t}{2\sigma^2}+\log\left(\frac{1-\beta}{2}\right), & s_2=+1, M=2,\\ -\frac{t}{2\sigma^2}+\log\left(\frac{1-\beta}{2}\right), & s_2=-1, M=2, \end{cases} \quad \text{and} \quad \sigma(t)=\frac{\sqrt{3t}}{\sigma}. \tag{42}$$

and for incompatible stimuli:

$$\mu(t)=\begin{cases} +\frac{5t}{2\sigma^2}+\log\left(\frac{\beta}{2}\right), & s_2=+1, M=1,\\ -\frac{t}{2\sigma^2}+\log\left(\frac{\beta}{2}\right), & s_2=-1, M=1,\\ +\frac{3t}{2\sigma^2}+\log\left(\frac{1-\beta}{2}\right), & s_2=+1, M=2,\\ -\frac{9t}{2\sigma^2}+\log\left(\frac{1-\beta}{2}\right), & s_2=-1, M=2, \end{cases} \quad \text{and} \quad \sigma(t)=\frac{\sqrt{3t}}{\sigma}. \tag{43}$$

Here, we also use the fact that all sample paths start with the initial conditions specified in Eq. (10) and that $\mu_0 = \mu(0) = \log(z(0))$.

As noted at the beginning of this section, normalization and averaging do not commute. This may be understood in terms of the distributions of Figure 7 as follows. While each sample path can be computed for the uncoupled processes and normalized at time t to yield the same result as a sample path of the coupled system (cf. Figure 3), *different* normalization factors must typically be applied to the values of different paths $z_{i,j}(t)$ at each time t. This would distort the distributions $p(z_{i,j}, t)$, thereby changing their means. However, we may appeal to the observation that the expected value of the denominator remains close to 1 (cf. Figure 5) to

conclude that this distortion is likely to be small, and proceed by dividing by the sums of the four mean probability trajectory values at time t to normalize the resulting expressions.

Typical results for mean posterior probabilities are shown in Figure 8. The approximate predictions developed above are shown as dashed curves and the results of averaging over 5000 simulated trials of the full inference model (9) are shown solid; compatible and incompatible trials are shown in red and blue respectively. As above, we compute 200 steps for the discrete iteration of the full system, and we evaluate the corresponding quantities for $t \in [0, 200]$ time units from the formulae above. For $P(M) = 0.5$ (not shown), joint posteriors for correct responses increase similarly for both compatible and incompatible cases, but $P(M)$=0.9 elicts markedly different behaviors (top left). The compatibility posteriors $P(M = 1|\mathbf{X}_t)$ show a general rise for compatible stimuli and a monotonic fall for incompatible stimuli, but the posterior probability $P(s_2 = 1|\mathbf{X}_t)$ shows a significant dip below 0.5 at early times for incompatible stimuli, while it rises monotonically for compatible stimuli. As discussed in Section 5, the resulting accuracies exhibit similar patterns to the experimental data, with the incompatible case showing a dip in accuracy for early responses. Evolutions of the four individual posterior probabilities are shown in the lower panels of Figure 8.

Figure 8 illustrates that, while the approximations developed here do not capture all the detailed behavior of the full model, they do provide reasonably good approximations to the average evolutions of the posteriors over the course of a trial. Time scales are slightly misestimated and the compatibility posterior $P(M=1|\mathbf{X}_t)$ (top right) fails to reproduce the slight dip below 0.9 that occurs for compatible trials at early times, but the relative orderings of all the posteriors are correctly predicted. Overall, absolute errors in mean posteriors, computed as described at the end of this section, lie between 0.002 and 0.05, the largest being for $P(M=1|\mathbf{X}_t)$ in the case of incompatible stimuli (top right, lower curves).

Predictions for the spatial uncertainty model follow from the formula (41) in a similar manner, upon the substitution of values for a and b from Eqs. (31-33), and using the initial conditions μ_0=log(1/4) for all four posteriors (Eq. (22)). For compatible stimuli, the function $\mu(t)$ is

$$\mu(t) = \begin{cases} \left[\frac{a_1^2+a_2^2}{\sigma_1^2+\sigma_2^2} - \frac{2a_1a_2}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=+1, M=1, \\ \left[-\frac{3a_1^2+8a_1a_2+3a_2^2}{\sigma_1^2+\sigma 2^2} - \frac{2(a_1+a_2)(a_1+4a_2)}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=-1, M=1, \\ \left[\frac{-3a_1^2+4a_1a_2+a_2^2}{\sigma_1^2+\sigma_2^2} + \frac{2(3a_1-4a_2)a_2}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=+1, M=2, \\ \left[\frac{a_1^2+4a_1a_2-3a_2^2}{\sigma_1^2+\sigma_2^2} - \frac{2a_1(a_1-3a_2)}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=-1, M=2; \end{cases} \tag{44}$$

for incompatible stimuli

$$\mu(t) = \begin{cases} \left[\frac{-3a_1^2-4a_1a_2+a_2^2}{\sigma_1^2+\sigma_2^2} - \frac{2(3a_1+4a_2)a_2}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=+1, M=1, \\ \left[\frac{a_1^2-4a_1a_2-3a_2^2}{\sigma_1^2+\sigma_2^2} - \frac{2a_1(a_1+3a_2)}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=-1, M=1, \\ \left[\frac{a_1^2+a_2^2}{\sigma_1^2+\sigma_2^2} + \frac{2a_1a_2}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=+1, M=2, \\ \left[-\frac{3a_1^2-8a_1a_2+3a_2^2}{\sigma_1^2+\sigma 2^2} - \frac{2(a_1-a_2)(a_1-4a_2)}{\sigma_1^2+2\sigma_2^2}\right]t + \log\left(\frac{1}{4}\right), & s_2=-1, M=2, \end{cases} \tag{45}$$

and in both cases

$$\sigma(t)=\begin{cases}\sqrt{\left[\frac{(a_1+a_2)^2}{\sigma_1^2+\sigma_2^2}+\frac{(a_1+2a_2)^2}{\sigma_1^2+2\sigma_2^2}\right]t}, & s_2=\pm 1, M=1,\\ \sqrt{\left[\frac{(a_1-a_2)^2}{\sigma_1^2+\sigma_2^2}+\frac{(a_1-2a_2)^2}{\sigma_1^2+2\sigma_2^2}\right]t}, & s_2=\pm 1, M=2.\end{cases} \tag{46}$$

The above results, presented in Figure 9, are not as good as those for the compatibility bias model. Nonetheless, the approximate model captures the key features of the evolving posteriors in the full model rather well, prediciting the relative ordering of the posteriors appropriately in all cases except the incorrect choices *P*(HHH) and *P*(SHS) for incompatible stimuli; in that case the approximation for *P*(SHS) diverges from the correct function, *increasing* rather than decreasing as *t* increases (lower right panel), for an absolute error of 0.12. Apart from this case, however, errors lie between 0.015 and 0.08.

The errors for both models were computed for each mean posterior using the L^1 norm as follows:

$$\text{Error}=\sum_{t=o}^{T}|p_t-\tilde{p}_t|, \tag{47}$$

where p_t and $\tilde{p}_t$ denote the posteriors predicted by the full and simplified models respectively.

4.4 Making use of explicit mean posteriors

In addition to providing explicit expressions for posterior probabilities, the continuum limit also yields approximations for accuracy and reaction time distributions. To estimate accuracy as a function of response time under the free response protocol assumed by [Yu et al., 2007], we compute the fraction of mass of the evolving probability density $p\left(z_{i,1}^t, z_{i,2}^t\right)$ that exceeds a given threshold $z_{i,1}^t+z_{i,2}^t=q$ at each time *t* (recall Eq. (11)). This procedure *overestimates* first passage times, since some of the sample paths that lie beyond the threshold *q* at time *t* may have crossed at earlier times, but it permits some analytical simplification. Without loss of generality, we shall assume that $s_2 = 1$.

