RUNNING HEAD: BAYESIAN FLANKER MODELS

A test of Bayesian observer models of processing in the Eriksen flanker task

Corey N. White[1], Scott Brown[2], Roger Ratcliff[1]

[1] The Ohio State University, [2] The University of Newcastle

Corresponding Author:
Corey N. White
Department of Psychology, University of Texas
5.316 Neural and Molecular Sciences Bldg.
2506 Speedway Blvd. Austin, TX
Email: white.1198@mail.texas.edu

Abstract

Two Bayesian observer models were recently proposed to account for data from the Eriksen flanker task, in which flanking items interfere with processing of a central target. One model assumes that interference stems from a perceptual bias to process nearby items as if they are compatible, and the other assumes that the interference is due to spatial uncertainty in the visual system (Yu, Dayan, & Cohen, 2009). Both models were shown to produce one aspect of the empirical data, the below-chance dip in accuracy for fast responses to incongruent trials. However, the models had not been fit to the full set of behavioral data from the flanker task, nor had they been contrasted with other models. The present study demonstrates that neither model can account for the behavioral data as well as a comparison spotlight-diffusion model. Both Bayesian observer models missed key aspects of the data, challenging the validity of their underlying mechanisms. Importantly, the results show that the shortcomings of the observer models stem from their assumptions about visual processing, not the use of a Bayesian decision process.

The Eriksen flanker task has been widely employed to explore visual processing and attention (Eriksen & Eriksen, 1974). In the basic task, an observer must identify a target item that is surrounded, or flanked, by other items that are to be ignored. Crucially, congruent trials include flankers that indicate the same response as the target, whereas incongruent trials indicate the opposite response. For example, if the observer had to identify whether a central target was a < or a >, a congruent trial would include >>> whereas an incongruent trial includes < > <. The standard finding is that incongruent flankers produce interference, leading to slower and less accurate responses compared to the congruent condition.

The influence of the incongruent flankers can best be seen by plotting the conditional accuracy function (CAF), in which accuracy is displayed for responses that fall within different bins (e.g., 300-325 ms). Figure 1 shows an idealized CAF that reflects the standard finding: responses for incongruent trials dip below chance for fast responses, but rise sharply towards asymptote for longer responses. This shows that the interference from the incongruent flankers is strongest early in the trial, but diminished later. This pattern can also be seen in the mean response times (RTs), with faster errors than correct responses for incongruent trials (e.g., White, Ratcliff, & Starns, 2011).

(Place Figure 1 about here)

Several models have been proposed to account for processing in the flanker task. The goal of the present study was to assess the validity of recent Bayesian observer models of flanker processing. Yu and colleagues (Liu, Yu, & Holmes, 2009; Yu et al., 2009) demonstrated that the qualitative patterns in the data could be produced by two variants of a Bayesian observer model. However, neither of these models was fit to the full set of behavioral data, thus it is uncertain whether they can adequately account for all of the data from the flanker task. Further, the models have not yet been compared to other extant models of flanker processing to assess the relative performance. To address this, the Bayesian observer models were fit to behavioral data from a simple flanker experiment and compared to a shrinking-spotlight diffusion model that was recently shown to account for flanker data across several experimental manipulations (White, et al., 2011). The next section introduces the models with focus on how they account for the effects of flanker interference.

Models of Flanker Processing

Four models of flanker processing were compared in the present study; a Bayesian spatial uncertainty model, a Bayesian compatibility bias model, a spotlight diffusion model, and a Bayesian spotlight model. Although the models have different structure, they all share the common assumption that noisy information is sampled and accumulated over time until a threshold is reached.

Bayesian Observer Models

The two Bayesian observer models proposed by Yu et al. (2009) share the same framework. For the remainder of the study, these models will be referred to as the observer models rather than the Bayesian models to emphasize that the Bayesian components of the models are not necessarily the most crucial to their performance (see Discussion). Both of the observer models assume that the stimulus produces noisy inputs to the perceptual system that are sampled over time. The observer accumulates these noisy samples and uses them to determine their belief about the identity of the stimulus. The left side of Figure 2 shows the basic structure of the models. The models assume that each item in the display produces noisy perceptual input, represented by the normal distributions in the left side of Figure 2. Leftward facing arrows produce input centered at -1 and rightward facing arrows produce input centered at +1. The variability of the perceptual inputs is determined by a free parameter, σ_1^2. At each unit of time (e.g., ~1 ms), a value is sampled from each item in the display and accumulated. The accumulated samples for each item determine the observer's belief about that item's identity. That is, what is the probability that this sample was obtained given a right or left arrow? Due to the noisy sampling, certainty about the item's identity increases as more samples are accumulated. The decision process is based on pattern-matching, and proceeds concurrently with the perceptual sampling to determine the observer's belief about which stimulus is being viewed. In essence, the observer decides at each time step whether the current perceptual information is more consistent with a display that is congruent with left target, congruent with right target, incongruent with left target, or incongruent with right target (see Figure 2). Then the decision involves summing over congruency to determine the relative probabilities that the display includes a left target (congruent or incongruent) or right target (congruent or incongruent). Finally, if either probability passes a threshold, q, then the corresponding response is selected.

In this framework, the observer updates their belief about the identity of the presented stimulus. The posterior probability in Equation 1 gives the ideal observer's belief about the target identity (s_{tar}) and stimulus congruency based on the sampled inputs (see Yu et al., 2009 for details).

$$P(s_{tar}, M|X_t) = \frac{p(x_t|s_{tar}, M)\ \ P(s_{tar}, M|X_{t-1})}{\sum_{s'_2 m'}\ p(x_t|s'_{tar}, M')\ \ P(s'_{tar}, M'|X_{t-1})'} \qquad (1)$$

(place Figure 2 about here)

In this equation, X_t indicates the vector of total inputs accumulated until time t, and s' refers to the other items in the display. The compatibility of the stimulus is denoted by M, where M=1 if the flankers are congruent and M=0 if they are not. In addition to providing belief about the identity of the target, Equation 1 also allows the observers to determine if the display is congruent or not. Although standard flanker tasks do not require decisions based on congruency, Yu and colleagues (2009) show how such decisions could be performed in this framework. Regardless, since the present study focuses on traditional flanker experiments where the decision is based on the identity of the target, the congruency calculation will not be discussed further. To make a perceptual decision about the target based on the posterior probability in Equation 1, the observer computes the marginal probability of s_{tar} being a left facing arrow (L) by summing over the uncertainty over compatibility. That is, the total probability of the display contained a left target is equal to the probability that the display is congruent with left target *or* incongruent with left target.

$$P(s_{tar}= L |X_t)=.5[P(s_{tar}= L , M=1|X_t)+P(s_{tar}= L , M=0|X_t)] \quad (2)$$

In Equation 2, .5 indicates the prior probability of the target facing left, which was 50% for the first experiment. Because the probabilities have to sum to 1, $P(s_{tar} = R|X_t)$ is given as $1-P(s_{tar} = L|X_t)$. A response is triggered whenever one of the two marginal probabilities exceeds a threshold value, q.

