

A linear cue combination framework for understanding selective attention

Richard F. Murray

Department of Psychology, University of Toronto,
Toronto, Canada



Allison B. Sekuler

Department of Psychology, McMaster University,
Hamilton, Canada



Patrick J. Bennett

Department of Psychology, McMaster University,
Hamilton, Canada



Using a linear cue combination framework, we develop a measure of selective attention that describes the relative weight that an observer assigns to attended and unattended parts of a stimulus when making perceptual judgments. We call this measure *attentional weight*. We present two methods for measuring attentional weight by calculating the trial-by-trial correlation between the strength of attended and unattended parts of a stimulus and the observer's responses. We illustrate these methods in three experiments that investigate whether observers can direct selective attention according to contrast polarity when judging global direction of motion or global orientation. We find that when observers try to judge the global direction or orientation of the parts of a stimulus with a given contrast polarity (white or black), their responses are nevertheless strongly influenced by parts of the stimulus that have the opposite contrast polarity. Our measure of selective attention indicates that the influence of the opposite-polarity distractors on observers' responses is typically 65% as strong as the influence of the targets in the motion task, and typically 25% as strong as the targets in the orientation task, demonstrating that observers have only a limited ability to direct attention according to contrast polarity. We discuss some of the advantages of using a linear cue combination framework to study selective attention.

Keywords: selective attention, contrast polarity, global motion, texture, signal detection theory

Introduction

When we make visual judgments about a scene, we can base our judgments on selected parts of the scene, and ignore other parts. This ability is called *selective visual attention*. We can direct visual attention according to simple stimulus properties, such as spatial location (Posner, Snyder, & Davidson, 1980), color (Brawn & Snowden, 1999), direction of motion (Ball & Sekuler, 1981), and spatial frequency (Davis & Graham, 1981), and perhaps also according to more complex criteria, such as the perceptual segmentation of a scene (Baylis & Driver, 1992; Duncan, 1984; Egly, Driver, & Rafal, 1994). However, selective attention is sometimes imperfect: if targets and distractors differ along certain dimensions, we find that even when we try to attend only to the targets, our judgments are nevertheless influenced by the distractors. This raises the question of how targets and distractors together determine an observer's responses, and the closely related question of how we should measure intermediate degrees of selective attention.

The problem of how observers combine information from two or more sources to arrive at a single response has a long history in perceptual psychology (Anderson, 1974). One particularly simple hypothesis is that

observers calculate a weighted sum of internal responses to individual sources of information. Such weighted sum models have been used to describe how observers perform many different tasks, including detecting an auditory signal with two frequency components that activate different auditory channels (Green, 1958), combining redundant stimulus properties in complex figures (Kinchla, 1977), combining multiple depth cues (Landy, Maloney, Johnston, & Young, 1995), and combining information across different senses (Ernst, Banks, & Bülthoff, 2000; Jacobs, 1999). Applied to the problem of selective attention, the weighted sum hypothesis suggests that if T is an internal response to targets and D is an internal response to wholly or partly unattended distractors, then the observer bases his responses on a decision variable of the form

$$s = T + kD . \quad (1)$$

The weighting factor k measures the influence of the distractors on the observer's responses, and we will call it the *attentional weight* that the observer assigns to the distractors.

Here we investigate some theoretical and empirical aspects of this weighted sum theory of selective attention. First, we discuss why we might expect selective attention to work this way. We present a general Bayesian

description of how observers perform discrimination tasks, and we show that in many circumstances, it is entirely natural for observers to combine information from attended and partly unattended sources in a weighted sum, as in Equation 1.

Second, we derive two methods for measuring the attentional weight k assigned to distractors in a wide range of tasks, and we show that these methods work even when we do not know how the observer computes the internal responses T and D to the targets and distractors. We illustrate these methods in three experiments that investigate whether observers can direct selective attention according to contrast polarity when judging global direction of motion, or when judging global orientation. Several recent studies have investigated the first question concerning global motion and have given conflicting results (Croner & Albright, 1997; Edwards & Badcock, 1994; Li & Kingdom, 2001; Snowden & Edmunds, 1999; van der Smagt & van de Grind, 1999). The methods that we introduce avoid some of the problems of these earlier studies, and so we hope to give a more convincing answer to the question whether observers can direct attention according to contrast polarity.

Third, we test an assumption that is implicit in the weighted sum hypothesis, namely that selective attention only affects the relative weight that an observer assigns to the internal responses to the targets and distractors, T and D , without changing the internal responses themselves. This issue is crucial for the problem of how to measure selective attention. If selective attention affects only the relative weight assigned to targets and distractors, then it can be described by a scalar, such as attentional weight. On the other hand, if selective attention qualitatively changes how an observer computes the internal responses T and D , then a more complex description may be necessary. We show how methods developed by Chubb and colleagues (Chubb, 1999; Chubb, Econopouly, & Landy, 1994) can be used to investigate how observers process attended and unattended stimuli, and we illustrate these methods by measuring directional selectivity for attended and partly unattended motion signals in a global direction discrimination task.

We begin with the question of why selective attention might take the form of a single weighting factor.

Why Attentional Weight?

When studying human performance in a perceptual task, it is often revealing to model observers as Bayesian decision-makers who are limited by simple degradations of the stimulus or by imperfect knowledge of the stimulus. For instance, in many shape discrimination tasks, human observers behave like Bayesian observers who view stimuli through a small amount of additive Gaussian noise and have an imperfect representation of the shapes to be discriminated (Barlow, 1956; Lu &

Dosher, 1998; Pelli, 1990). Bayesian models are often illuminating, because they make explicit claims about what information observers use to perform a task, and about what types of inefficiencies limit observers' performances (Geisler, 1989; Watson, 1987). We follow a similar approach to define a measure of selective attention.

Consider a task in which the observer discriminates between two classes of stimuli, A and B . A Bayesian decision-maker performs this task by viewing the stimulus U on each trial, and evaluating the probability that the stimulus was drawn from class A or class B , given that the observed stimulus was U . Bayes' theorem shows that these probabilities are

$$P(A|U) = \frac{P(U|A)P(A)}{P(U)} \quad (2)$$

$$P(B|U) = \frac{P(U|B)P(B)}{P(U)} \quad (3)$$

Equivalently, the observer can base his responses on the likelihood ratio L :

$$L = \frac{P(U|A)}{P(U|B)} \quad (4)$$

If stimulus types A and B appear equally often, and if the observer's goal is to maximize the number of correct responses, then the optimal strategy is to respond 'A' if $L \geq 1$, and 'B' otherwise (Green & Swets, 1974).

If the stimulus U is composed of many independently varying elements U_i (e.g., a noisy N pixel stimulus, or a random dot cinematogram with N independent dot displacements), then the likelihood ratio L is the product of many subsidiary likelihood ratios u_i computed from the stimulus elements U_i :

$$L = \prod_{i=1}^N u_i, \text{ where } u_i = \frac{P(U_i|A)}{P(U_i|B)} \quad (5)$$

Equivalently, the observer can calculate the logarithm of this likelihood ratio, which is the sum of the subsidiary log likelihood ratios:

$$\ln L = \sum_{i=1}^N \ln u_i \quad (6)$$

A likelihood ratio $u_i > 1$ makes it more likely that U belongs to A , and a likelihood ratio $u_i < 1$ makes it more likely that U belongs to B . A likelihood ratio $u_i = 1$ does not shift the overall likelihood ratio L either way.

We should point out that the observer's estimates of the likelihood ratio L may be correct or incorrect. Often we use a Bayesian framework to derive the *ideal* observer for a particular task, and certainly the ideal observer must compute the relevant likelihood ratios correctly. More generally, though, a Bayesian framework allows us to

model an observer's *beliefs* about what can be inferred from an observation, and these beliefs may be correct or incorrect. In other words, just because we describe an observer in a Bayesian framework, we need not assume that the observer follows an ideal strategy.

How could we represent selective attention in this well-known Bayesian pattern classification framework? Suppose that a stimulus contains two classes of elements, U_i and V_j . When the observer selectively attends to U_i , he takes these elements as being more relevant to the task than V_j , and he reduces the influence of V_j on his responses. Another way of saying this is that the observer discounts the evidence provided by V_j , and assigns it a smaller weight in his decision. If we regard the observer as basing his responses on a likelihood ratio as in Equation 5, this amounts to his adjusting the likelihood ratios u_i and v_j that are computed from the two classes of stimulus elements, U_i and V_j . For instance, if on a particular trial an element V_1 would contribute a likelihood ratio of $v_1=1.2$ if attended to, hence biasing the observer's response toward 'A', an observer who selectively attends away from V_1 can be thought of as adjusting the likelihood ratio v_1 toward 1.0, so that V_1 has less influence on his response. That is, when the observer selectively attends to U_i , he adjusts the likelihood ratios v_j by some function f :

$$L = \prod_{i=1}^{N_1} u_i \prod_{j=1}^{N_2} f(v_j) \quad (7)$$

We will assume that selective attention affects only the likelihood ratios v_j corresponding to the elements V_j that the observer selectively attends away from. Later in this section we show that this makes our model only slightly less general than if we allow selective attention to affect both sets of likelihood ratios, u_i and v_j .