The integral that we need to evaluate is

$$P(s_2=1|\mathbf{X}_t)_{\text{est}}=\int_0^\infty\int_{q-z_2}^\infty p\left(z_{1,1}^t, z_{1,2}^t\right)dz_{1,1}^t dz_{1,2}^t \approx \int_0^\infty\int_{q-z_2}^\infty p(z_1,t)\,p(z_2,t)\,dz_1 dz_2, \tag{48}$$

where we have used the shorthand notation $p\left(z_j,t\right)=p\left(z_{1,j}^t\right)$, and the approximation comes from assuming $p\left(z_{1,1}^t, z_{1,2}^t\right) \approx p\left(z_{1,1}^t\right)p\left(z_{1,2}^t\right)$ for the uncoupled and linearized approximate dynamical system - this assumption greatly simplifies the computations, although the uncoupled processes are not entirely independent since they are activated by common inputs (x_1, x_2, x_3), albeit in different linear combinations. We also note that the variables z_j should be non-negative (cf. Figure 7). The domain of integration is pictured in Figure 12. The $p(z_j, t)$'s take the forms derived in §4.3 above and since each is a normalized Gaussian in the logarithmic *y* variables, the integral of their product over the entire positive quadrant is 1. Hence we have

$$P(s_2=1|\mathbf{X}_t)_{\text{est}}=1-\int_0^q\int_0^{q-z_2} p(z_1,t)\,p(z_2,t)\,dz_1 dz_2, \tag{49}$$

which is evaluated in Appendix A to yield:

$$P(s_2=1|\mathbf{X}_t)_{\text{est}}=\frac{3}{4}-\frac{1}{4}\text{erf}\left(\frac{\log(q)-\mu_2(t)}{\sqrt{2\sigma_2(t)^2}}\right)-\frac{1}{2}\int_0^q p(z_2,t)\,\text{erf}\left(\frac{\log(q-z_2)-\mu_1(t)}{\sqrt{2\sigma_1(t)^2}}\right)dz_2, \tag{50}$$

where

$$\overline{2\sigma_2(t)^2}. \tag{51}$$

Unfortunately, the final integral in Eq. (50) cannot be computed analytically, but it can be evaluated accurately and rapidly by numerical methods.

Response accuracy is approximated by the fraction of correct responses that exceed threshold:

$$\frac{P(s_2=1|\mathbf{X}_t)_{\text{est}}}{P(s_2=1|\mathbf{X}_t)_{\text{est}}+P(s_2=2|\mathbf{X}_t)_{\text{est}}}, \tag{52}$$

where the denominator approximates the sum of all four probabilities $z^t_{1,1}+z^t_{1,2}+z^t_{2,1}+z^t_{2,2}$. (The term $P(s_2=2|\mathbf{X}_t)_{\text{est}}$ is computed in a similar manner to Eq. (50), with the appropriate expressions for $\mu(t)$, $\sigma(t)$ from §4.3.) The denominator is the cumulative reaction time and so its derivative with respect to t provides the reaction time distribution. Hence, both accuracy and reaction time distributions can be approximated semi-analytically. Figure 10 shows the resulting approximations to the mean posteriors for the compatibility bias model, for a particular setting of model parameters. The dip in accuracy for incompatible trials is reproduced, and after an initial rise in accuracy for compatible trials, accuracy slowly declines.

As we have noted, sample paths of the SDE (35) may pass across q and back, possibly repeatedly, in the interval $(0, t)$, so these results do not directly correspond to the first-passage decision policy of the Bayesian models in [Yu et al., 2007]. This accounts for differences between the accuracy curves and reaction time distributions of Figure 1 and the free response results of §3.3. For example, the compatibility bias free response data of Figure 4 do not show the mild decline in accuracy for later compatible trials of Figure 10, although the spatial uncertainty simulations of Figure 4 do show such a decline. Nonetheless, the qualitative agreement between Figures 10 and 4 is quite good, and since the semi-explicit expression Eqs. (50-51) replaces lengthy Monte-Carlo simulations of §3.3, it may be helpful in guiding parameter fits to data.

The posterior probability expressions can also be used to constrain parameter choices, by requiring the derivative of $P(s_2=1|\mathbf{X_t})=z^t_{1,1}+z^t_{1,2}$ at time $t = 0$ to be negative and finding corresponding conditions on the parameters. The results of this computation (details not shown) agree closely with those in Section 3.

4.5 Fitting the models to data

We now briefly describe the results of fitting the full models of Section 2 and the reduced DD processes of Sections 4.2-4.4 to the data of [Servan-Schreiber et al., 1998], reproduced in Fig. 1B. For the compatibility bias model the parameters fitted are the noise level σ, prior β, threshold q and step durations δt (for DDM) and ΔT (for the full model), which determine the overall timescale. For spatial uncertainty, they are σ_1, σ_2, a_1, q and δt, ΔT (as in §3.2, we set $a_2 = 1 - a_1$). To these we add one further parameter, T_0, to account for time occupied by sensory

decoding and motor response mechanisms, which superimposes a rightward shift on the RT distributions. (Such an "overhead time" might approximate the mean RT on a simple target detection task).

We employ the same weighted Euclidean error norm as in [Liu et al., 2007] (see Appendix B for details). The parameter values obtained are as follows. Compatibility bias: $\sigma = 6.5$, $\beta = 0.87$, $q = 0.98$, $\delta t = 0.95$ ms, $\Delta T = 1.04$ ms, and $T_0 = 90$ ms. Spatial uncertainty: $\sigma_1 = 6.9$, $\sigma_2 = 5.1$, $a_1 = 0.71$, $q = 0.92$, $\delta t = 3.4$ ms, $\Delta T = 0.33$ ms, and $T_0 = 95$ ms. Note that the noise levels are consistent with the assumptions of Sections 3 and 4.1-4.2: e.g., $1/\sigma^4 \ll 1/\sigma^2$ (cf. Equation (13)). The fitting errors are as follows: Compatibility bias: full model 2.5; DDM 2.3. Spatial uncertainty: full model 2.1; DDM 1.8. In fitting we excluded data points in the first (0 - 100 ms) and the last (900 - 1000 ms)of the 10 RT bins, since no accuracy data is available for the former, and all trials in which responses exceeded 1000 ms were placed in the latter (note the uptick in RT distributions at the rightmost data point). However, we computed model data in that bin and in the next one (1000 - 1100 ms). Since our fitted values of the overhead time T_0 push even the shortest model RTs beyond 100 ms, accuracies cannot be computed for the 0-100 ms bin, unless we assume some premature responses that are initiated before stimulus onset. For such premature responses, the equal prevalence of H and S in the experiments ensure that accuracy approaches chance at very short decision times (cf. upper left panels of Figs. 8 and 9). Indeed, this chance performance is unavoidable, independent of the inference or decision strategy, since the response is deprived of stimulus information and cannot possibly correlate with stimulus identity.

These results are shown in Fig. 11. Fit qualities are slightly better for the spatial uncertainty model, and in both cases, perhaps surprisingly, fit errors are slightly smaller for the reduced DDM than for the full Bayesian procedure. The fit errors are similar to that of 2.4 obtained in [Liu et al., 2007] for the [Gratton et al., 1988] data (Fig. 1A), using a DDM with variable drift rates derived from the neural network model of [Cohen et al., 1992]. That model contains 8 free parameters, compared with 5 and 6 respectively in the present cases. Indeed, in [Liu et al., 2007] 6 parameters are required to describe drift rates in the compatible and incompatible cases, modeling progressive increase in attention to the central stimulus, and these cases are fitted separately. In the present study compatible and incompatible trials are fitted simultaneously, and a single parameter in each model (the compatability prior β, or the weight a_1), along with Bayesian updating, serves to describe the accumulation of evidence.