The two observer models are implemented in this general framework and share several components. The amount of evidence required for the response is reflected by the response threshold, q. The threshold can range from .5-1, and a higher value (e.g., .95) will lead to slower but more accurate responses because more evidence is required to make the decision. The variance parameter, σ_1^2, determines the amount of noise in the perceptual inputs. A large value of σ_1^2 means that many samples are required to overcome the perceptual noise and form an accurate belief about the stimulus. The weighting parameter, a_1, determines the relative weight of the perceptual inputs. A large value of

a_1 means that each sample of input contributes heavily to the calculation of the posterior odds.

The compatibility bias model assumes that our perceptual systems are biased to process nearby items as if they are compatible or congruent with each other. The bias parameter, β, can range from .5 (no bias) to 1 (complete bias), and leads to preferential weighting of the compatible stimulus arrays discussed above (Figure 2, top right). With a high value of β (e.g., .9), much more weight is given to the compatible than the incompatible stimulus. To see how this influences the decision process, we take an example trial with a left, incongruent stimulus (> > < > >). Early in the trial, the noisy perceptual input will favor (> > > > >) and (> > < > >) over (< < < < <) and (< < > < <) because of the greater total overlap between the patterns. If there is significant compatibility bias (high β), the congruent patterns will be weighted more heavily than the incongruent patterns. When coupled with the noisy perceptual information, the compatibility bias leads the observer's overall belief to be strongest for the congruent pattern that highly overlaps the stimulus (> > > > >) rather than the actual stimulus (> > < > >), which can lead to the incorrect response ("right"). The impact is most pronounced early in the trial when the sampled perceptual input is relatively poor, which leads to fast errors and the characteristic below-chance dip in accuracy for fast responses. As more inputs are sampled, the perceptual input eventually outweighs the bias and favors the correct pattern, leading to the correct response. Thus early responses for incongruent trials can dip below chance, consistent with the empirical CAFs. In the general framework of the observer models, the compatibility bias model is implemented by assuming that $\beta > .5$.

In contrast, the spatial uncertainty model assumes that each item in the display partially affects its neighbors due to overlapping spatial representations (Figure 2, bottom right). Because neural receptive fields can cover large areas of the visual field (up to 20 degrees, Desimone & Gross, 1979), some neurons will respond to both an item and its neighbors, resulting in overlap in the perceptual representations. Variations of this concept have been proposed previously in this domain (e.g., Logan, 1996) and in others (e.g., Gomez, Ratcliff, & Perea, 2008; Ratcliff, 1981; Ratcliff & Starns, 2009). With overlapping representations, the input for each item is a combination of the item itself and some portion of the neighboring items. For this model, compatibility bias is set at .5 (no bias), and parameters are added for the contribution of the neighboring items to the mean and standard deviation of the perceptual samples. Equation 3 gives the perceptual inputs for the spatial uncertainty model, where each x_i is the momentary perceptual sample from the item.

$$x_{outer}(t) \sim N(a_1\mu_{outer} + a_2\mu_{inner}, \sigma_1^2 + \sigma_2^2) \quad (3)$$

$$x_{inner}(t) \sim N(a_1\mu_{inner} + a_2\mu_{outer} + a_2\mu_{target}, \ \sigma_1^2 + 2\sigma_2^2)$$

$$x_{target}(t) \sim N(a_1\mu_{target} + a_2\mu_{inner} + a_2\mu_{inner}, \ \sigma_1^2 + 2\sigma_2^2)$$

Thus a_1 and σ_1^2 determine the relative contribution of the item itself to the perceptual input, and a_2 and σ_2^2 determine the contribution of its neighbors. As a_2 increases relative to a_1, the perceptual input from each item becomes more contaminated by its neighbors. The neighboring items also contribute to the variance of x_i by the parameter σ_2^2. Thus if a right arrow is neighbored by a left arrow, the distribution of perceptual input for the target will be on a value less than +1. That is, the contamination from the neighbor causes the target to appear less “rightward”. Consequently, the target input is less accurate for incongruent compared to congruent trials, increasing the probability of selecting the wrong response. Similar to the bias model, early errors in the incongruent condition arise because early perceptual input is relatively poor and influenced by neighboring items, but later responses are less influenced because more accurate perceptual information has been accumulated.

To allow the models to fit the data a scaling parameter, *s*, was added to translate the model steps into RT, and a nondecision time parameter, *Ter*, was added to account for processing time outside of the decision process. The scaling parameter was allowed to freely vary and set at a starting value of 2 based on a rough estimate from the data. The original models included a fast-guess parameter, γ, to account for the effects of a response deadline (see Yu et al., 2009). Since none of the experiments in this study employed such a deadline, γ was not included in the models.

Spotlight Diffusion Model

For comparison, the shrinking spotlight diffusion model of White et al. (2011) was included. A schematic of the model is shown in Figure 3. The model incorporates the principles of a shrinking spotlight into a diffusion model framework to produce predicted RT distributions and accuracy values. The standard diffusion model (Ratcliff, 1978, Ratcliff & McKoon, 2008) assumes that noisy information is sampled and accumulated until one of two boundaries is reached, similar to sampling in the observer models. Importantly, the standard diffusion model assumes a constant source of evidence during the decision, but this assumption is replaced in the spotlight diffusion model with time-varying evidence that is governed by the spotlight component. Thus similar to the observer models, the evidence driving the decision early in a trial is not necessarily the same as that later in the trial.

Spotlight or zoom-lens models assume that visual attention can be conceptualized as a spotlight in neural space (e.g., Eriksen & St. James, 1986), whereby any items within the spotlight get processed. The spotlight is represented as a normal distribution in Figure 3 to capture the idea that attentional resources drop off gradually near the edge of the spotlight. We assume that the spotlight is always centered properly on the target. The size of the spotlight at stimulus onset can be thought to result from attentional capture. Thus even though the instructions are to focus solely on the target, all of the items receive some attention early in the trial. Importantly, the size of the spotlight can be modulated by engaging attention. Over time the spotlight can be narrowed to focus on the target item, eliminating the interference from the flankers. In this framework, attention is diffuse at stimulus onset and thus the flankers can produce significant interference, leading to early errors. But as the spotlight shrinks, the interference is eliminated, leaving accurate information based on the target alone. White et al. (2011) incorporated these spotlight components into a diffusion model framework and demonstrated that it could account for flanker data from a number of different manipulations.