For this description of selective attention to be meaningful, the attenuating function f must satisfy a simple constraint: the likelihood ratio L computed in Equation 7 should not depend on how we conceptually divide the stimulus into independently varying elements U_i and V_j . In particular, our predictions concerning the effects of selective attention should not change if we reformulate our model so that two elements V_1 and V_2 with likelihood ratios v_1 and v_2 are now regarded as a single element $V_{1,2}$ with likelihood ratio $v_1 v_2$. It follows that

$$f(v_1 v_2) = f(v_1) f(v_2) . \quad (8)$$

The theory of functional equations (Falmagne, 1985) shows that Equation 8 implies that f is a power function,

$$f(x) = x^k . \quad (9)$$

Hence, a reasonable guess for the form of selective attention is

$$L = \prod_{i=1}^{N_1} u_i \prod_{j=1}^{N_2} v_j^k \quad (10)$$

$$= \prod_{i=1}^{N_1} u_i \left(\prod_{j=1}^{N_2} v_j \right)^k . \quad (11)$$

The corresponding log likelihood ratio is

$$\ln L = \sum_{i=1}^{N_1} \ln u_i + k \sum_{j=1}^{N_2} \ln v_j . \quad (12)$$

If $k=0$, all likelihood ratios v_j are mapped to 1, and the distractor elements V_j have no effect on the observer's responses. If $k=1$, the likelihood ratios v_j are unaffected, and V_j have their full effect. Note the similarity of Equation 12 to Equation 1, where we defined k as the attentional weight assigned to the distractors.¹

The idea that observers combine information from different sources in a weighted sum has been proposed by many authors for many different tasks, as we discussed in the 'Introduction.' This derivation shows that in tasks where observers selectively attend to one information source rather than another, there are good reasons why they might combine information this way. This formulation leads directly to the notion of attentional weight, which provides a very general way of measuring selective attention, and even gives a meaningful way of comparing the efficacy of selective attention across different tasks.

Finally, suppose that we allow selective attention to affect the likelihoods computed from both targets and distractors:

$$L = \prod_{i=1}^{N_1} f_1(u_i) \prod_{j=1}^{N_2} f_2(v_j) \quad (13)$$

$$= \prod_{i=1}^{N_1} u_i^{k_1} \prod_{j=1}^{N_2} v_j^{k_2} \quad (14)$$

If we set the attentional weight in Equation 10 to $k=k_2/k_1$, then the likelihood ratio in Equation 10 exceeds 1 if and only if the likelihood ratio in Equation 14 exceeds 1, so an unbiased observer would give the same response regardless of which expression that he used. Hence, for an unbiased observer, we can assume that selective attention affects only the likelihood ratios corresponding to unattended stimuli. If an observer is biased (i.e., adopts a likelihood ratio criterion different from 1), then models (10) and (14) are not equivalent, and we might be able to compare these models experimentally by persuading the observer to use an extreme criterion. Here we do not consider the case of a biased observer.²

An Illustration: Selective Attention and Contrast Polarity

As an illustration, we will apply this framework to the question of whether observers can direct attention according to contrast polarity when judging global direction of motion. [Edwards and Badcock \(1994\)](#) argued that this question is relevant to whether signals in ON and OFF pathways merge before reaching cortical area MT, which plays an important role in computing global direction of motion ([Newsome & Paré, 1988](#)). The question is also interesting from a purely psychological point of view, as it addresses a basic question about the capabilities of selective attention.

In [Edwards and Badcock's \(1994\)](#) experiments, observers viewed random dot cinematograms that contained an equal number of white target dots and black distractor dots. A small number of white target dots all moved either directly upward or directly downward, whereas the remaining white target dots and all the black distractor dots moved in random directions. Observers judged whether all the white dots moved on average upward or downward. The question [Edwards and Badcock \(1994\)](#) posed was, "Can observers judge the direction of only the white dots, or do the black dots disrupt the ability to discriminate between upward and downward motion of the white dots?" (In the following section, we will assume that the dots move on average to the left or to the right, rather than upward or downward, as this was the case in the experiments we report later in this work.)

In this task, a Bayesian observer could take each dot displacement as a piece of evidence that the correct answer is "left" or "right," as in [Equation 5](#). Such an observer would compute the product of the likelihood ratios corresponding to the individual dot displacements, and set a criterion to discriminate between movement to the left and to the right. Equivalently, the observer could calculate the sum of the log likelihood ratios corresponding to the dot displacements, as in [Equation 6](#). This sum of quantities corresponding to individual dot displacements can often be redescribed more intuitively. For instance, if the observer assumes that the distribution of dot directions is Gaussian, then the sum of log likelihood ratios simply measures the total horizontal displacement of all the target dots; an unbiased observer who follows this strategy responds "left" if the total displacement is leftward, and "right" if the displacement is rightward ([Watamaniuk, 1993](#)). Alternatively, the observer could base his responses on the output of more narrowly tuned motion channels, perhaps considering only the number of dots that move directly to the left or to the right. To be concrete, we will assume that observers base their responses on the total horizontal displacement of all the dots, but in a later section ("A More General Model") we show that our results do not depend on this assumption.³

We can plot the total horizontal displacements of the white target dots and the black distractor dots on orthogonal axes ([Figure 1](#)). In this plot, each point represents a single trial. The x-component of each point is the total horizontal displacement of all the target dots on that trial (i.e., the sum of the horizontal displacements of the individual target dots), and the y-component is the total horizontal displacement of all the distractor dots. The cluster on the left represents trials on which the correct answer is "left" and the cluster on the right represents trials on which the correct answer is "right." Because the dots take finite random walks, there is trial-to-trial variability in their horizontal displacements.

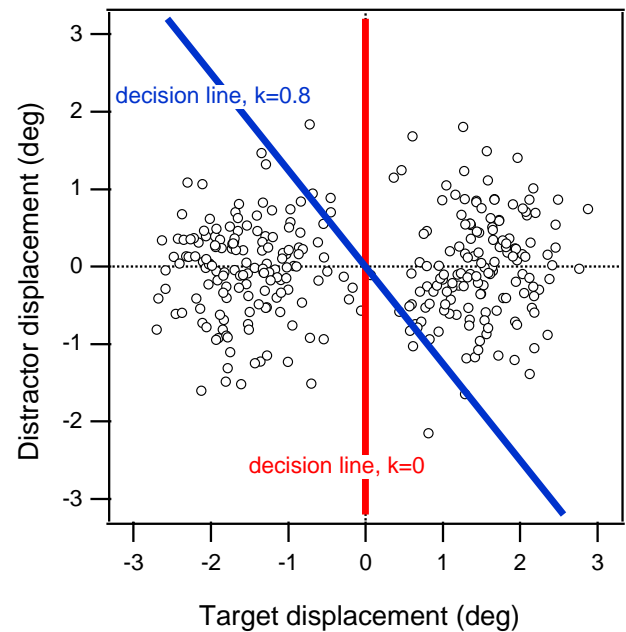


Figure 1. A hypothetical observer's decision space in Experiment 1. Each point represents a single trial. The x-coordinate of each point is the total horizontal displacement of all the target dots on a trial, and the y-coordinate is the horizontal displacement of the distractor dots. The red and blue lines are illustrative decision lines.

This plot represents the decision space of an observer who bases his responses on the total horizontal displacements of the target and distractor dots. Ideally, the observer should ignore the displacement of the distractors, as this quantity gives no information as to the correct response. For such an observer, the decision variable, which we will call s , is equal to the target displacement, which we will call T . An unbiased observer of this type responds "right" if s is greater than zero, and "left" if s is less than zero. This strategy can be represented as a vertical decision line that divides the decision space in two (e.g., the red line in [Figure 1](#)). On the other hand, if the observer cannot selectively attend to the target dots, his responses will be based on some combination of the total horizontal target displacement T and the total horizontal distractor displacement, which we will call D .