Both models underestimate mean RTs for compatible trials, producing an excess of points in the 200-250 ms RT bin. They are also unable to capture the drop in accuracy at the shortest RTs on compatible trials (left panels), due to the T_0 behavior noted above. They do reproduce this drop on incompatible trials, although the full compatibility bias model does not exhibit the dip below 50%. The spatial uncertainty model is substantially better in this regard (lower right panel), although it underestimates accuracy in the 400 - 900 ms part of the RT range for both the compatible and incompatible cases. In preliminary work we also tried a modified norm that preferentially weights low RT data: this slightly improved fits of RT distributions, but did not affect compatible accuracy fits. We also fitted the full and DD models to the data of [Gratton et al., 1988] (Fig. 1A), obtaining similar fit qualities, although the failure to capture the steady rise from 50% accuracy at low RTs for compatible trials was more striking in that case (model results not shown here).

We note that individual subjects exhibit large differences in signal-to-noise ratios and thresholds (in DDM fits, cf. [Ratcliff et al., 1999, Bogacz et al., 2007]), and that here we have averaged over all subjects to produce single sets of fit parameters for each model. As illustrated in Fig. 1, there is also substantial variability in Eriksen data, perhaps due to differing deadlining protocols. (Deadlines are necessary to produce enough short reaction times and hence obtain

a significant dip in accuracy on incompatible trials.) The resulting variability in motor preparation times can affect reaction times, and no allowance for this is made in the inference model, which describe only cognitive processing. Our additional parameter T_0 only partially accounts for this, and as we have remarked, in the present case it deprives us of accuracy data in the smallest RT bin.

5 Discussion and conclusions

In earlier work [Liu et al., 2007] a neural network model of the Eriksen task [Cohen et al., 1992, Servan-Schreiber et al., 1998] was linearized and reduced to a DDM with time-varying drift, allowing relatively complete analysis that reveals how parameters influence accuracy curves such as those of Figure 1. However, this network model involves somewhat arbitrary assumptions on architecture and parameters, and it is not clear how the DDM reduction of [Liu et al., 2007], with its variable drift rate, relates to the optimal decision theory for the constant drift case [Bogacz et al., 2006]. The present paper addresses this issue by offering analytically tractable approximations to two Bayesian inference models (compatibility bias and spatial uncertainty) proposed in [Yu et al., 2007].

Specifically, the joint signal probability distribution of Eq. (4) is approximated as a linear sum, and then, by assuming that the sum of the non-normalized posteriors remains close to one and taking a continuum limit, we obtain analytical expressions for the mean posterior probabilities. Employing a further approximation in which the net probabilities of having answered correctly or incorrectly at time t are computed, we derive semi-analytical approximations for accuracy and reaction time distributions. While the latter correspond more closely to an "interrogation protocol" [Bogacz et al., 2006,Liu et al., 2007] in which subjects are cued to respond at specific times, and so differ quantitatively from those computed numerically for free responses (compare Figures 10 with Figure 4), the overall accuracy curves and individual posteriors derived from the continuum model reproduce those of the Bayesian model quite well (see Figures 8-9).

We therefore expect that our analytical approximations will be useful in guiding parameter selection when fitting models to experimental data. In Section 3, we provide an example of this by deriving simple parametric constraints that must hold to obtain the dip below 50% in the posterior probability for early responses. Moreover, although the coefficients differ, the linearized update rules of both Eqs. (14) and (20) demonstrate that the flanker inputs x_1 and x_3 work *with* the target input x_2 for the compatible hypotheses, and *against* it for the incompatible hypotheses. This underlying computational architecture gives rise to the same basic ability of both the compatibility bias and spatial uncertainty models to account for the dynamics of flanker interference in behavioral data. In Section 4.5 we show that both the original models and DDM approximations derived from them can be fitted to experimental data, further strengthening our case.

Our analysis also reveals that a particularly simple stochastic differential equation, the constant-drift diffusion (DD) process of Eq. (35), approximately describes the evolution of Bayesian posteriors in log probability space. As described in [Bogacz et al., 2006], this is a continuum limit of the sequential probability ratio test [Wald, 1947], which is known to be optimal for identifying noisy signals in two-alternative choice tasks [Wald and Wolfowitz, 1948]. Moreover, it has been shown [Bogacz et al., 2006,Liu et al., 2007] that DD and related Ornstein-Uhlenbeck processes emerge naturally in linearized reductions of competing leaky accumulator models [Usher and McClelland, 2001] for 2AFC. In these neural networks the *difference* between activities in a pair of units at the output decision or response stage behaves like the accumulating variable $y(t)$ in Eq. (35)[1] [Gold and Shadlen, 2001]. DD models can also

capture bottom-up (stimulus-driven) and top-down influences such as attention and expectation of rewards via variable drift rates [Liu et al., 2007,Eckhoff et al., 2007]

Since accumulator models may be derived from biophysical models of spiking neurons [Wang, 2002, Wong and Wang, 2006], in which their activities represent short-term averages of collective firing rates, this suggests a mechanism by which neural substrates may be able to perform Bayesian computations. Specifically, in reducing the coupled Bayesian inference model (9) to a DD process we see how prior information maps into initial conditions, and evolving posteriors in log probability space are represented by spike rates of groups of neurons. In connection with the latter, we note that [Bogacz and Gurney, 2007] present computational and experimental evidence that Bayesian computations involving exponentiation and taking logarithms (cf. [Yu and Dayan, 2005]), as in Section 4, can be approximated by neurons in the basal ganglia.

Acknowledgments

This work was supported by PHS grants MH58480 and MH62196 (Cognitive and Neural Mechanisms of Conflict and Control, Silvio M. Conte Center). YL benefited from studentship support from the School of Engineering and Applied Science at Princeton University and AY received funding from an NIH NRSA institutional training grant. We thank the referees for perceptive and helpful comments.

Appendix

Appendix: Mathematical and data fitting details

A Evaluation of integrals

To evaluate the integrals of Eq. (39) we employ the change of variables

$$x=\frac{(\log(z)-\mu)}{\sqrt{2\sigma^2}},\ z=\exp\left(\mu+\sqrt{2\sigma^2}x\right), \tag{53}$$

so that $dx=\frac{dz}{z\sqrt{2\sigma^2}}$ and the integrals become

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\frac{-\mu}{\sqrt{2\sigma^2}}}\exp\left(-x^2+\mu+\sqrt{2\sigma^2}x\right)dx \quad \text{and} \quad \frac{1}{\sqrt{\pi}}\int_{\frac{-\mu}{\sqrt{2\sigma^2}}}^{\infty}\exp\left(-x^2\right)dx. \tag{54}$$

The second expression is a standard error function integral, and the first may be put into the same form by completing the square in the argument of the exponent:

$$x^2-\mu-\sqrt{2\sigma^2}x=\left(x-\sqrt{\frac{\sigma^2}{2}}\right)^2-\left(\mu+\frac{\sigma^2}{2}\right), \tag{55}$$

followed by the further change of variables

[1]In *N*-alternative choice models, linear combinations of variables approximate (*N* - 1)-dimensional DD processes [Usher and McClelland, 2001, McMillen and Holmes, 2006].

$$u=\left(x-\sqrt{\frac{\sigma^2}{2}}\right). \tag{56}$$

This process results in the expressions of Eq. (40).