(place Figure 3 about here)

In the spotlight diffusion model, the total evidence that drives the decision (i.e., the drift rate) is a function of the input strength of each item, +- p (depending on the direction of the arrow), weighted by the proportion of attentional spotlight that is allocated to it. The spotlight component of the model is reflected by two parameters, sd_a and r_d. The initial width of the spotlight, sd_a, determines the relative contribution of the flankers and targets. A wide spotlight means that the flankers have more influence than the target, which leads to below-chance responses for incongruent trials. The parameter r_d indexes the rate or speed of attentional focusing. Once the decision process begins, the width of the spotlight decreases linearly by r_d at each time step. Thus the spotlight parameters determine the initial influence of the flankers and the speed at which their impact is reduced.

For simplicity, we define each item's region as one unit wide with the target centered at 0 (see Figure 3). To ensure that the total attentional area always summed to 1, any portion of the attentional distribution that exceeds the outer flanker range is allocated to the outer flanker. This also allows the models to capture potential edge effects, because a wide spotlight results in more attentional capacity allocated to the outer compared to the inner items. The original allocation of attention for the target, right inner flanker, and right outer flanker (left flankers are symmetrical), is given as:

$$a_{outer}(t)=\int_{1.5}^{\infty}\phi[0,sd_a(t)];\quad a_{inner}(t)=\int_{.5}^{1.5}\phi[0,sd_a(t)];\quad a_{target}(t)=\int_{-.5}^{.5}\phi[0,sd_a(t)] \tag{4}$$

In equation 4, ϕ is the density function for a normal distribution with mean 0 and standard deviation sd_a. Over time the width of the spotlight, sd_a(t), approaches 0, and thus the influence of the flankers is eliminated.

Although the spotlight diffusion model is formulated in a different framework than the observer models, it shares many of the same components. The diffusion model component assumes that noisy input is sampled until a criterial amount is reached, similar to the observer models. The response threshold is represented by the distance between the two boundaries, a. Similar to the threshold, q, for the observer models, a larger value of a leads to slower but more accurate responses. The input provided by each item in the stimulus is represented by $\pm p$ (depending on the direction of the arrow). Similar to the weighting parameter a_1 in the observer models, a large value of p indicates that each arrow provides strong evidence for the respective response, whereas a small value indicates weak evidence. The starting point, z, reflects response bias, but was fixed at $a/2$ for this study (no bias). The nondecision time parameter, *Ter,* is the same as for the observer models. The within-trial noise in the evidence accumulation, *s,* can act as a scaling parameter and was fixed at .1, consistent with previous diffusion model applications (Ratcliff & Tuerlinckx, 2002; but see Donkin, Brown, & Heathcote, 2009).

Bayesian Spotlight Model

An important distinction to note is that the observer models differ from the spotlight model in two regards: the decision mechanism and the assumptions about spatial attention. Thus differences among the models could stem from the decision mechanism rather than the attentional assumptions. A natural way to test this would be to implement the attentional assumptions from one model into the decision process of the other. However, the attentional components in the observer models are closely tied to the pattern-matching component of decision process, making them difficult to implement in a diffusion model. For example, the compatibility bias parameter exerts its influence during the pattern-matching components of the decision process, and thus can only be implemented when combined with pattern matching. Thus it is difficult to implement the mechanisms for compatibility bias and spatial overlap without retaining the pattern-matching component.

However, we argue that the pattern-matching component of the observer models is not consistent with the nature of the standard flanker task. That is, typical instructions are to identify the target while ignoring the flankers. With the pattern-matching framework, the decision is based on all the items in the display, even late in the decision process. Even if the stimulus has been presented for several seconds and the perceptual representation is completely stable and accurate, the participant would still base their decision on all of the items simultaneously rather than the just target. Further, if different stimuli are introduced, like (< > > > < ; see White et al., 2011), the pattern-matching becomes increasingly more complicated because a larger set of patterns needs to be assessed. Finally, pattern matching would be difficult to implement within the framework of a spotlight model because all of the items are not always in the focus of the spotlight. For example, after some amount of time the spotlight would narrow to include only the three inner items (target and inner flankers), and it is not clear how this input would be directly matched to the 5-item patterns.

Thus there are potential limitations to the pattern-matching aspect of the observer models. An alternative approach that many models of flanker processing assume, including the spotlight diffusion model, is that even though there is interference from the flankers, the participant attempts to base their decision solely on the target rather than the whole pattern of items in the display. In other words, although there is early interference from the flankers, attentional narrowing or selection allows for the decision to be based on the target alone for longer decisions (e.g., Eriksen & St. James, 1986; Hubner et al., 2010; White et al., 2011). We argue that this framework is more consistent with the task instructions, and thus we chose to implement a model that implements decisions in this manner rather than with pattern-matching.

The most straightforward model for this framework is one in which the spotlight attentional components are combined with the Bayesian probability decision process. We created a Bayesian spotlight model to test the relative impact of the attentional and decision components. Note that this model will mimic the spotlight diffusion model in many ways. The model uses the same spotlight components as the diffusion model, but the resulting decision evidence is fed into the probability-based framework used in the observer models. In brief, the spotlight component produces a value of the perceptual evidence (p_e) between -1 (only left arrows in the spotlight) and +1 (only right arrows in the spotlight), that changes over time as the spotlight is narrowed. This is identical to the component in the diffusion model, where the spotlight value is scaled into a drift rate by p. In the Bayesian spotlight model, this value is scaled into a probability by the parameter a_p, and accumulated in the same manner as for the original observer models. Equation 5 gives the scaling of perceptual evidence into decision evidence.

$$p(R \mid p_e) = .5 + (a_p * p_e) \qquad (5)$$

To retain the concept of noise in the accumulation process, the momentary evidence at each time step is sampled from a normal distribution with standard deviation, e (because the values are given as probabilities, they were truncated to remain between 0 and 1). Thus if the spotlight evidence at time t was .55 (more likely a right target), the actual value sampled in the decision process might be .48 or .57. At each time step, the momentary evidence is multiplied with the current total evidence to update the observer's belief about the target, just as with the other observer models. The Bayesian spotlight model includes the following parameters: response threshold (q), spotlight width (sd_a); narrowing rate

(r_d), probability scaling (a_p), nondecision time (*ter*), within-trial noise (*e*), and RT scaling (*s*).