As in Equation 12, we will model the observer's decision variable s as a weighted sum of the internal responses to the target and distractor dots:

$$s = T + kD \quad (15)$$

The attentional weight k assigned to the distractor dots determines the influence of the distractors on the observer's responses. For an observer for whom $k \neq 0$, the decision line is not vertical, but rather has slope $-1/k$ (e.g., the blue line in Figure 1).

The weighted sum of target and distractor displacements in Equation 15 would be a natural first attempt at modeling selective attention in this task, even on grounds of simplicity. We wish to emphasize, though, that we arrived at Equation 15 by a different, less pragmatic argument. First, we noted that several studies support the notion that observers judge the global direction of a random dot cinematogram by summing internal responses to individual dot displacements (Watamaniuk, 1993; Watamaniuk, Sekuler, & Williams, 1989; Williams, Tweten, & Sekuler, 1991; Zohary, Scase, & Braddick, 1996). Second, we reasoned that any observer who arrives at a response by summing several quantities computed from a stimulus can be regarded as a Bayesian observer who sums log likelihood ratios, as in Equation 6). Third, our derivation of attentional weight showed that one simple and plausible way of describing the effects of selective attention is with a single weighting factor in a sum of log likelihood ratios, as in Equation 12. These considerations show that the weighting factor k in Equation 15 is not just an arbitrary free parameter, but that it is actually the attentional weight that our hypothetical observer assigns to the distractors. Hence Equation 15 results from a direct application of our account of attentional weight to the task of judging the direction of target dots mixed with distractor dots in a random dot cinematogram.

How to Measure Attentional Weight

According to the account we have outlined, a key problem in the study of selective attention is measuring the attentional weight that an observer assigns to distractors. We now describe a simple method of doing this.

Figure 2 is a plot of our hypothetical observer's decision space, showing only trials on which the correct answer is "right." The large black dot M shows the mean total horizontal displacement of the target and distractor dots over all trials where the target signal dots move right, indicating that on average the target dots move to the right and the distractor dots have zero displacement. The green dot M_R shows the mean target and distractor displacements over all trials where the target moves right and the observer responds "right." As indicated by the dashed line in Figure 2, this conditional mean is shifted from the unconditional mean along a line that is perpendicular to the decision line. This follows from the

fact that the distribution of target and distractor displacements is radially symmetric: the part of the distribution that falls on one side of the decision line is mirror-symmetric about a line that is perpendicular to the decision line and passes through M , so the mean over all trials where the observer responds "right" must lie along this line. Similarly, the mean displacement M_L over all trials where the target moves right and the observer responds "left" is shifted from the overall mean along the same line in the opposite direction, as indicated by the large red dot. The slope of the decision line is $-1/k$, so the slope of the perpendicular line connecting the two conditional means is k , the attentional weight that the observer assigns to the distractor dots.

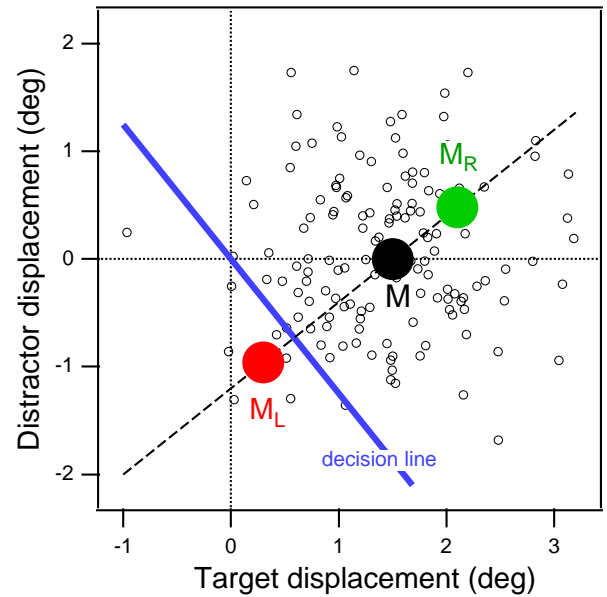


Figure 2. Part of a hypothetical observer's decision space, showing trials on which the target signal dots move to the right. M is the mean over all trials, M_R is the mean over trials where the observer responds "right," and M_L is the mean over trials where the observer responds "left."

Let the random variable C represent the correct response on a given trial, taking the value $+1$ or -1 on trials where the correct response is "right" or "left," respectively. Similarly, let the random variable R represent the observer's responses, taking the value $+1$ or -1 on trials where the observer responds "right" or "left," respectively. With this notation, the coordinates of the conditional mean displacements M_R and M_L are

$$\begin{aligned} M_R &= (E[T | C = 1, R = 1], E[D | C = 1, R = 1]) \\ M_L &= (E[T | C = 1, R = -1], E[D | C = 1, R = -1]) \end{aligned} \quad (16)$$

and we have just shown that the slope of the line connecting these points is k :

$$k = \frac{E[D | C = 1, R = 1] - E[D | C = 1, R = -1]}{E[T | C = 1, R = 1] - E[T | C = 1, R = -1]} \quad (17)$$

We can obtain a second, independent estimate of k by using Equation 17 with $C=-1$ (i.e., finding the slope of the line connecting the conditional means over all trials where the correct answer is “left”).

We could stop here, as Equation 17 shows how to calculate k from measurable quantities, but a reformulation makes the meaning of this expression much clearer. First, note that the coordinates of M_R and M_L with respect to an origin $M = (\mu_T, \mu_D)$ at the mean of the distribution of target and distractor displacements are

$$\begin{aligned} M_R' &= (E[T | C = 1, R = 1] - \mu_T, E[D | C = 1, R = 1] - \mu_D) \\ M_L' &= (E[T | C = 1, R = -1] - \mu_T, E[D | C = 1, R = -1] - \mu_D) \end{aligned} \quad (18)$$

and, of course, we obtain the same value of k if we calculate the slope of the connecting line in this coordinate frame. Second, because M_R' , M_L' , and M are collinear, we obtain the same value for k if we multiply M_R' by $P(R = 1)(1 - \mu_R)$ and multiply M_L' by $-P(R = -1)(-1 - \mu_R)$, where $\mu_R = E[R]$. These transformations convert Equation 17 into a ratio of covariances:

$$k = \frac{P(R = 1)E[(D - \mu_D)(1 - \mu_R) | C = 1, R = 1] + P(R = -1)E[(D - \mu_D)(-1 - \mu_R) | C = 1, R = -1]}{P(R = 1)E[(T - \mu_T)(1 - \mu_R) | C = 1, R = 1] + P(R = -1)E[(T - \mu_T)(-1 - \mu_R) | C = 1, R = -1]} \quad (19)$$

$$= \frac{E[(D - \mu_D)(R - \mu_R) | C = 1]}{E[(T - \mu_T)(R - \mu_R) | C = 1]} \quad (20)$$

$$= \frac{\text{cov}[D, R | C = 1]}{\text{cov}[T, R | C = 1]} \quad (21)$$

Hence to find the attentional weight that the observer assigns to the distractor dots, we can measure the covariance between the total horizontal target and distractor dot displacements and the observer's responses, over all trials where the correct answer is “right,” and take the ratio of these two covariances. That is, the attentional weight is equal to the influence of the distractor dots on the observer's responses, as a proportion of the influence of the target dots on the observer's responses.

Strictly speaking, Equation 21 requires a small correction. We have assumed that the distribution (T, D) is radially symmetric over all trials where the target dots move in a given direction. The random variables T and D are independent, so this is true only if they are Gaussian and their variances are equal. Both T and D are the sum of many horizontal dot displacements, so the central limit theorem ensures that they will be approximately Gaussian. However, in the random dot cinematograms in the experiments we report below, there are an equal number of target and distractor dots, and a small number of target dots always move in a given direction, so there are slightly fewer randomly moving target dots than

randomly moving distractor dots. Consequently, the variance of T is actually slightly less than the variance of D . In Appendix A, we show that we can correct for this difference by adjusting k by a factor $(N - n_T)/N$, where N is the total number of target dots, and n_T is the number of target dots that move directly left or right. The corrected expressions are

$$k = \frac{(N - n_T) E[D | C = 1, R = 1] - E[D | C = 1, R = -1]}{N E[T | C = 1, R = 1] - E[T | C = 1, R = -1]} \quad (22)$$

$$= \frac{(N - n_T) \cdot \text{cov}[D, R | C = 1]}{N \cdot \text{cov}[T, R | C = 1]} \quad (23)$$

When the coherently moving target dots make up only a small proportion of the dots in the cinematogram, as is usual, this correction is negligible compared to experimental error.