To evaluate the integral of Eq. (49) we proceed as follows, dropping the explicit reference to time dependence, which enters the expressions through the mean and standard deviations $\mu(t)$, $\sigma(t)$. Figure 12 indicates the domain of integration.

$$\begin{aligned} P(s_2=i|\mathbf{X}_t)_{\text{est}} &= 1-\int_0^q\int_0^{q-z_2} p(z_1)\,p(z_2)\,dz_1dz_2 = 1-\int_0^q p(z_2,t)\int_0^{q-z_2} p(z_1)\,dz_1dz_2 \\ &= 1-\int_0^q p(z_2,t)\,\tfrac{1}{2}\left[1+\operatorname{erf}\left(\frac{\log(q-z_2)-\mu_1(t)}{\sqrt{2\sigma_1(t)^2}}\right)\right]dz_2 \\ &= 1-\tfrac{1}{2}\int_0^q p(z_2)\,dz_2-\tfrac{1}{2}\int_0^q p(z_2)\operatorname{erf}\left(\frac{\log(q-z_2)-\mu_1(t)}{\sqrt{2\sigma_1(t)^2}}\right)dz_2 \\ &= 1-\tfrac{1}{4}\left[1+\operatorname{erf}\left(\frac{\log(q)-\mu_2(t)}{\sqrt{2\sigma_2(t)^2}}\right)\right]-\tfrac{1}{2}\int_0^q p(z_2)\operatorname{erf}\left(\frac{\log(q-z_2)-\mu_1(t)}{\sqrt{2\sigma_1(t)^2}}\right)dz_2. \end{aligned} \tag{57}$$

Here we have added subscripts to the time-varying means and standard deviations $\mu_j(t)$, $\sigma_j(t)$, using the same shorthand $z_j = z_{1,j}$ as in §4.4 to indicate which of the four cases $s_2 = \pm 1$; $M = 1$, 2 enumerated in §4.3 is intended.

B Data fitting method

Data fits were performed using the fmincon() function in MATLAB. Parameters were determined by adjusting them while seeking minima of a error function, described by a weighted Euclidean norm, which averages over accuracy and RT data for both compatible and incompatible trials. The usual Euclidean (L^2) distance between vectors **u** and **v** with components u_j and v_j is

$$\|\mathbf{u}-\mathbf{v}\|=\sqrt{(u_1-v_1)^2+(u_2-v_2)^2+\ldots+(u_n-v_n)^2}. \tag{58}$$

Vectors describing accuracies and RT histograms were first formed from the data ($\mathbf{AC}_d$, $\mathbf{RT}_d$) and corresponding model predictions ($\mathbf{AC}_m$, $\mathbf{RT}_m$) were formed and their differences computed by (58). Since the units of accuracy and RT differ, each of these was then weighted by dividing it by the mean of the data, as indicated by an overbar below. This produces the nondimensional quantity:

$$\text{Error}=\sum_{\text{comp.,incomp.}}\left[\frac{\|\mathbf{AC}_d-\mathbf{AC}_m\|}{\|\bar{\mathbf{AC}}_d\|}+\frac{\|\mathbf{RT}_d-\mathbf{RT}_m\|}{\|\bar{\mathbf{RT}}_d\|}\right]. \tag{59}$$

This error term, representing the sum of percentage differences in accuracy and RT, was then minimized. Note that the resulting value depends on the number of RT bins in the data, and so should be normalized with respect to this when comparing fits of data sets with differing numbers of bins.

References

Bogacz R, Brown E, Moehlis J, Holmes P, Cohen J. The physics of optimal decision making: A formal analysis of models of performance in two alternative forced choice tasks. Psychological Review 2006;113(4):700–765. [PubMed: 17014301]

Bogacz R, Gurney K. The basal ganglia and cortex implement optimal decision making between alternative actions. Neural Computation 2007;19:442–477. [PubMed: 17206871]

Bogacz, R.; Hu, P.; Cohen, J.; Holmes, P. Submitted for publication. 2007. Do humans select the speed-accuracy tradeoff maximizing reward rate?.

Cohen J, Dunbar K, McClelland J. On the control of automatic processes: A parallel distributed processing model of the Stroop effect. Psychological Review 1990;97(3):332–361. [PubMed: 2200075]

Cohen J, Servan-Schreiber D, McClelland J. A parallel distributed processing approach to automaticity. American Journal of Psychology 1992;105:239–269. [PubMed: 1621882]

Eckhoff P, Holmes P, Law C, Connolly P, Gold J. On diffusion processes with variable drift rates as models for decision making during learning. New Journal of Physics. 2007??:?? In press

Eriksen B, Eriksen C. Effects of noise letters upon the identification of a target letter in a non-search task. Perception and Psychophysics 1974;16:143–149.

Gardiner, C. Handbook of Stochastic Methods. Vol. Second Edition. Springer; New York: 1985.

Gold J, Shadlen M. Neural computations that underlie decisions about sensory stimuli. Trends in Cognitive Science 2001;5(1):10–16.

Gold J, Shadlen M. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 2002;36:299–308. [PubMed: 12383783]

Gratton G, Coles M, Sirevaag E, Eriksen C, Donchin E. Pre- and poststimulus activation of response channels: a psychophysiological analysis. J. Exp. Psychol. Hum. Percept. Perform 1988;14:331–344. [PubMed: 2971764]

Gratton G, Coles MGH, Donchin E. Optimizing the use of information: The strategic control of the activation of responses. J. Exp. Psych. General 1992;121:480–506.

Higham D. An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM Rev 2001;43(3):525–546.

Holmes P, Shea-Brown E, Moehlis J, Bogacz R, Gao J, Aston-Jones G, Clayton E, Rajkowski J, Cohen J. Optimal decisions: From neural spikes, through stochastic differential equations, to behavior. IEICE Transactions on Fundamentals on Electronics, Communications and Computer Science 2005;E88A(10):2496–2503.

Laming, D. Information Theory of Choice-Reaction Times. Academic Press; New York: 1968.

Liu Y, Blostein S. Optimality of the seqeuntial probability ratio test for nonstationary observations. IEEE Transactions on Information Theory 1992;38(1):177–82.

Liu Y, Holmes P, Cohen J. A neural network model of the Eriksen task: Reduction, analysis, and data fitting. Neural Computation. 2007??:?? In press

McMillen T, Holmes P. The dynamics of choice among multiple alternatives. J. Math. Psych 2006;50:30–57.

Oksendal, B. Stochastic Differential Equations. Springer; New York: 2002.

Platt M, Glimcher P. Neural correlates of decision variable in parietal cortex. Nature 2001;400:233–238. [PubMed: 10421364]

Ratcliff R. A theory of memory retrieval. Psych. Rev 1978;85:59–108.

Ratcliff R, Smith P. A comparison of sequential sampling models for two-choice reaction time. Psychol. Rev 2004;111:333–46. [PubMed: 15065913]

Ratcliff R, Van Zandt T, McKoon G. Connectionist and diffusion models of reaction time. Psych. Rev 1999;106(2):261–300.

Roitman J, Shadlen M. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci 2002;22:9475–9489. [PubMed: 12417672]

Schall J. Neural basis of deciding, choosing and acting. Nature Reviews in Neuroscience 2001;2:33–42.

Schall J, Stuphorn V, Brown J. Monitoring and control of action by the frontal lobes. Neuron 2002;36:309–322. [PubMed: 12383784]

Servan-Schreiber D, Bruno R, Carter C, Cohen J. Dopamine and the mechanisms of cognition: Part I. A neural network model predicting dopamine effects on selective attention. Biological Psychiatry 1998;43:713–722. [PubMed: 9606524]

Shadlen M, Newsome W. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J. Neurophysiology 2001;86:1916–1936.

Usher M, McClelland J. On the time course of perceptual choice: The leaky competing accumulator model. Psych. Rev 2001;108:550–592.