Model Overview

Yu et al. (2009) showed that both the compatibility bias and spatial uncertainty models can produce the characteristic below-chance dip for incongruent trials in the CAF. Unfortunately, the data for the CAFs are noisy and highly variable because relatively few responses have terminated by the earliest bins (e.g., 300-325ms). Further, fitting the CAF does not guarantee good accord with all of the behavioral data. For example, Yu et al. (2009) chose a particular set of parameter values for the compatibility bias model that produced a below-chance dip in the CAF, and used this accordance as support for the viability of the model. However, those same parameter values predict accuracy near 65% for incongruent trials, which is substantially lower than the 80%-90% typically found (e.g., Gratton et al., 1988). Thus although those parameter values produced the correct form of the CAF, they did not produce the correct accuracy values. Because the early portion of the CAF relates to a small proportion of the data, fitting the CAF ignores the majority of the data.

Since neither of the observer models were fit to the full set of behavioral data, their validity is yet to be established. In contrast, the spotlight diffusion model was shown to account for the form of the CAF, the accuracy values, and the distributions of correct and error RTs for each condition across several experimental manipulations in the flanker task (White et al., 2011). Thus the spotlight model provides a good benchmark against which to compare the observer models. We used data from a basic flanker experiment to contrast the processing models. Each model was formulated for a five-item display to correspond to the experiment. In terms of model complexity, the spatial uncertainty and Bayesian spotlight models have seven free parameters, the bias model has six, and the spotlight model has five. Thus the observer and Bayesian spotlight models could be considered more flexible than the spotlight diffusion model, implying the need for measures to correct for differences in flexibility (e.g., Schwarz, 1978). However, the results will later show that the more-flexible models provide worse fits, precluding the need to correct for complexity.

Experiment

Data were taken from Experiment 1 of White et al. (2011), which involved a simple flanker task with an equal number of congruent and incongruent arrow stimuli. The authors fit the spotlight diffusion model to those data, and the best-fitting parameters and fit index from those fits are used for comparison with the observer models.

Method

Procedure

A standard flanker experiment was conducted in which participants were instructed to decide if the central arrow in a display of five arrows faced left or right, and to respond quickly and accurately. No other speed instructions were given. They were informed that the surrounding arrows might face the same or opposite direction as the target, but they were supposed to base their responses on the target only. Responses were collected from the keyboard, with the "/" key indicating a right-facing target, and the "z" key indicating a left-facing target. No error feedback was provided. The stimuli were presented uncued and remained on screen until a response was given, followed by 350 ms of blank screen until the next trial. The number of congruent and incongruent trials was equal, as were the number left and right target trials. Participants completed 48 practice trials followed by 8 blocks of 96 trials. Each block of trials had an equal number of left and right targets, and congruent and incongruent stimuli. The entire experiment lasted approximately 40 minutes.

Stimuli

Each stimulus array contained five arrows (< or >) displayed on top of each other in a vertical column (e.g., Figure 2, left), with the central (target) arrow in the center of the screen. Each arrow subtended .7 degree of visual angle, with .4 degree separation between the arrows. For congruent trials, the flankers faced the same direction as the target, whereas for incongruent trials the flankers faced the opposite direction. The flanking arrows were always symmetrically displayed around the target.

Participants

Twenty-five Ohio State undergraduates participated for credit in an introductory psychology course.

Results

Responses shorter than 300 ms or longer than 1500 ms were excluded from analyses (less than .9% of the data). Data were collapsed across trials with right and left facing targets. The accuracy values, mean RTs for correct and error responses, and number of observations for each condition are shown in Table 1. As expected, accuracy was lower and RTs were longer for incongruent compared to congruent trials.

(Place Table 1 here)

The data are graphically displayed in two ways. Figure 4 (top) shows the quantile probability function (QPF) from Experiment 1 (along with the predictions from the models). For clarity, the same data are replotted in each column of Figure 4 to show the relation to the predictions of each of the models. The QPF displays the accuracy values and the correct and error RT distributions simultaneously. The position on the x-axis indicates the probability of a response, with correct responses on the right and errors on the left. The points in the figure are the quantiles (.1, .3, .5, .7, .9) of the RT distributions, which provide a summary of the distribution shape. Thus the lowest point for a condition represents the fastest 10% of responses, or leading edge of the distribution, and the highest point represents the slowest 10% of responses, or tail of the distribution. For all of the QPFs presented in this study, data from congruent trials are represented by the column of circles nearest 1 for correct responses and the column nearest 0 for error responses. Due to the low number of errors for congruent trials, only the median quantile is presented in the graphs. Incongruent trials are represented by the columns nearer the center of the graphs. The QPFs show that the incongruent condition led to slower and less accurate responses, and the errors for incongruent trials were substantially faster than correct responses.

(place Figure 4 here)

The data are also displayed as CAFs with 25ms bins (e.g., 300-325ms, Figure 4 bottom). The CAF provides useful information about the relative speed of correct and error responses for each condition. However, while early responses for incongruent trials are below chance, overall accuracy for that condition is still around 90% because relatively few trials had terminated by 350 ms. Accordingly, data for the fastest bins are sparse and more variable. Further, not all participants had responses that terminated in the earliest bins. Together, these aspects of the CAF suggest that it should be used primarily to judge qualitative, rather than quantitative, patterns. It is also important to note that since CAFs display the relative proportions of trials that have terminated at different time points, they only provide an indirect display of the time-varying evidence that is driving the responses.

There is an apparent difference in the CAF from this experiment and some previous studies, namely that the earliest responses in Experiment 1 were not at chance (50%). Previous studies have used response deadlines and time pressure to increase the number of fast responses (e.g., Gratton et al., 1988). In fact, one of the original presentations of a CAF from this task showed responses for congruent and incongruent trials were at chance for the fastest RT bins, indicating that the speed pressure induced fast-guessing by the participants. Since the design of Experiment 1 relied on free responses, the CAFs in Figure 4 do not show effects of fast guessing. Accordingly, the earliest responses are not at 50%, and none of the models required mechanisms for fast guesses (see Yu et al., 2009).