This correlation method is closely related to the classification image method used in psychophysics to characterize the computation that an observer uses to perform a perceptual task (Ahumada & Lovell, 1971; Beard & Ahumada, 1998; Gold, Murray, Bennett, & Sekuler, 2000; Neri, Parker, & Blakemore, 1999), and to the reverse correlation method used in neurophysiology

to map receptive fields (Chichilnisky, 2001; Pinter & Nabet, 1992). Our method reduces the stimulus to two numbers, the total horizontal target and distractor displacements, and measures the correlation of these quantities with the observer's responses. As in the classification image and reverse correlation methods, these correlations reveal the linear component of the computation that the observer uses to perform the task.

It should be clear that this correlation method could be useful even outside the linear cue combination framework. If we measure the correlation of targets and distractors with an observer's responses, and find that the distractors have as strong an influence on an observer's responses as the targets do, then clearly we can conclude that the observer has little ability to selectively attend to the targets, even if we have no reason to believe that the observer uses a decision variable that is a weighted sum of internal responses, as in Equation 1. That is, regardless of how the observer makes his responses, the correlation ratio gives a rough measure of how much an observer's responses are influenced by distractors.

In Experiment 1, we illustrate this correlation method by measuring the attentional weight that observers assign to distractor dots in a global direction discrimination task.

A More General Model

Up to now, we have assumed that the observer's decision variable is a weighted sum of the total horizontal displacements of the targets and distractors, $s = T + kD$. This allowed us to calculate the exact value of the random variables T and D on each trial, directly from the stimulus. With this information, we were able to locate each trial in the observer's decision space, as in Figure 2, and recover the attentional weight k by finding the slope of the line connecting the mean internal responses over all trials where the observer responded "left" or "right." However, real observers' decision variables are certainly not $s = T + kD$. First of all, real observers have internal noise, and, second, observers might compute some quantity other than the horizontal displacement of the target and distractor dots (e.g., an observer might count the number of dots that move directly to the left or right, or monitor the activation of 30°-wide motion channels). This seems to pose a problem for our method of measuring attentional weight, as this method apparently relies on our knowing the observer's internal responses to the target and distractor dots on every trial.

In fact, the methods given by Equations 17 and 21 are valid under a much broader range of conditions than we have shown so far. In Appendix A, we show that we need assume only that the observer's decision variable fits the following model, which is similar to the very general Bayesian decision variable in Equation 12, except that it explicitly introduces noise into the observer's decisions.

First, we assume that the observer's decision variable is a weighted sum of a quantity T^* computed from the target dots and a quantity D^* computed from the distractor dots:

$$s = T^* + kD^* \quad (24)$$

Second, we assume that T^* and D^* are computed by summing responses to individual target and distractor dot displacements, and that the observer has the same selectivity f for target and distractor dot displacements. We also assume that T^* and D^* are contaminated by independent, equal-variance internal noise sources Z_T and Z_D . Thus we can write the internal responses T^* and D^* as

$$T^* = \sum_i f(t_i) + Z_T \quad (25)$$

$$D^* = \sum_i f(d_i) + Z_D \quad (26)$$

Here t_i and d_i are random variables, perhaps multi-dimensional, that describe the relevant properties of individual target and distractor dot displacements, respectively. For instance, to describe an observer who performs the direction discrimination task using 30°-wide motion channels, but is less affected by dots at greater eccentricities, the random variables t_i and d_i would report

the direction and eccentricity of each dot displacement, and the function f would describe the observer's selectivity to dots in each direction, at each eccentricity. Such noisy linear-filter models have been found to give a good account of global motion perception under a wide range of conditions (Watamaniuk, 1993; Zohary et al., 1996).

One straightforward way of testing this model is by measuring the observer's psychometric function, which the following argument shows should be linear when plotted as d' versus the number of signal dots. Let f_R and f_L be the mean value of $f(t_i)$ when t_i is a dot that steps directly to the right or to the left, respectively. If an observer can perform the direction discrimination task at all, then $f_R \neq f_L$, and in a task with n_T target signal dots, the difference in the mean of T^* when the dots move to the right or to the left is $n_T(f_R - f_L)$. Furthermore, if n_T is much smaller than the total number of dots in the cinematogram, then the variance σ_s^2 of the observer's decision variable is largely independent of n_T . Consequently, the observer's sensitivity is $d' = n_T(f_R - f_L) / \sigma_s$, indicating that the psychometric function is linear when plotted as d' versus the number of signal dots. In Experiment 1, we measured psychometric functions in a global direction discrimination task to test the linearity assumption implicit in this model.

Same Selectivity for Attended and Unattended Stimuli?

According to Equation 24, the observer computes the same internal response D^* from the distractors, regardless of whether the distractors are fully attended ($k=1$) or partially or completely unattended ($k<1$); selective attention merely modulates the influence of this internal response on the decision variable. In other words, this account implies that selective attention does not qualitatively change how the observer processes the distractors, but only attenuates the influence that the distractors have on the observer's responses. Of course, we cannot know a priori whether this is true of human observers, and it may be that in some tasks, processing of attended and unattended stimuli is qualitatively different. For instance, it may be that when observers judge global direction of motion in random dot cinematograms, the directional selectivity of motion channels is different for attended and partly unattended dots. Accordingly, we cannot be certain that attentional weight is an appropriate measure of selective attention until we compare how observers process attended and unattended stimuli.

Chubb and colleagues (Chubb, 1999; Chubb et al., 1994) have developed a method of characterizing observers' strategies in perceptual tasks by measuring the influence of small stimulus elements on the observers' responses. They call this method histogram contrast analysis (HCA). In Appendix B, we describe a version of HCA that allows us to measure the directional selectivity of the motion channels that an observer applies to attended and

unattended stimuli. We show that if the observer bases his responses on a linear motion channel with directional selectivity $f(\theta)$, then we can estimate the directional selectivity function $f(\theta)$ by measuring the influence that each dot moving in direction θ has on the observer's responses. Specifically, we show that the conditional probability that an observer responds "right" when an arbitrarily chosen dot moves in direction θ is related to the directional selectivity function $f(\theta)$ as follows:

$$P(\text{'right'} | \theta) = uf(\theta) + v, \quad (27)$$

where u and v are constants. In Experiment 1, we used the HCA method to compare direction selectivity for attended and unattended dots in a global direction discrimination task.

Experiment 1

In the first experiment, we applied the methods we have described in the previous sections to a global direction discrimination task. First, we measured psychometric functions in a task where observers judged the global direction of black or white random dot cinematograms, in order to see whether observers met the linearity assumption of the model given by Equations 24 through 26, which underpins our other methods. Second, we measured the attentional weight that observers assigned to distractors, in a task where observers judged the global direction of motion of white target dots in a random dot cinematogram. In one condition, the white target dots were mixed with black distractor dots. This condition tested whether observers could direct attention according to contrast polarity. In a second condition, the white target dots were mixed with white distractor dots. This condition served as a validation condition for our method of measuring attentional weight, as observers could not distinguish between targets and distractors,⁴ and so we knew in advance that the correct value of attentional weight was $k=1$. Finally, we used the HCA method developed by Chubb et al. (1999) to measure directional selectivity for target and distractor dots, to see

whether selective attention led to qualitative differences in processing of targets and distractors, or merely reduced the influence of distractors on observers' responses.

Methods

Participants

One author (R.F.M.) and four University of Toronto students participated. Two observers (R.F.M. and C.P.T.) were practiced at direction discrimination in random dot cinematograms and were aware of the hypotheses being tested. The other three observers were not practiced at this task and were unaware of the hypotheses. All observers in all experiments reported in this paper had normal or corrected-to-normal Snellen acuity.

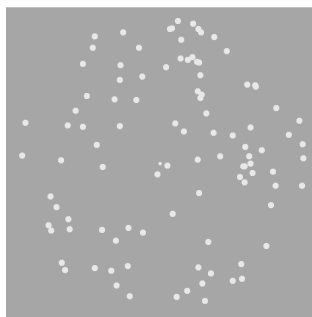
Stimuli

Psychometric Function Conditions (100L, 100D)

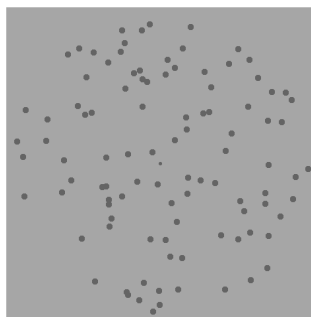
The stimuli in the psychometric function conditions were eight-frame random dot cinematograms (Figure 3). Each frame lasted 45 ms, and the entire cinematogram lasted 360 ms. In each frame, 100 dots of radius 0.10 deg of visual angle appeared in a circular aperture of radius 6.0 deg. Between successive frames, a number of dots (the "signal dots") moved 0.30 deg to the left or to the right, and the remainder (the "noise dots") moved an equal distance in random directions. On a given trial, all the signal dots moved in the same direction. On each frame, a new random subset of dots was chosen as signal dots. The lifetime of each dot was eight frames. In the 100L condition, the dots were white (Weber contrast 0.40; Figure 3a), and in the 100D condition, the dots were black (Weber contrast -0.40; Figure 3b). Weber contrast is defined as $c_W = (L - L_{bg}) / L_{bg}$, where L is the luminance of the point of interest, and L_{bg} is background luminance. The stimuli were shown on a gray background of luminance 40 cd/m².