Wald, A. Sequential Analysis. John Wiley & Sons; New York: 1947.

Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. Ann. Math. Statist 1948;19:326–339.

Wang X-J. Probabilistic decision making by slow reverberation in cortical circuits. Neuron 2002;36:955–968. [PubMed: 12467598]

Wong K-F, Wang X-J. A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci 2006;26:1314–1328. [PubMed: 16436619]

Yu A, Cohen J, Dayan P, Center for the Study of Brain, Mind and Behavior; Princeton University. A Bayesian view of sensory conflicts in decision-making. submitted to J. Exp. Psych. Human Perception and Performance. 2007Preprint

Yu, A.; Dayan, P. Inference, attention and decision in a Bayesian nueral architecture. In: Saul, L.; Yair, W.; Bottou, L., editors. Advances in Neural Information Processing Systems. Vol. 17. MIT Press; Cambridge, MA: 2005. p. 179-196.

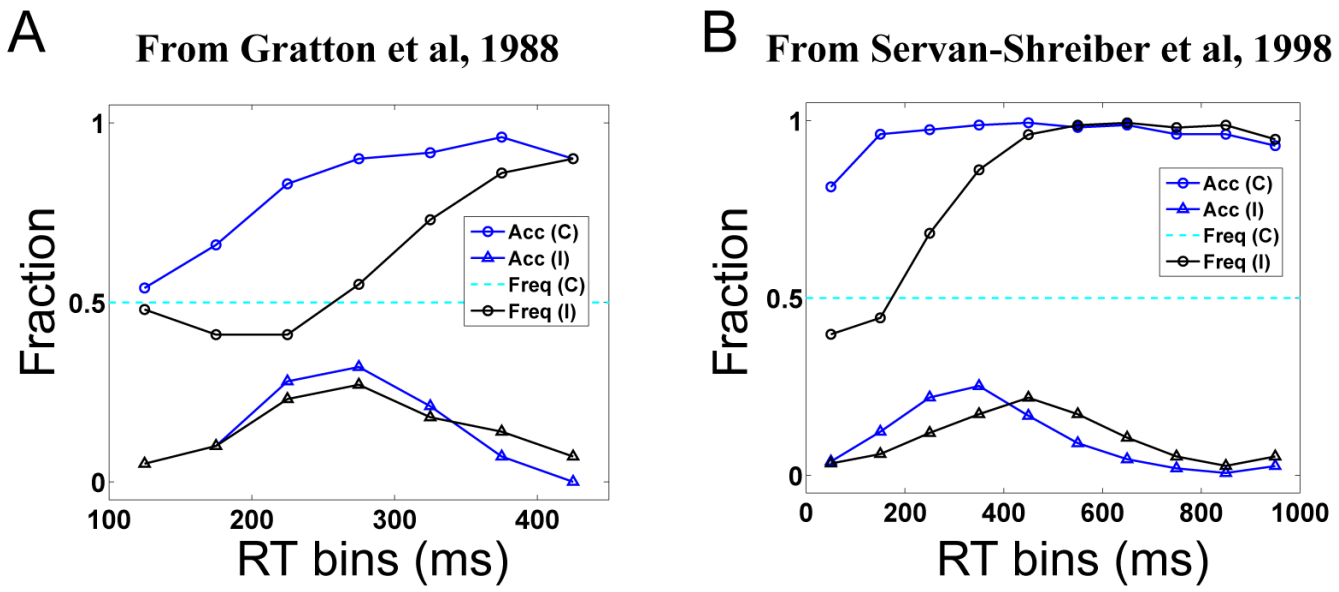

Figure 1.

Accuracy vs. RT in the Eriksen task. Human subjects respond slower and less accurately in the incompatible condition. In particular, accuracy is below chance (.50) for short RT's, but approaches 1 for longer RT's. (A) Reaction times gauged by electromyographic activities (EMG), adapted from [Gratton et al., 1988]. (B) Behavioral data from [Servan-Schreiber et al., 1998]. Details differ, but the compatibility effect and "dip" in accuracy for short-reaction incompatible trials, are obvious in both data sets.

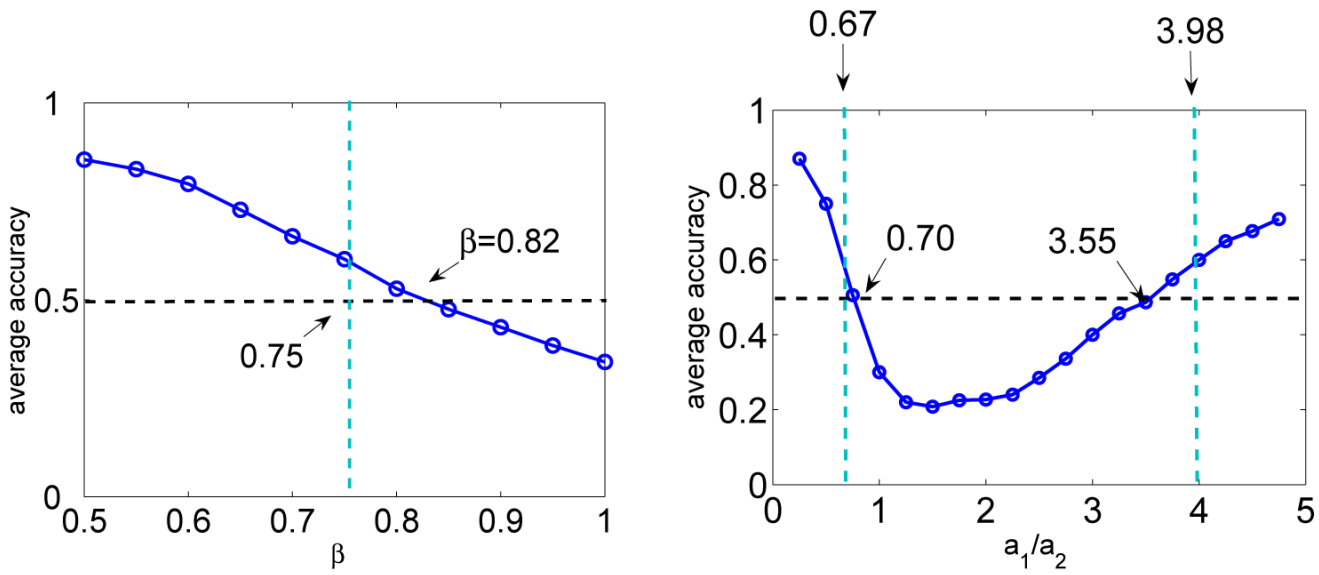

Figure 2.

Simulated and analytical approximations of parameter values that produce dips in accuracy vs. reaction time for incompatible trials. Graphs show accuracy averaged over trials with simulated reaction times under 20 timesteps, as a function of for the compatibility bias model (left), and the ratio of means a_1/a_2 for the spatial uncertainly model (right). Crossings with the 0.5 accuracy line indicate numerically obtained estimates of the "true" parameter constraints; dashed lines show the approximate constraints of Eqs. (18) and (24).