(place Table 2 here)

Model Fitting

Each model was fit to the data using a SIMPLEX routine for χ^2 minimization. Fits were performed to both individual and grouped data with similar results, so results are presented for the averaged data. The data entered into the fitting routine were the accuracy values, RT quantiles (.1, .3, .5, .7, .9) for correct and error responses, and the number of observations for each condition (see Ratcliff & Teurlinckx, 2002). For each model, a set of parameter values was chosen, simulated data were generated (40,000 observations), the predicted quantile proportions were compared against the observed quantile proportions, and the difference was weighted by the number of observations to produce a χ^2 value. This χ^2 value was then minimized by the SIMPLEX routine. Each model was simulated several times with different starting values to ensure the results were not due to local minima in the parameter space.

This method of model fitting ensured that the models had to account for the entire data set from the experiment, including the accuracy values, RT distributions, and number of observations for correct and error responses. The χ^2 values from the fitting routine provide a measure of model fit. However, the χ^2 value is sensitive to the number of observations. The conditions in Experiment 1 have nearly 10,000 observations each, meaning that even a small misfit from the model would produce a significant χ^2 value. Further, because group data were fit rather than individuals, the resulting χ^2 values do not follow a typical χ^2 distribution (see Ratcliff & Smith, 2004). In light of this, the χ^2 values are used as a measure of relative fit quality among the models. In addition to the quantitative comparison based on χ^2, qualitative fit quality can be assessed the graphical display in Figure 4.

No parameters were allowed to freely vary between conditions, so the models would need to account for data from the congruent and incongruent conditions with only the perceptual input values differing (which were constrained depending on the direction of the arrows). The best fitting parameters for each model are shown in Table 2. The fit quality was significantly poorer for the both of the observer models compared to the spotlight models, both in terms of X^2 and qualitative correspondence with the data.

The model predictions are shown in Figure 4 for each model. As was shown in White et al. (2011), the predicted values for the spotlight diffusion model fall within the confidence ellipses in the QPF, showing good accord with the data. The CAF plots also show that the spotlight diffusion model captured the main trends in the data, showing below-chance accuracy for the fastest responses and a sharp rise to asymptote for the incongruent conditions. Note that the models are not fit to the CAF but rather the data in the QPFs (along with the number of observations in each condition). The Bayesian spotlight model also captured the main trends in the data, though the fit quality was numerically worse than the spotlight diffusion model.

The fit quality was much poorer for the observer models, particularly the spatial uncertainty model. The observer models fail to capture the qualitative patterns in the data, specifically the relative speed of correct and error responses for the incongruent trials, which can be seen in Figure 4. They fail primarily because the mechanism for improving evidence, accumulating more perceptual information through sampling, is relatively slow compared to the dynamic attention mechanism in the spotlight model. Consequently, if the models assume a large influence of the flankers early on (from a strong compatibility bias or large overlap), they capture the early dip in the CAF, but predict incongruent responses that are too inaccurate and slow because the perceptual sampling mechanism takes too long to overcome the early interference. Conversely, if the models assume a small influence of the flankers early on, the accuracy values are better predicted, but the models fail to produce the below-chance dip in the CAF, and they fail to predict significantly faster errors than correct responses for incongruent trials. As Figure 4 shows, the best fitting parameters from both observer models predict very little influence from the flankers. For the compatibility bias model, β was estimated at .72, which is slightly greater than the value necessary (.675) to produce a below-chance dip in the posterior odds (i.e., decision evidence) for a 5-item display (see Yu et al. 2009 for details). However, this slight dip in the posterior odds was not sufficient to produce below-chance responses for the incongruent trials when coupled with the other parameters. For the spatial uncertainty model, the overlap parameter (a_2) was relatively low as well, predicting about 20% overlap from the neighboring item that also fails to produce a below-chance dip in the CAF. Thus in addition to the quantitative disadvantage of the observer models, the failure of the observer models to capture the qualitative patterns in the data suggests significant limitations to the models (see Heinke & Backhaus, 2011).

Discussion

Overall, the observer models were unable to account for the accuracy values and the relative speed of correct and error responses for congruent and incongruent trials. The primary shortcoming of the models is their inability to produce high early interference and yet high overall accuracy for incongruent trials. As mentioned above, this is because the method of improving the decision evidence, sampling more noisy information, is too slow to overcome a large amount of influence in time to lead to accurate responses. In contrast, the spotlight models have a fast mechanism for improving the evidence (i.e., narrowing the spotlight), allowing for better accord with the data. Although there was a numerical advantage for the diffusion model, both versions of the spotlight model provided a better quantitative fit than the observer models, and importantly captured qualitative patterns that the observer models could not. Further, the relative success of the Bayesian spotlight model suggests that the shortcoming of the observer models is not due to their probability-based decision sampling.

Yu and colleagues (2009) did mention the possibility of adding a conflict-based mechanism in the models that could reduce flanker interference more effectively than the standard models. In brief, the observer might calculate the amount of conflict in the trial in a similar manner as the decision evidence, and then adjust their processing based on the conflict. Specifically, if the amount of conflict exceeds an observer-set threshold, the observer could switch to independent processing of each item in the display (i.e., the overlap is reduced to zero). Such a mechanism could improve the fit quality of the observer models by providing a stronger mechanism to improve the decision evidence (in addition to the original mechanism of sampling more inputs). However, the addition of a conflict mechanism would increase the complexity of the models and lead to a less parsimonious account of the flanker effect. Whereas the models tested in the present study attribute the effect to a single phenomenon (narrowing attention, compatibility bias, or spatial uncertainty), the conflict model would require two mechanisms (bias/overlap and conflict adaptation). Further, the conflict mechanism would result in a discrete, all-or-none improvement in decision evidence once the conflict threshold is reached, which would make it a variant of a dual process model. It has been recently shown that dual process models do not account for certain types of flanker data as well as single process models like the spotlight model presented in this study (White et al., 2011), though they might be more appropriate for situations involving more complex flanker stimuli (see Hübner et al., 2010).

Beyond Basic Flanker Effects

The present study employed very basic flanker stimuli and focused solely on congruent and incongruent trials. Consequently the results of this study do not address many of the nuances of visual attention and flanker processing. There is a rich literature on flanker effects that includes a myriad of experimental manipulations. For example, different stimuli (e.g., A or E) can be mapped onto the same response to disrupt the association between stimulus and response (e.g., Eriksen & Eriksen, 1974), item spacing and grouping can be manipulated to affect flanker interference (Eriksen & Eriksen, 1974; Hubner et al, 2010), and features other than visual proximity, like motion, can be used to affect flanker interference (Driver & Baylis, 1989). Further, neutral conditions (- - > - -) and more complex congruency conditions (> < < < >) can be employed to constrain models of flanker processing.