Stimuli were displayed on an AppleVision 1710 monitor (640 × 480 resolution, pixel size 0.467 mm, refresh rate 67 Hz). Observers viewed the stimuli

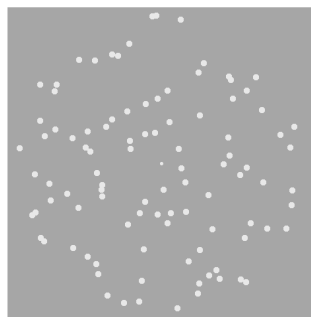
(a) 100L



(b) 100D



(c) 50L50L



(d) 50L50D

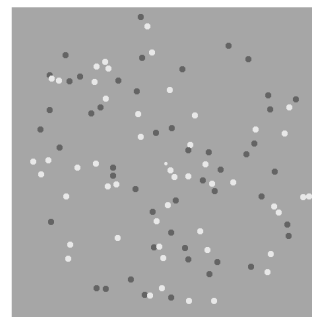


Figure 3. Stimuli in Experiments 1 and 2. Click to view movies.

binocularly from a distance of 1 m, and head position was stabilized using a chin-and-forehead rest.

Attention Conditions (50L50L, 50L50D)

The stimuli in the attention conditions were similar to those in the psychometric function conditions, but the dots were divided into two 50-dot subsets. Fifty dots were target dots: between successive frames, a number of dots in this subset (the signal dots) moved 0.30 deg to the left or to the right, and the remainder (the noise dots) moved an equal distance in random directions. From frame to frame, a new random subset of the 50 target dots was chosen as signal dots. The other 50 dots in the cinematogram were distractor dots: between successive frames, all the dots in this subset moved 0.30 deg in random directions. In both the target and the distractor subsets, the lifetime of each dot was eight frames. In the 50L50L stimulus, both the targets and the distractors were white (Weber contrast 0.40; Figure 3c). In the 50L50D stimulus, the targets were white and the distractors were black (Weber contrast ± 0.40 ; Figure 3d). These stimuli are similar to those used by Edwards and Badcock (1994), the main difference being that in Edwards and Badcock's 100L stimulus, any of the 100 dots could become a signal dot, whereas in our 50L50L stimulus, only the 50 target dots could become signal dots, and all 50 distractor dots took unbiased random walks.

Procedure

Psychometric Function Conditions

Two observers (J.A.P. and S.U.M.) participated in two to three 1-hr sessions. Each session consisted of 18 blocks of 100 trials. One half the blocks were 100L blocks, one half were 100D blocks, and the session alternated between the two types of blocks. Each trial began with a 500-ms fixation interval, followed by a 360-ms random dot cinematogram, followed by a response interval in which the observer pressed one of two keys to indicate whether the mean direction of the dots was to the left or to the right. Auditory feedback indicated whether the observer's response was correct. A small white fixation dot appeared at the center of the screen throughout the trial in the 100L condition, and a small black fixation dot appeared in the 100D condition. The number of signal dots varied across trials according to the method of constant stimuli. The numbers of signal dots were chosen to span each observer's psychometric function, based on a short pilot session. For observer J.A.P., the signal levels were 2, 4, 8, 12, and 16 signal dots per frame, and for observer S.U.M., they were 5, 10, 15, 20, and 25 signal dots per frame.

Attention Conditions

Three observers (A.N.C., C.P.T., and R.F.M.) participated in four to eight 1-hr sessions. Each session consisted of eight blocks of 300 trials. One half the blocks were

50L50L blocks, one half were 50L50D blocks, and the session alternated between the two types of blocks. The sequence of events in a trial was the same as in the 100L and 100D conditions. For each observer, the number of signal dots per frame was fixed at a number found during a pilot session to give approximately 70% correct performance. For observer A.N.C., this was eight signal dots per frame, for C.P.T., six signal dots per frame, and for R.F.M., two signal dots per frame.

In both the 50L50L and 50L50D conditions, observers were instructed to indicate the mean direction of the white dots. In the 50L50L condition, the targets and distractors were indistinguishable, so we assumed that instructions to selectively attend to the target dots would merely frustrate the observers. Furthermore, the purpose of the 50L50L condition was to measure attentional weight in a condition where observers attended equally to the targets and the distractors, and instructions to judge the mean direction of all the white dots encouraged observers to follow this strategy.

Results and Discussion

Psychometric Functions

Figure 4 shows psychometric functions for both observers in the 100L and 100D conditions. The functions were approximately linear, supporting our hypothesis that the observers' decision variable is a linear sum of responses to individual dot displacements.

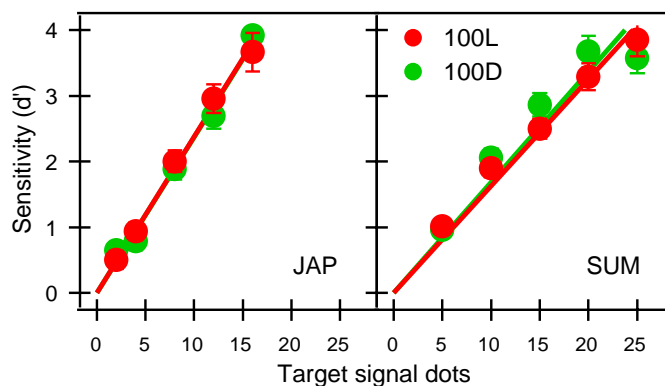


Figure 4. Psychometric functions in the 100L and 100D conditions. The error bars are SEs, and are often smaller than the data points.

50L50L Condition

Figure 5 shows the results of the 50L50L condition for all three observers. Each small X represents a single trial on which the observer responded "left," and each small O represents a trial on which the observer responded "right." The x-coordinate of each small X and O shows the total horizontal displacement of the target dots on that trial, and the y-coordinate shows the total horizontal

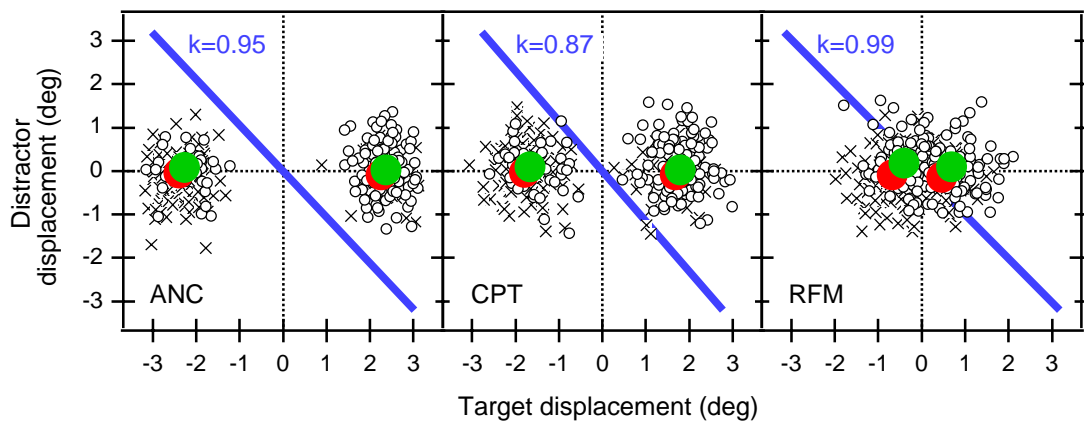


Figure 5. Results of Experiment 1, 50L50L condition. Each X represents a trial on which the observer responded “left,” and each O represents a trial on which the observer responded “right.” The x-coordinate of each X and O shows the total horizontal displacement of the target dots on that trial, and the y-coordinate shows the total horizontal displacement of the distractor dots. The cluster on the left represents trials on which the correct answer was “left,” and the cluster on the right represents trials on which the correct answer was “right.” The red and green dots show the mean displacements over all trials on which the observer responded “left” and “right,” respectively. The blue lines are the observers’ decision lines, $T+kD=0$.