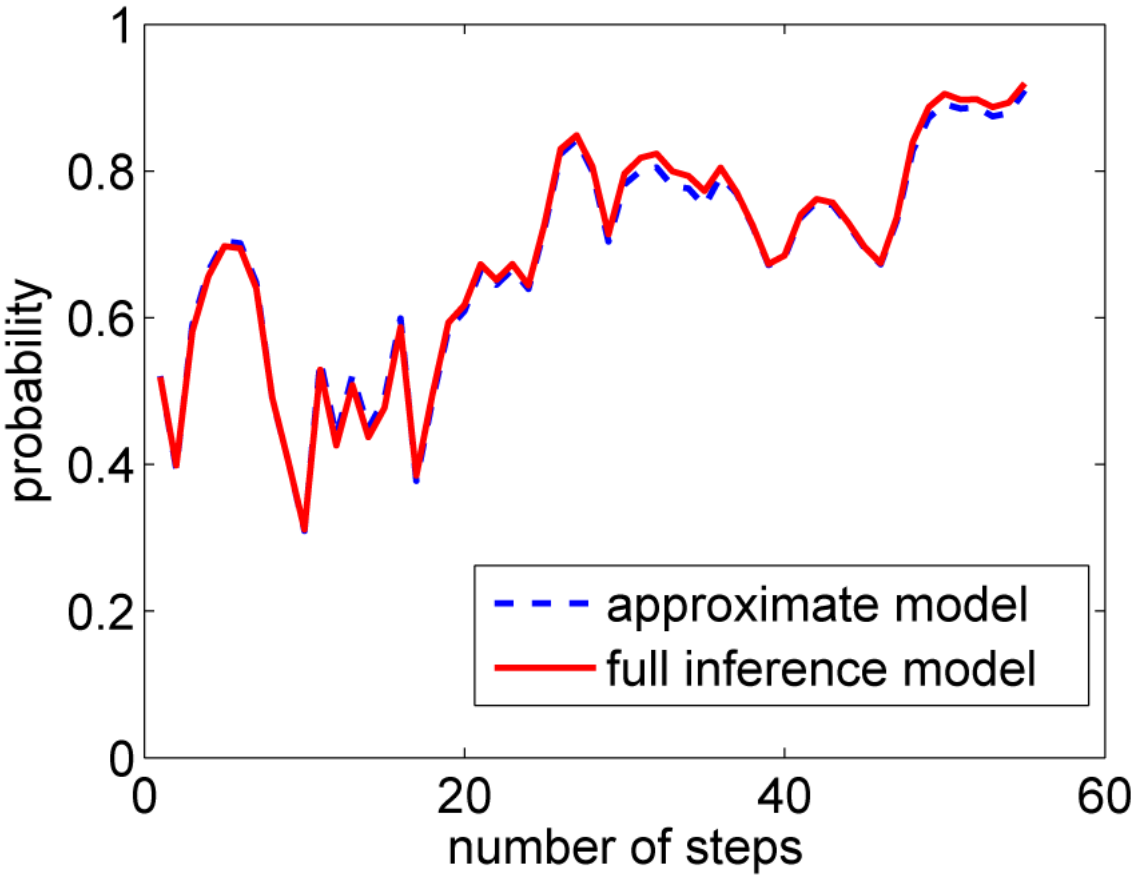

Figure 3.
Posterior probability $P(s_2 = 1|\mathbf{X}_t)$ for one sample path of the approximate compatibility bias model (Eq. (17), dashed), compared with a sample path from the original inference model (Eq. (9), solid). The same sequence $\mathbf{x}(t)$ of observations was used in both cases.

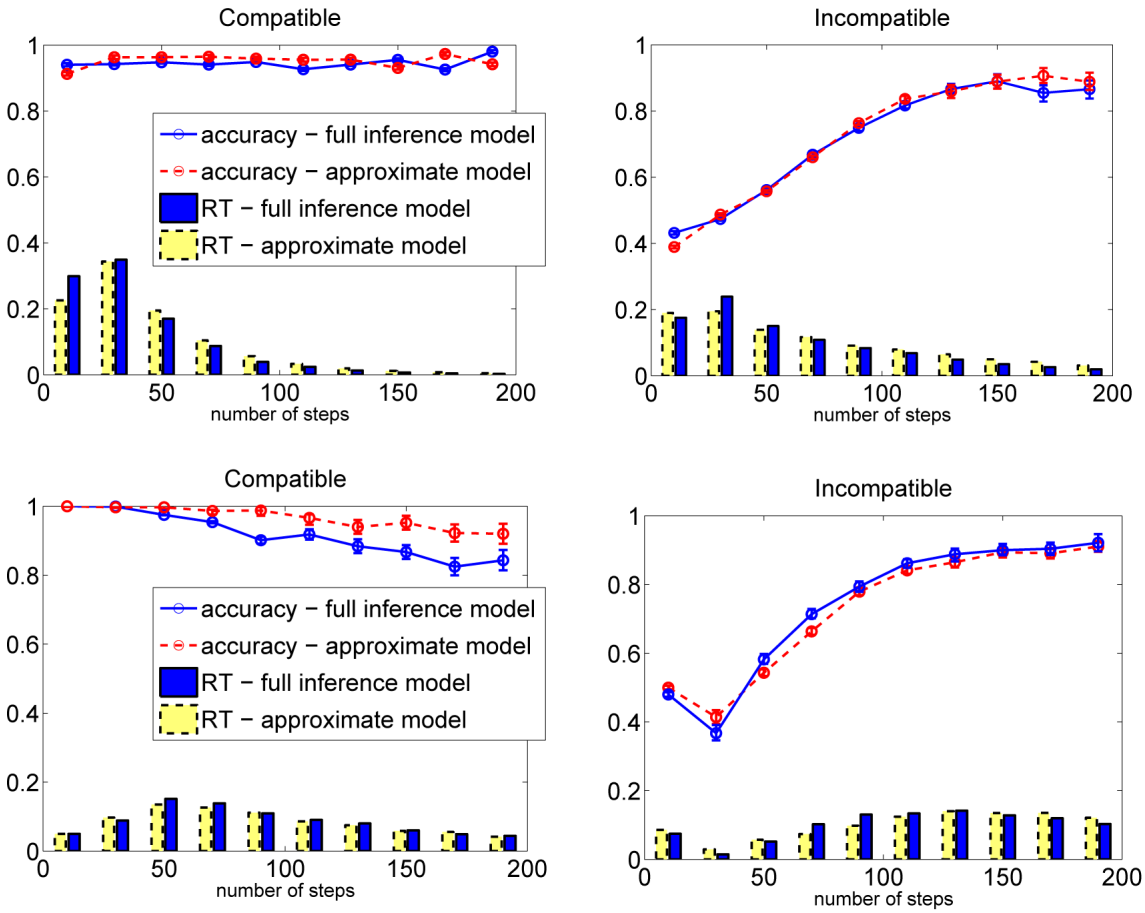

Figure 4.

Top panels: Accuracy and reaction time distributions for the compatibility bias model for compatible stimuli (left) and incompatible stimuli (right). Solid and right hand (blue) bar of each RT bin pair from full inference model of [Yu et al., 2007]; dashed and left hand (yellow) bars from approximate linearized likelihood model. Bottom panels: Accuracy and reaction time distributions for the spatial uncertainty model. Results obtained by averaging over 2, 000 simulated trials in each case.

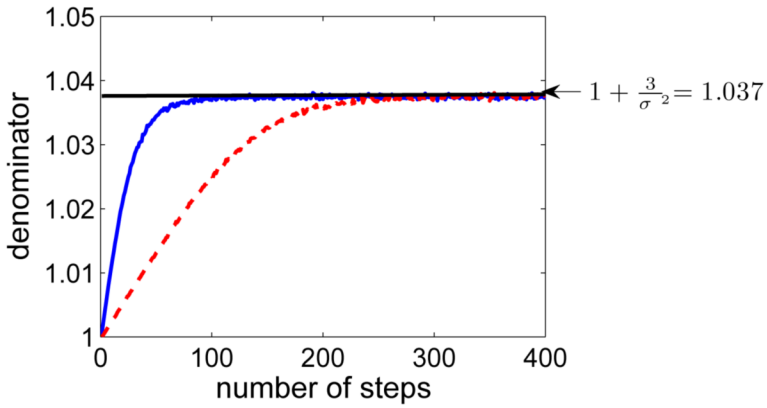

Figure 5.