Thus the results of this study only speak to a small subset of flanker effects. However, the stimuli and conditions used in this study are still useful for assessing the models' ability to capture the basic flanker effects. The below-chance dip and relative speed of correct and error responses from the standard conditions in this study are robust effects that are relatively consistent across different manipulations. Thus while increasing the spacing between items might decrease the magnitude of the flanker interference, the overall pattern of early errors and late correct responses for incongruent trials is still observed (e.g., Hubner et al., 2010). Additional assumptions would have to be incorporated into the spotlight models to account for the full range of flanker effects (see White et al., 2011 for related discussion). However, the results of this study are still informative for models of flanker processing, as the inability of the observer models to account for the basic flanker effects suggests fundamental inadequacies in the assumptions of those models.

Finally, the models tested in this study could be expanded to incorporate trial-by-trial variability in the components. Standard applications of the diffusion model include across-trial variability in nondecision time, starting point, and drift rate to capture fluctuations of those values (see Ratcliff & McKoon, 2008). We purposefully excluded those parameters in our model comparison to focus on the primary components of each model. However, the fit quality would be improved by incorporating these variability parameters. White et al. (2011) took this approach and showed how the addition of these variability parameters improved the fit of the spotlight diffusion model. Importantly, the values of the primary model components were not greatly affected by the addition of the variability parameters, suggesting that there are no substantial parameter tradeoffs. Thus their exclusion in this study did not likely affect the overall conclusions.

Conclusion

In summary, the present study demonstrates that the observer models proposed by Yu et al. (2009) are unable to account for data from the standard flanker task. Although both the compatibility bias and spatial uncertainty models are capable of producing the characteristic below-chance dip in the CAF that is observed for incongruent trials, the models failed to account for the overall pattern of empirical data. In contrast, the spotlight diffusion model accounted for both the functional form of the CAF and the overall pattern of accuracy and RTs. We further showed that the spotlight component could be successfully implemented in a Bayesian probability decision process, suggesting that the shortcomings of the observer models stem from their assumptions about attentional processing. Thus the data do not support the observer models' assumptions of compatibility bias or spatial uncertainty in flanker processing. Importantly, the results of this study should not be taken to challenge the concepts of compatibility bias, spatial uncertainty, or Bayesian models of cognition in general, but rather their specific implementation to account for flanker data.

References

Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research, 178,* 363-380.

Driver, J. & Baylis, G. C. (1989). Movement and visual attention: the spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Psychophysics, 15*, 448-456.

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

Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143-149.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225-240.

Gratton, G., Coles, M. G. H., Sirevaag, E. J., & Eriksen, C. W. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 331-344.

Gomez, P., Ratcliff, R., & Perea, M. (2008). A model of letter position coding: The overlap model. *Psychological Review, 115,* 577-601.

Heinke, D. & Backhaus, A. (2011). Modelling visual search with the Selective Attention for Identification Model (VS-SAIM): A novel explanation for visual search asymmetries. *Cognitive Computation, 3*, 185-205.

Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review, 117,* 759-784.

Liu, Y. S., Yu, A., & Holmes, P. (2009). Dynamical analysis of Bayesian inference models for the Eriksen task. *Neural Computation, 21,* 1520-1553.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603-649.

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

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88,* 552-572.

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

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333-367.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116,* 59-83.

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

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

White, C. N., Ratcliff, R., & Starns, J. J (2011). Diffusion models of the flanker task: Discrete versus gradual attentional narrowing. Manuscript in revision.

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 700-717.

Figure Captions

Figure 1. Idealized conditional accuracy function from a flanker task (see text).

Figure 2. Bayesian observer models of flanker processing. Left: General framework for both models. Noisy perceptual input from each item is sampled and used to perform pattern-matching. As more samples are accumulated, the observer's belief about the items' identities becomes more certain. The decision is based on the total probability of the display having a left target (congruent or incongruent) versus the probability of a right target. Right: how the two versions of the observer models are implemented in the general framework. The compatibility bias model (top right) assumes that preferential weight is given to compatible stimuli ($\beta > .5$), essentially increasing the evidence for those stimulus patterns in the decision process. The spatial uncertainty model assumes that neighboring items interfere with each other (bottom right), with the amount contributed by the neighbors to the mean and variance of the item's input denoted by a_2 *and* ∂_2*, respectively*.

Figure 3. Spotlight diffusion model. Left: Attention is represented by a spotlight over the items, which can be narrowed on the target to improve performance. Right: The total evidence from the spotlight component drives a diffusion process, where evidence is sampled over time until a boundary is reached. The curved arrow represents the decision evidence that changes over time. In the diffusion process, *a* is the total distance between the boundaries, and *z* is the starting point of evidence accumulation.

Figure 4. Behavioral data and predictions from the best-fitting parameters for each model. Each

column shows the same behavioral data plotted with the predictions of a specific model. The top panel shows quantile probability functions with 95% confidence ellipses that were constructed by bootstrapping the data (see White et al., 2011). The bottom panel shows condition accuracy functions.

Table 1. Accuracy and reaction times averaged across subjects for Experiment 1

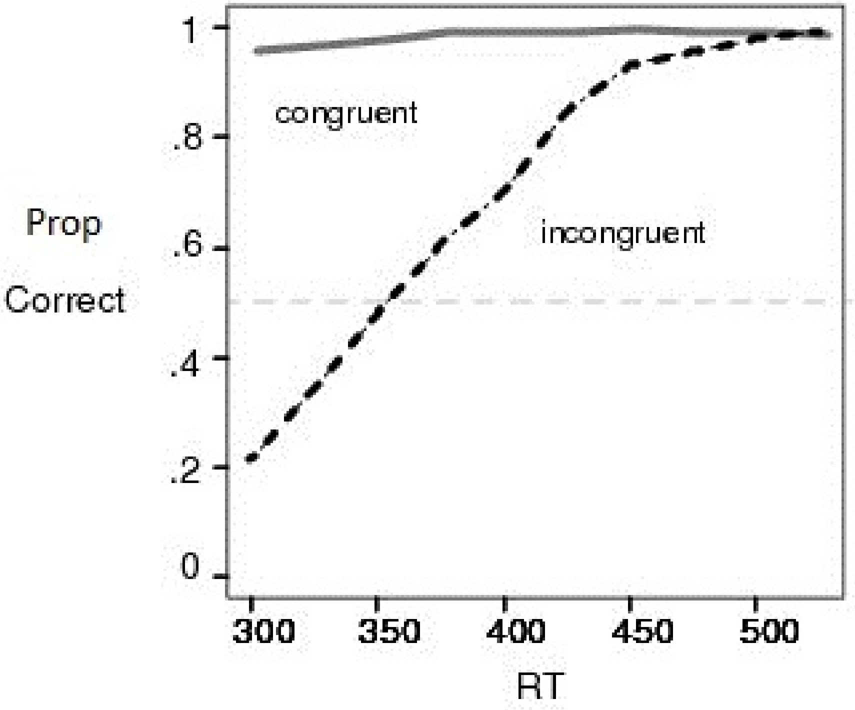
Condition	Accuracy	Correct RT	Error RT
Congruent	.985 (.01)	466 (32)	473 (89)
Incongruent	.909 (.07)	544 (34)	408 (31)