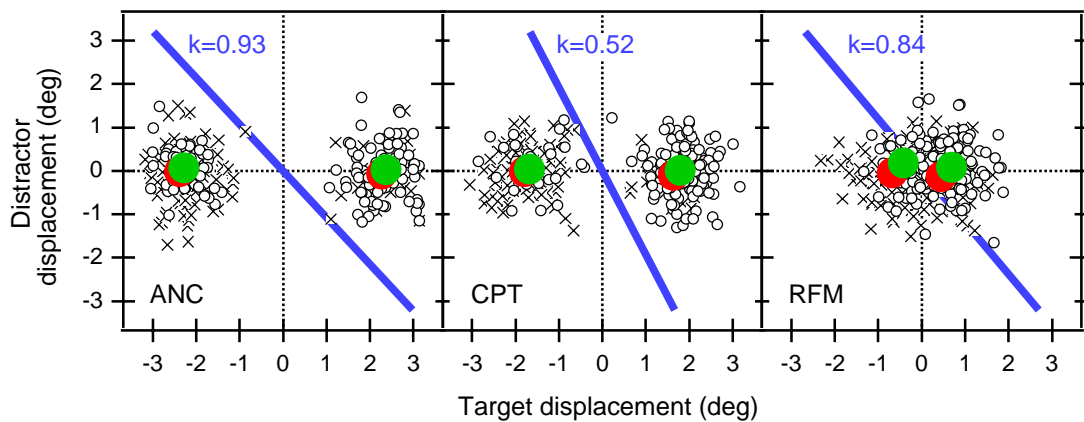


Figure 6. Results of Experiment 1, 50L50D condition. See caption of Figure 5 for details.

displacement of the distractor dots. The cluster on the left represents trials on which the correct answer was “left,” and the cluster on the right represents trials on which the correct answer was “right.” Only 150 randomly chosen trials are shown, to keep the graphs from being too cluttered. The red and green dots represent the mean displacements over all trials on which the observer responded “left” and “right,” respectively. The pair of red and green dots on the left side of each observer’s plot represents the means over all trials on which the correct answer was “left,” and the pair on the right represent the means over all trials where the correct answer was “right.”

The left and right clusters of data points are separated by different distances in each observer’s plot, because each observer required a different number of signal dots per frame to maintain 70% correct performance. For instance, observer A.N.C. required eight signal dots, whereas R.F.M. required only two, so the distance between the clusters is four times larger for A.N.C. than for R.F.M.

For the highly practiced observer R.F.M., a large number of the trials on which the nominally correct answer was “right” actually had mean target displacements to the left, and vice versa. This indicates that R.F.M. used such an efficient strategy that his performance was largely limited by statistical noise in the stimulus itself.

Note that the mean displacements on trials where the observer responded “right” (the green dots) are shifted upward and to the right of the mean displacements on trials where the observer responded “left” (the red dots). The horizontal shift indicates that the target displacement was correlated with observers’ left/right responses, and the vertical shift indicates that the distractor displacement was also correlated with observers’ responses. Furthermore, the vertical displacement was typically just as large as the horizontal displacement, indicating that the distractors influenced observers’ responses as much as the targets did. This is precisely what we expect in the 50L50L condition, as observers had no way of distinguishing

between target and distractor dots. Because these shifts are small, and not easily seen in some observers' plots, we have listed the conditional mean displacements of the target and distractor dots in Table 1.

We calculated the attentional weight that observers assigned to the distractor dots using Equation 22 and the conditional mean displacements in Table 1. For observer A.N.C., $k=0.95 \pm 0.16$, for C.P.T., $k=0.87 \pm 0.19$, and for R.F.M., $k=0.99 \pm 0.09$. The error values are SEs. None of these estimates of k is significantly different from the anticipated value of 1 ($p > .40$ for all comparisons in a two-tailed test). The slanted blue lines in Figure 5 show the decision lines, $T + kD = 0$, corresponding to these values of k .

50L50D Condition

Figure 6 shows the results of the 50L50D condition for all three observers, and Table 1 lists the conditional mean displacements of the target and distractor dots. Again, both target and distractor displacements were correlated with observers' responses, indicating that observers were unable to restrict their attention to the white target dots. For observer A.N.C., $k=0.93 \pm 0.20$, for C.P.T., $k=0.52 \pm 0.15$, and for R.F.M., $k=0.84 \pm 0.09$. All these estimates of k are significantly greater than zero ($p < .001$ for all comparisons), none is significantly less than the observer's corresponding value in the 50L50L condition ($p > .10$ for all comparisons), and only C.P.T.'s is significantly less than 1 ($p < .01$).

Clearly, observers' abilities to direct attention according to contrast polarity were limited at best: two of the three observers were not influenced significantly less by opposite-polarity distractors than by same-polarity distractors, and the third observer was influenced 52% as

much by opposite-polarity distractors as by same-polarity distractors. These results are consistent with previous findings that opposite-polarity distractors have a large influence on observers' responses (Edwards & Badcock, 1994; Li & Kingdom, 2001; Snowden & Edmunds, 1999; van der Smagt & van de Grind, 1999). These results do not mean that observers *misperceive* distractors as targets. At a Weber contrast of $\pm 40\%$, the targets and distractors are highly discriminable. Rather, these results show that observers cannot make global direction judgments based solely on the directions of white target dots, in the presence of black distractor dots.

Directional Selectivity

We compared observers' directional selectivity for attended and unattended stimuli in the 50L50D condition, using the version of HCA presented in Appendix B. For each target and distractor dot in the cinematogram, we measured the probability of the observer responding "right" when the dot moved in direction θ , and we averaged these direction selectivity functions separately over all target dots and over all distractor dots. Figure 7 shows the influence of target dots and distractor dots on the observers' responses, as a function of dot direction, averaged across all three observers. The directional tuning was approximately sinusoidal for both attended and unattended dots, varying as $\cos \theta$, indicating that observers based their responses on the horizontal displacements of both target and distractor dots. Evidently observers processed attended and unattended stimuli in the same way, at least in terms of their directional selectivity. Furthermore, the best-fitting sinusoids had slightly different amplitudes, reflecting the fact that unattended dots had less overall

Table 1. Results of Experiment 1

		Mean target displacement (deg)				Mean distractor displacement (deg)			
		Target R		Target L		Target R		Target L	
		Response R	Response L	Response R	Response L	Response R	Response L	Response R	Response L
50L50L	A.N.C.	2.364	2.263	-2.277	-2.389	0.022	-0.087	0.083	-0.047
	C.P.T.	1.786	1.679	-1.658	-1.795	0.033	-0.081	0.093	-0.038
	R.F.M.	0.688	0.453	-0.417	-0.682	0.104	-0.142	0.182	-0.089
50L50D	A.N.C.	2.364	2.277	-2.284	-2.376	0.024	-0.074	0.077	-0.025
	C.P.T.	1.794	1.640	-1.680	-1.799	0.017	-0.094	0.038	-0.011
	R.F.M.	0.691	0.441	-0.420	-0.672	0.084	-0.124	0.179	-0.054

This table shows the mean total rightward displacement of the target and distractor dots, conditional on the target signal dots moving left or right and the observer responding "left" or "right." For example, the top left entry shows that for observer A.N.C. in the 50L50L condition, the average total target dot displacement was 2.364 deg to the right on trials where the target signal dots moved right and the observer responded "right." The values in this table are the coordinates of the conditional mean displacements shown in Figures 5 and 6 as red and green dots. Note that both the target and the distractor displacements were correlated with observers' responses: all mean displacements were further to the right when observers responded "right" than when observers responded "left." This was true even in the 50L50D condition, where observers tried to ignore the distractor dots.

influence on the observer's responses. These results support the notion that observers have the same selectivity for attended and unattended stimulus elements, and that selective attention operates by uniformly reducing the influence of distractor elements on observers' responses.

Recently, [Eckstein, Shimozaki, and Abbey \(2001\)](#) and [Shimozaki, Abbey, & Eckstein \(2001\)](#) used the response classification method to compare processing of attended and partly unattended stimuli in a very different task, namely a detection task in which observers were given a partially valid cue as to where the target would appear, if it appeared at all. Eckstein et al. also found that observers processed cued and uncued locations similarly, and simply gave more weight to cued locations in their responses: classification images at cued and uncued locations had the same spatial profile, and differed only in amplitude. This finding strongly supports [Kinchla's \(1974\)](#) and [Kinchla and Collyer's \(1974\)](#) weighted sum account of cued detection tasks, and is persuasive evidence that attentional weight is an appropriate measure of attention in such tasks.