Mean values of the denominator $\langle D_t \rangle$ for compatible (blue solid) and incompatible (red dashed) stimuli, each averaged over 10^5 trials. In both cases the $\langle D_t \rangle$ rises monotonically toward its upper bound $1+\frac{3}{\sigma^2}=1.0370\ldots$.

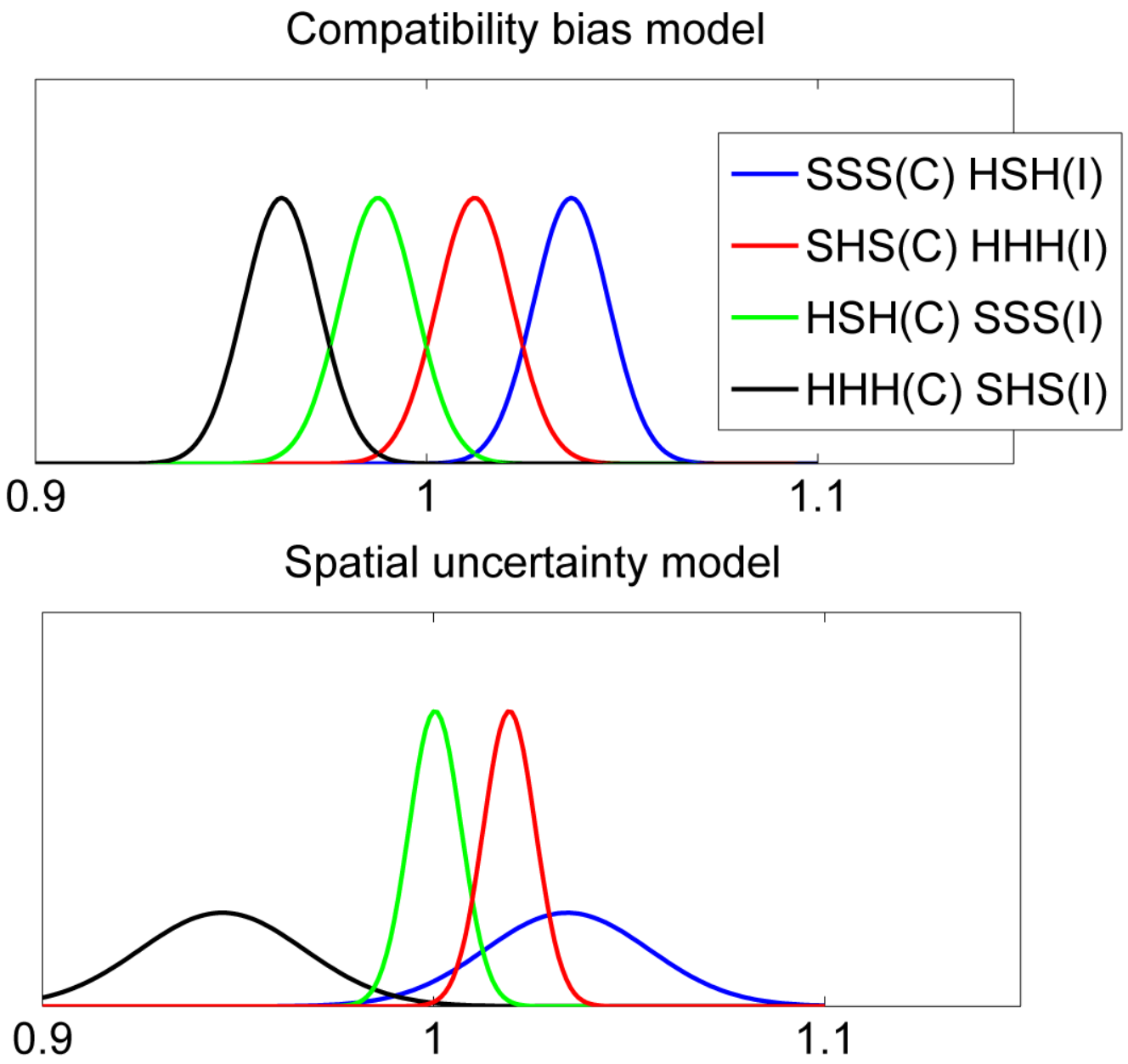

Figure 6.

Typical distributions from which the multiplicative factors $a_{i,j} + b_{i,j}\eta(t)$ in Eq. (28) are drawn on each time step. Parameter values are $\sigma = 1.8$ (top) and $a_1 = 0.7$, $a_2 = 0.3$, $\sigma_1 = 1.4$, $\sigma_2 = 1$ (bottom). For illustrative purposes, standard deviations σ, σ_1, σ_2 are 20% of those used in the text to reduce overlap of distributions.

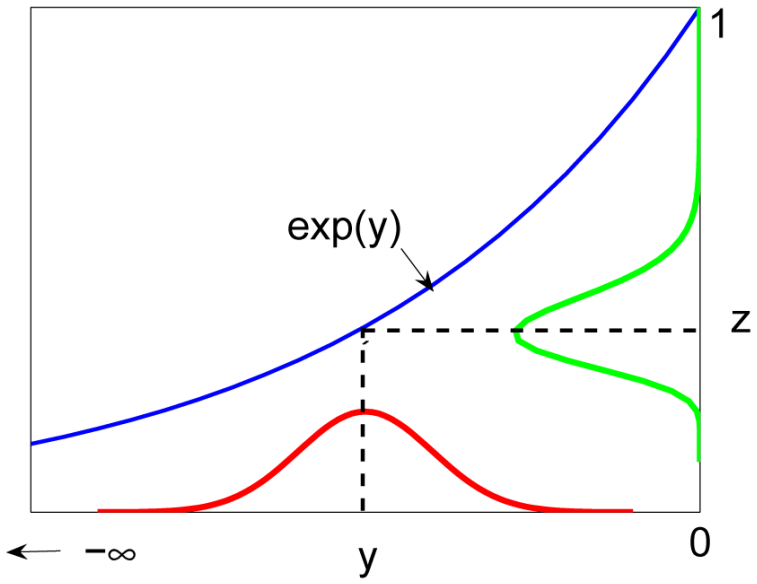

Figure 7.
Probability density functions in logarithmic *y*-space and the original *z*-space.

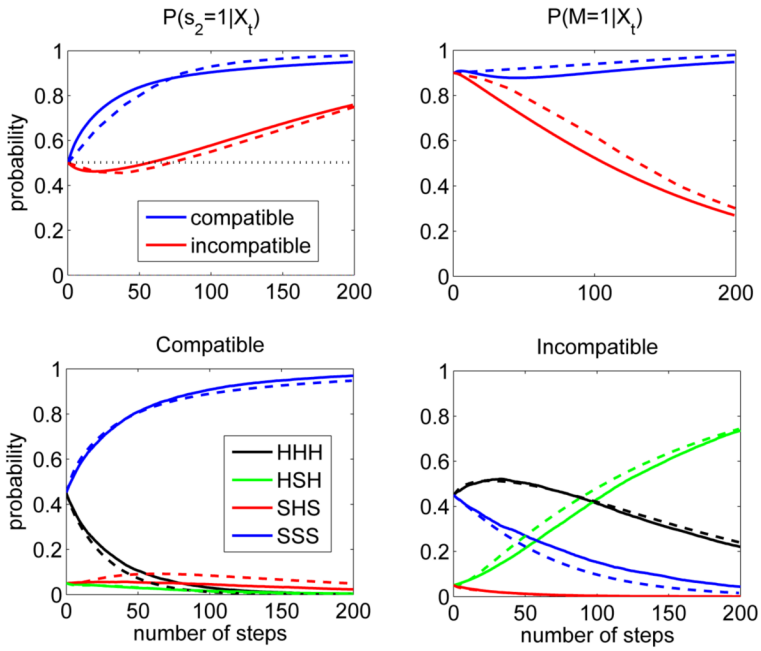

Figure 8.