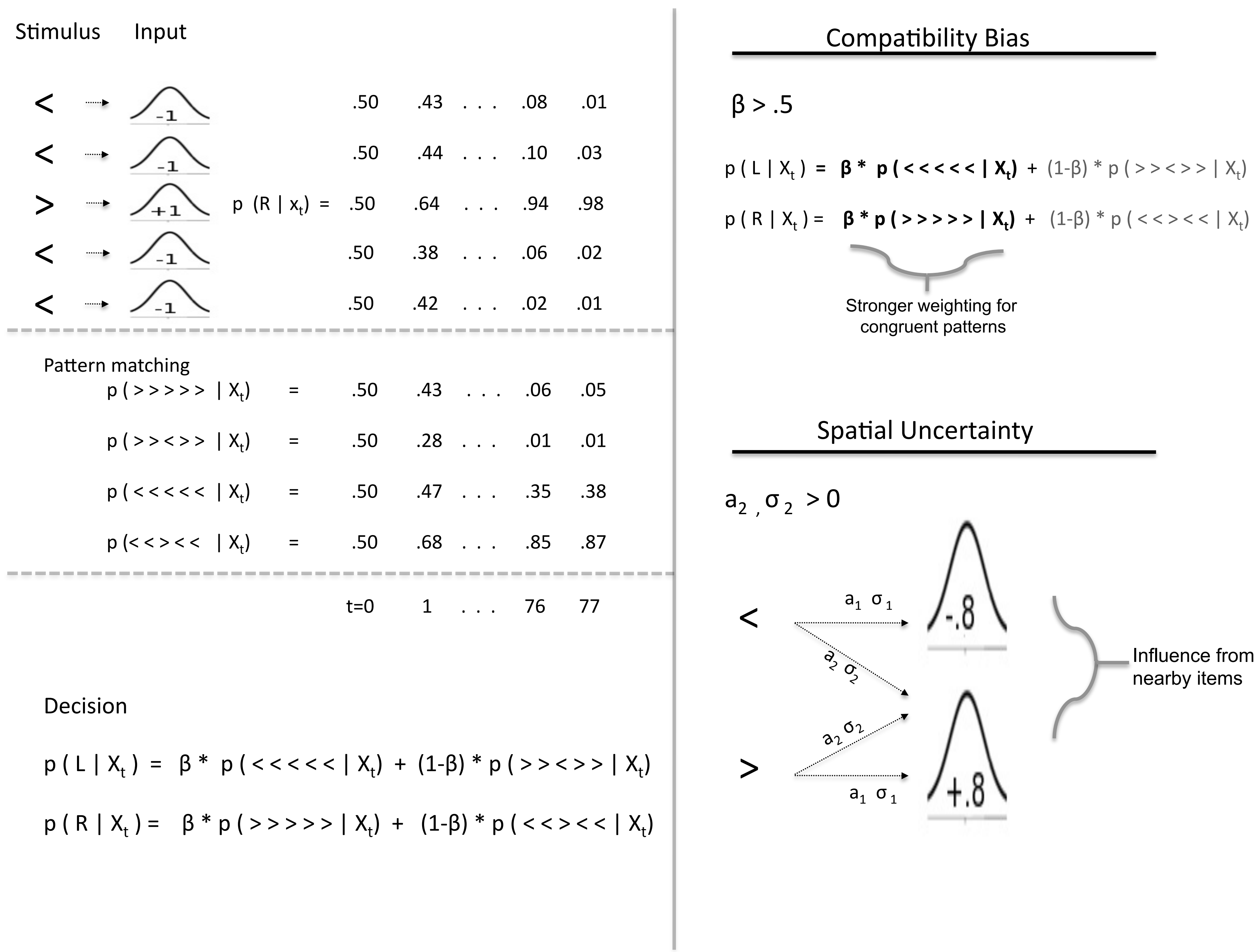
Note. SDs are shown in parenthesis

Table 2. Best fitting model parameters for Experiment 1.

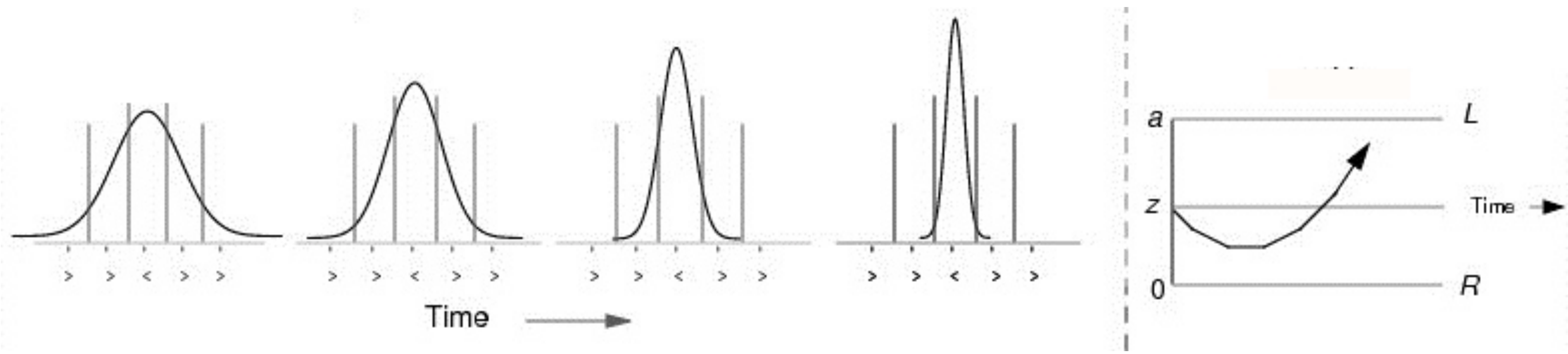
Spotlight Diffusion	a	Ter	p	r_d	sd_a			X^2
	.129	301	.383	.018	1.86			631.5
Compatibility Bias	q	Ter	B	a_1	σ_1	s		
	.936	314.3	.726	1.12	8.91	2.00		1552.5
Spatial Uncertainty	q	Ter	a_1	a_2	σ_1	σ_2	s	
	.939	288.1	1.53	.348	6.19	3.60	2.70	4411.1
Bayesian Spotlight	q	Ter	a_p	r_d	sd_a	e	s	
	.958	305.1	.011	.033	1.53	.063	2.24	658.8

Note. a = boundary separation; Ter = mean nondecision time; p = perceptual input; sd_a = spotlight width; r_d = rate of decrease in spotlight; q = response threshold; B = compatibility bias; a_1 = weight for each item; a_2 = weight for neighboring items; σ_1 = variance for item; σ_2 = variance for neighboring items; a_p = probability scaling; e = within-trial variability; s = RT scaling (see text for details).