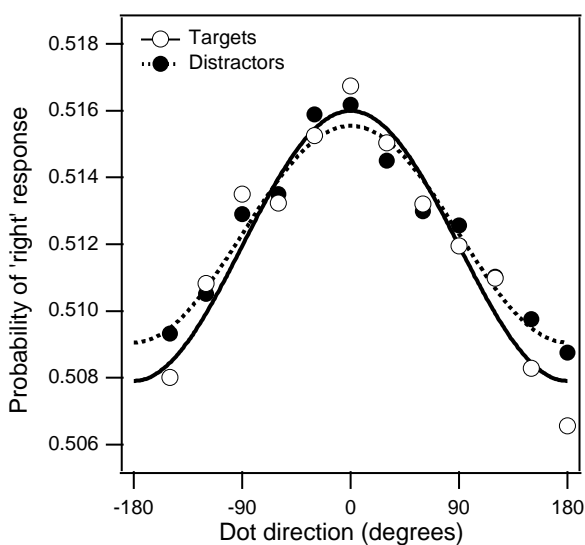


Figure 7. Histogram contrast analysis of Experiment 1, 50L50D condition. The plot shows the probability of a rightward response, as a function of the direction of each target or distractor dot, averaged across observers. The solid line is the best-fitting sinusoid to the target dot data, and the dotted line is the best-fitting sinusoid to the distractor dot data. The mean, amplitude, and phase of the sinusoids were chosen to give the best sum-of-squares fit. The amplitude of the distractor sinusoid is 0.88 times the amplitude of the target sinusoid, which is approximately the same as the mean value of attentional weight measured in Experiment 1, $k=0.86$.

Contrast Magnitude

A technical but potentially troublesome issue is whether we have properly equated the contrast magnitudes of black and white dots in this experiment.

We showed black and white dots with equal Weber contrast magnitudes ($\pm 40\%$), but there are other ways of measuring contrast besides Weber contrast. For instance, the Michelson contrast of the black and white dots was -25% and $+17\%$, respectively, so according to this measure, the contrast of the black dots was 1.5 times too high, compared to the white dots. (Michelson contrast is defined as $c_M = (L_{\max} - L_{\min}) / (L_{\max} + L_{\min})$, where L_{\max} is the maximum luminance in the region of interest and L_{\min} is the minimum luminance.) A mismatch like this might lead us to underestimate observers' abilities to direct attention according to contrast polarity, as the black dots might evoke a stronger response in motion channels than the white dots, and therefore be more difficult to ignore than black dots with properly equated contrast magnitudes.

However, for the following reasons, we believe that we have correctly matched the contrasts of the white and black dots by setting them to $\pm 40\%$ Weber contrast. First, [Edwards, Badcock, and Nishida \(1996\)](#) found that performance in an up-down direction discrimination task only improved with stimulus contrast up to about 15% Weber contrast, suggesting that at our contrast level of $\pm 40\%$, moderate differences in contrast should have little effect on performance. Second, in the psychometric function conditions of this experiment, performance was approximately the same for white and black cinematograms at $\pm 40\%$ contrast. These two facts are not conclusive, however, as [Edwards et al. \(1996\)](#) found that even though direction discrimination performance saturated at about 15% contrast when all dots in a cinematogram had the same contrast, performance worsened when the contrast of selected noise dots was increased, and continued to worsen as the contrast of the noise dots was increased up to 80% contrast. Similarly, in a previous study, we found that when observers judged small differences in the global direction of a random dot cinematogram, rather than 180° left-right or up-down direction differences, performance improved with stimulus contrast up to at least 80% contrast ([Murray, Sekuler, Bennett, & Sekuler, 1998](#)). These studies show that perceptual responses to global motion stimuli do not always saturate at low contrasts, so it is important to properly match the contrasts of white and black dots. The most persuasive evidence, therefore, is that [Murray et al. \(1998\)](#) measured performance in global direction discrimination tasks over a wide range of positive and negative contrasts, and found that observers performed equally well with white and black cinematograms that were equated for Weber contrast magnitude (e.g., $\pm 40\%$ Weber contrast). All these factors indicate that we have correctly equated the strength of the targets and distractors in our stimuli, so that we can use the influence of the distractors on observers' responses as an unbiased measure of the attentional weight that observers assign to the distractors.

A Faster Correlation Method

The correlation method we used in Experiment 1 requires a large number of trials, because it measures the effect of small statistical variations in the targets and distractors on the observer's responses. One way of measuring attentional weight more quickly would be to introduce larger trial-to-trial variations into the target and distractor displacements, and to measure the effect of these variations on the observer's responses. Here we describe a method that takes this approach.

Figure 8 shows a plot of a hypothetical observer's decision space for a task in which we vary both the mean target displacement and the mean distractor displacement from trial to trial. In this task, signal dots in the target distribution move left or right, and signal dots in the distractor distribution also move left or right. The directions of the target and distractor signal dots are chosen independently on each trial, so the decision space has four clusters of points corresponding to the four types of trials: target right, distractor right; target right, distractor left; target left, distractor right; and target left, distractor left. The observer's task is to judge the mean direction of the target dots. Note that the distractor signal dots contain no information as to the correct response; we call them signal dots only because they move coherently, rather than moving in random directions. In the task depicted in Figure 8, there are twice as many target signal dots as distractor signal dots, as indicated by the fact that the mean of each of the four clusters of trials is twice as far along the target axis as along the distractor axis.

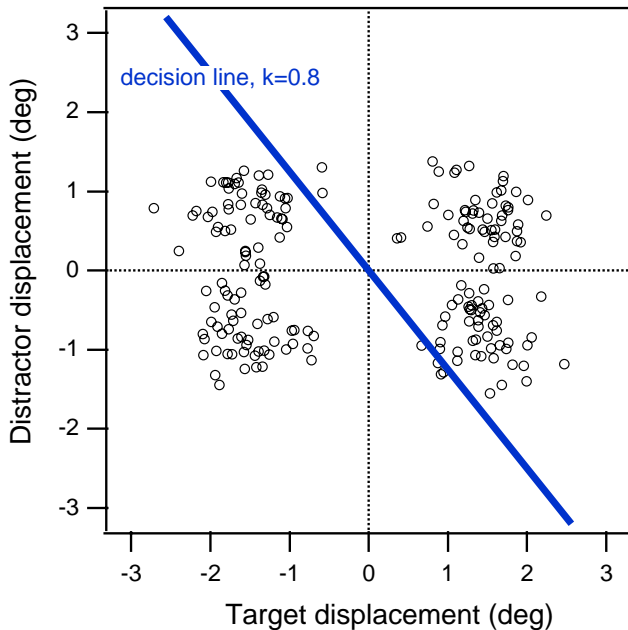


Figure 8. A hypothetical observer's decision space in Experiment 2.

If the observer's responses are influenced by the distractor dots in this task, he will give more rightward responses on trials where both the targets and the distractors move right than on trials where the target moves right and the distractor moves left. In the decision space, this is represented by the fact that a greater proportion of trials falls on the right side of the decision line when both the targets and the distractors move right (the top-right cluster in Figure 8) than when the targets move right and the distractors move left (the bottom-right cluster). By measuring the difference in the proportion of rightward responses, depending on whether the distractors move right or left, we can determine how much influence the distractors have on the observer's responses, and we can estimate the attentional weight assigned to the distractors.

The Probe Method of Measuring Attentional Weight

Let us consider in more detail how to measure attentional weight this way. Again, we will assume that the observer uses a decision variable, $s = T^* + kD^*$, described by Equations 24 through 26. We will derive expressions that show how attentional weight is related to the probability that the observer responds "right," depending on whether the target and distractor signal dots move left or right.

First, consider the statistics of T^* . Let μ_{T^*} be the expected value of T^* over all trials, and let $\Delta\mu_{T^*}$ be the difference in the expected value of T^* between trials where the target signal dots move right and trials where they move left. There are an equal number of signal-left and signal-right trials, so the overall mean μ_{T^*} lies midway between the means over signal-left trials and signal-right trials, and we can write the mean of T^* as $\mu_{T^*} \pm 0.5\Delta\mu_{T^*}$, where the sign depends on whether the target signal dots move right or left. Furthermore, the variance of T^* is the same over trials where the target signal dots move right and trials where they move left, because the signal dot displacements are constant within the signal-left and signal-right classes of trials, and do not contribute to the variance. We will denote the variance of T^* on signal-left or signal-right trials as $\sigma_{T^*}^2$. For later convenience, we define $d'_{T^*} = \Delta\mu_{T^*} / \sigma_{T^*}$, which is the sensitivity of T^* to the difference between signal-left and signal-right trials.