Predictions of the full and simplified compatibility bias models in the case that the central symbol is S (s_2=1) and with prior compatibility bias $P(M)$=0.9. Top left: marginal mean posterior probabilities $P(s_2 = 1|M)$ (correct response) for compatible and incompatible conditions. Top right: marginal mean posterior $P(M = 1)$ for compatibility. Bottom row: individual mean posteriors for compatible (left) and incompatible (right) trials. Results from full inference model, averaged as in Figure 8, shown solid and predictions of the continuum approximation (41-43) shown dashed. Keys identify individual curves.

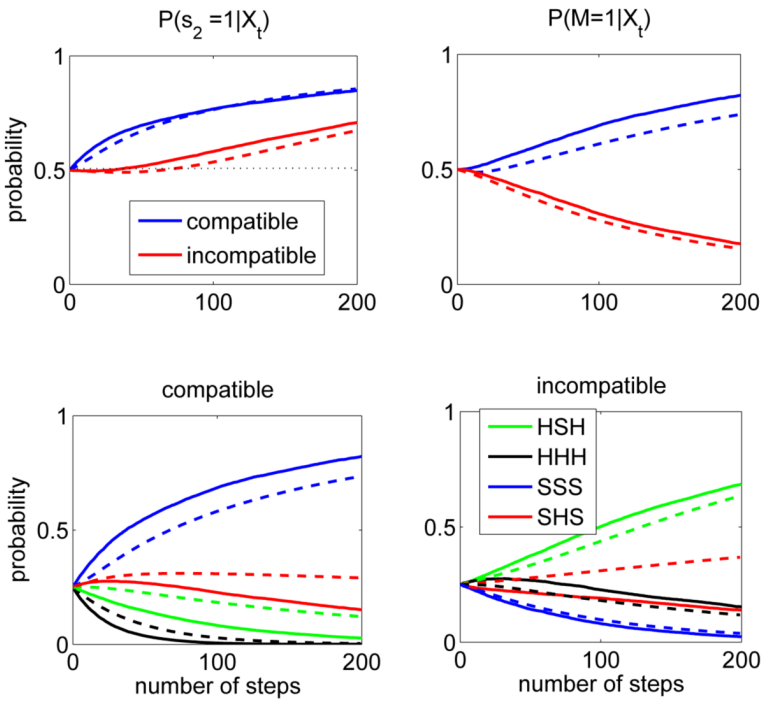

Figure 9.

Predictions of the full and simplified spatial uncertainty models. Top left: marginal mean posterior probabilities $P(s_2 = 1|M)$ (correct response) for compatible and incompatible conditions. Top right: marginal mean posterior $P(M = 1)$ for compatibility. Bottom row: individual mean posteriors for compatible (left) and incompatible (right) trials. Results from full inference model, averaged as in Figure 8, shown solid and predictions of the continuum approximation (41) and (44-46) shown dashed. Keys identify individual curves.

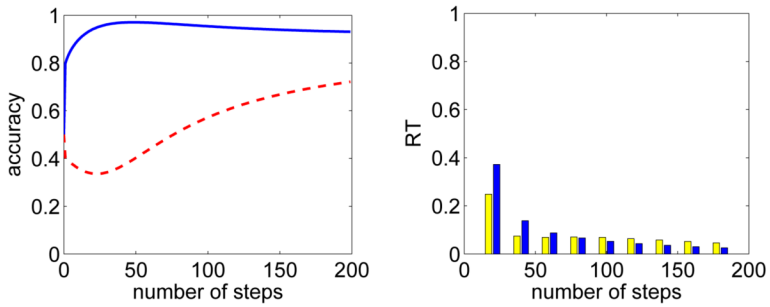

Figure 10.
Predictions of accuracy (left) and reaction time histograms (right) computed under the approximation of Section 4.4. Solid curve and dark boxes indicate compatible trials; dashed curve and light boxes indicate incompatible trials.

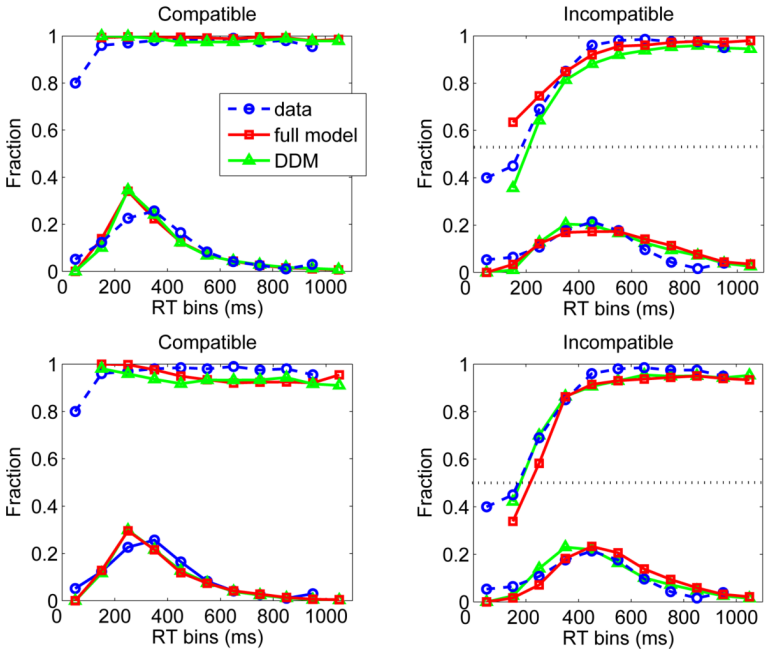

Figure 11.

Accuracy (upper curves in each panel) and reaction time distributions (lower curves) from the full (squares) and reduced DD (triangles) models for compatible (left) and incompatible (right) trials. Upper panels show compatibility bias and lower panels spatial uncertainty model results respectively. Parameters were fitted to the data of [Servan-Schreiber et al., 1998] (dashed curves with circles, cf. Fig 1B).

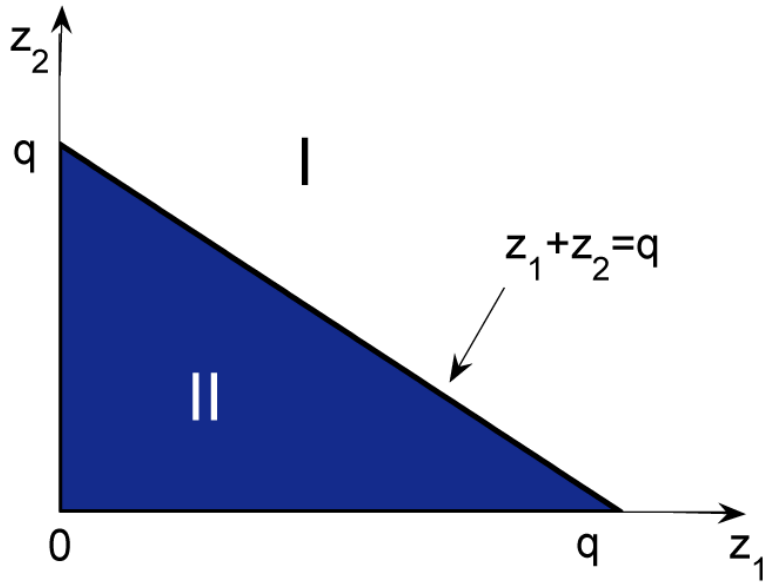

Figure 12.
The integral of the joint posterior probability distribution is taken over the positive (z_1, z_2)-quadrant less the shaded triangular region.