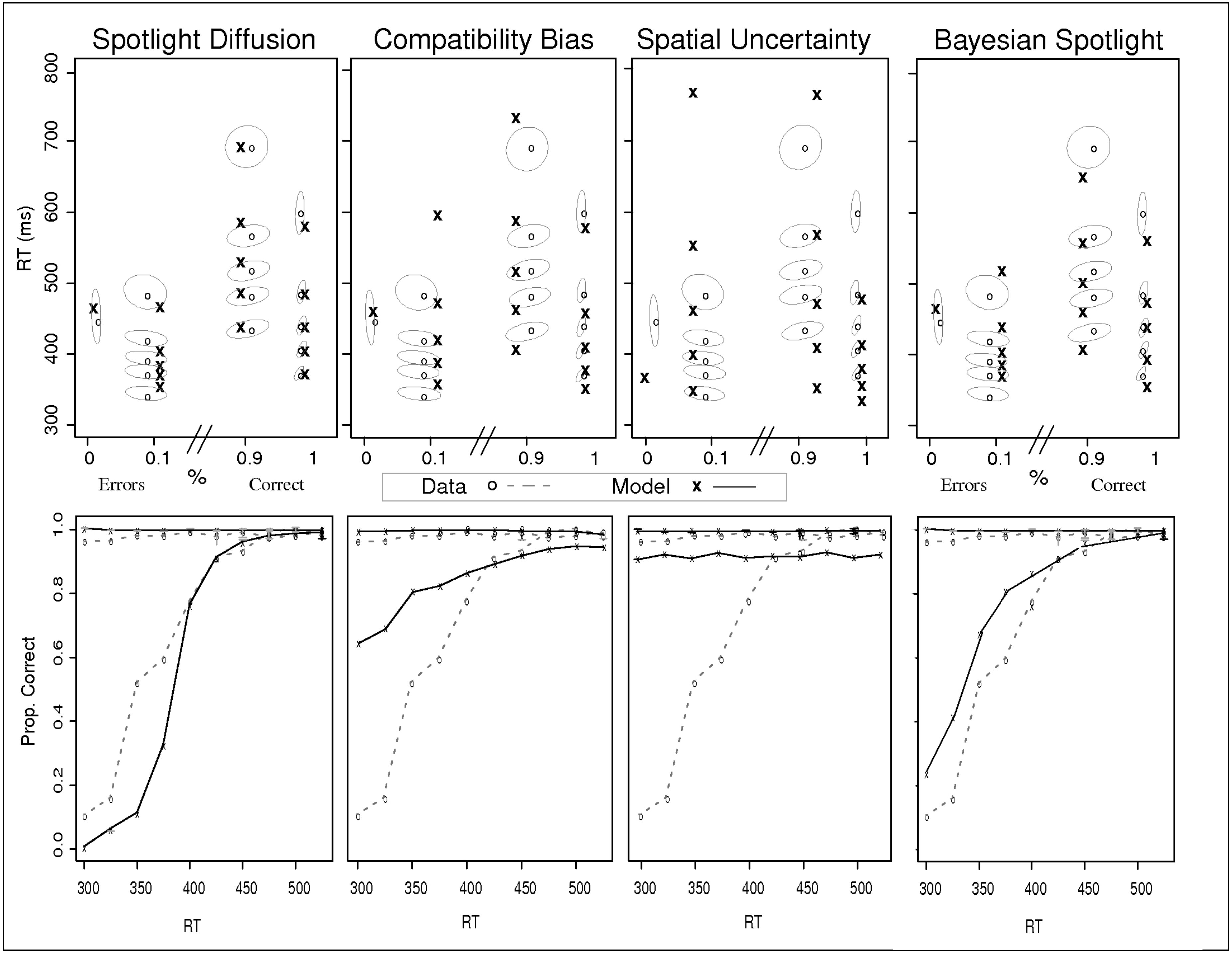

1
.8
.6
.4
.2
0
Prop
Correct
congruent
incongruent
300
350
400
450
500
RT



Stimulus
Input
<
-1
.50 .43 . . . .08 .01
<
-1
.50 .44 . . . .10 .03
>
+1
p (R | x_t) = .50 .64 . . . .94 .98
<
-1
.50 .38 . . . .06 .02
<
-1
.50 .42 . . . .02 .01
Pattern matching
p (> > > > > | X_t) = .50 .43 . . . .06 .05
p (> > < > > | X_t) = .50 .28 . . . .01 .01
p (< < < < < | X_t) = .50 .47 . . . .35 .38
p (< < > < < | X_t) = .50 .68 . . . .85 .87
t=0 1 . . . 76 77
Decision
p (L | X_t) = β * p (< < < < < | X_t) + (1-β) * p (> > < > > | X_t)
p (R | X_t) = β * p (> > > > > | X_t) + (1-β) * p (< < > < < | X_t)
Compatibility Bias
β > .5
p (L | X_t) = β * p (< < < < < | X_t) + (1-β) * p (> > < > > | X_t)
p (R | X_t) = β * p (> > > > > | X_t) + (1-β) * p (< < > < < | X_t)
Stronger weighting for
congruent patterns
Spatial Uncertainty
a_2 , σ_2 > 0
<
a_1 σ_1
-.8
a_2 σ_2
a_2 σ_2
>
a_1 σ_1
+.8
Influence from
nearby items



> > < > >
> > < > >
> > < > >
> > < > >
Time
a
L
z
Time
0
R



Spotlight Diffusion
Compatibility Bias
Spatial Uncertainty
Bayesian Spotlight
RT (ms)
Errors
%
Correct
Data
Model
Prop. Correct
RT