Second, consider the statistics of D^* . Just as with T^* , we can write the mean of D^* as $\mu_{D^*} \pm 0.5\Delta\mu_{D^*}$, where the sign depends on whether the distractor signal dots move left or right. Also, the variance of D^* is the same regardless of whether the distractor signal dots move left or right, and we will denote this variance by $\sigma_{D^*}^2$.

Third, consider how the statistics of T^* and D^* are related. Both T^* and D^* are calculated by summing internal responses to individual dot displacements, so $\Delta\mu_{T^*}$ and $\Delta\mu_{D^*}$ are proportional to the number of target and distractor signal dots, respectively. In the task we are

considering (Figure 8), there are twice as many target signal dots as distractor signal dots, so $\Delta\mu_{T^*} = 2\Delta\mu_{D^*}$. Furthermore, there are approximately the same number of target noise dots as distractor noise dots, so the variances of T^* and D^* are approximately equal: $\sigma_{T^*}^2 \approx \sigma_{D^*}^2$. (We will return to this approximation shortly.)

Finally, consider the statistics of the decision variable, $s = T^* + kD^*$. The mean of s is $\mu_{T^*} + k\mu_{D^*} \pm 0.5\Delta\mu_{T^*} \pm 0.5k\Delta\mu_{D^*}$, where the signs depend on whether the target and distractor signal dots move left or right. In this task, $\Delta\mu_{T^*} = 2\Delta\mu_{D^*}$, so the mean of s is $\mu_{T^*} + k\mu_{D^*} \pm 0.5\Delta\mu_{T^*} \pm 0.25k\Delta\mu_{T^*}$. The variance of s is $\sigma_{T^*}^2 + k^2\sigma_{D^*}^2$, and to a close approximation $\sigma_{T^*}^2 \approx \sigma_{D^*}^2$, so we can rewrite the variance as $(1+k^2)\sigma_{T^*}^2$. The midpoint of the distribution of s is $\mu_{T^*} + k\mu_{D^*}$, which is therefore the response criterion of an unbiased observer.

Now we are in a position to see how attentional weight is related to the probability of a “right” response, depending on the direction of the target and distractor signal dots. On trials where both the target and the distractor signal dots move to the right, which we will call RR trials, the mean of the decision variable is $\mu_{T^*} + k\mu_{D^*} + 0.5(1+k)\Delta\mu_{T^*}$ and the variance is $(1+k^2)\sigma_{T^*}^2$. Hence on an RR trial, the probability that the decision variable exceeds the observer’s criterion, and the observer responds “right,” is

$$p_{RR} = P(T^* + kD^* \geq \mu_{T^*} + k\mu_{D^*}) \quad (28)$$

$$= 1 - G(\mu_{T^*} + k\mu_{D^*}, \mu_{T^*} + k\mu_{D^*} + 0.5(1+0.5k)\Delta\mu_{T^*}, \sqrt{1+k^2}\sigma_{T^*}) \quad (29)$$

$$= G\left(\frac{(1+0.5k)d'_T}{2\sqrt{1+k^2}}\right). \quad (30)$$

Here $G(x, \mu, \sigma)$ is the normal cumulative distribution function, and when we omit arguments μ and σ , they default to 0 and 1, respectively.

On trials where the target signal dots move right and the distractor signal dots move left, which we will call RL trials, the mean of the decision variable is $\mu_{T^*} + k\mu_{D^*} + 0.5(1-k)\Delta\mu_{T^*}$, and the variance is the same as on RR trials. Hence the probability of the observer responding “right” on an RL trial is

$$p_{RL} = G\left(\frac{(1-0.5k)d'_T}{2\sqrt{1+k^2}}\right). \quad (31)$$

Similarly, the probabilities of the observer responding “right” when the target moves to the left and the distractor moves to the left (p_{LL}) or to the right (p_{LR}) are

$$p_{LL} = 1 - G\left(\frac{(1+0.5k)d'_T}{2\sqrt{1+k^2}}\right) \quad (32)$$

$$p_{LR} = 1 - G\left(\frac{(1-0.5k)d'_T}{2\sqrt{1+k^2}}\right). \quad (33)$$

We could solve Equations 30 and 31 to find k and d'_T as a function of the conditional response probabilities, and solve Equations 32 and 33 to give another independent estimate. However, when analyzing data from simulated model observers with known values of attentional weight, we have found the estimates of k and d'_T to be less variable when we solve all four equations simultaneously, using a simplex search to find the values of k and d'_T that minimize the sum-of-squares error between the left- and right-hand sides of Equations 30 through 33. This is the method that we recommend, so we will not derive explicit expressions for k and d'_T as a function of the conditional response probabilities.

Equations 30 through 33 use the approximation that the variances of T^* and D^* are equal, $\sigma_{T^*}^2 \approx \sigma_{D^*}^2$. In fact, the variance of T^* is slightly less than the variance of D^* in our experiments, because there are more target signal dots than distractor signal dots, and hence fewer target noise dots than distractor noise dots. We could derive exact expressions for k and d'_T that do not use this assumption, but we will not do so for two reasons. First, the approximation $\sigma_{T^*}^2 \approx \sigma_{D^*}^2$ is very accurate. In the following experiment, the target distribution had on average only three or four more signal dots per frame than the distractor distribution, so the numbers of target and distractor noise dots were approximately equal, and the bias introduced by this approximation is small compared to experimental error. Second, we used stimuli with different numbers of target and distractor signal dots only in order to make our stimuli as similar as possible to those in earlier studies (e.g., Edwards & Badcock, 1994), and it would be easy to do away with the approximation $\sigma_{T^*}^2 \approx \sigma_{D^*}^2$ simply by using an equal number of target and distractor noise dots. In any case, in a task where this approximation is inadequate, it should be clear how Equations 30 through 33 could be rederived without the approximation.

We will refer to the correlation method we used in Experiment 1 as the sampling noise method, because it measures the effect of statistical fluctuations in the targets and distractors on the observer’s responses, and we will refer to the method we just described as the probe method, because it measures the effect of small target and distractor signals on the observer’s responses. We will refer to both methods as correlation methods because they determine whether small variations in the distractors are correlated with observers’ responses. The probe method is similar to a perturbation method developed by Kinchla (1977) to measure the influence of two or more redundant stimulus properties on an observer’s responses, and it is similar in principle to Landy, Maloney, Johnston, and Young’s (1995) method of using signal perturbations

to study how observers combine several different estimates of an object's depth.

In Experiment 2, we used the probe method to make another estimate of the attentional weight that observers assign to black distractor dots when judging the global direction of motion of white target dots.

Methods

Participants

One author (R.F.M.) and four University of Toronto students participated. One observer (R.F.M.) was practiced at direction discrimination with random dot cinemato-grams, had participated in Experiment 1, and was aware of the hypotheses being investigated. The other four observers were not practiced at this task, had not participated in Experiment 1, and were unaware of the hypotheses.

Stimuli

The stimuli were the same as in the 50L50L and 50L50D conditions of Experiment 1, except that the distractor dots included a number of signal dots that moved to the left or to the right. The number of distractor signal dots was half the number of target signal dots, and on each trial the direction of the distractor signal dots was chosen independently of the direction of the target signal dots.

Procedure

The procedure was the same as in Experiment 1,

except that each observer participated in only two 1-hr sessions. As in Experiment 1, the number of target signal dots per frame was fixed at a number found during a pilot session to give approximately 70% correct performance. The numbers of target signal dots per frame were observer A.J.R., 10 dots; K.E.H., 8 dots; R.F.M., 4 dots; S.A.K., 8 dots; and T.F.S., 6 dots.

Results and Discussion

50L50L Condition

Table 2 shows the probability of each observer responding "right," conditional on the target and distractor signal dots moving left or right, in the 50L50L condition. Table 2 also shows the estimates of k and d'_T that we calculated from these conditional response probabilities, using a simplex search to find the values of k and d'_T that minimized the sum-of-squares error between the left- and right-hand sides of Equations 30 through 33. The estimates of k ranged from 0.86 to 1.12, and the mean estimate across observers was 0.99 ± 0.04 . Neither the individual estimates nor the mean estimate were significantly different from the anticipated value of 1 ($p > .20$ for all comparisons).

The observers' target sensitivities d'_T ranged from 0.66 to 2.02. Although we chose the number of signal dots to maintain 70% correct performance based on a pilot session, some observers performed markedly better or worse than this in the main experiment. This is reflected in the wide range of values of d'_T .

Table 2. Results of Experiment 2, 50L50L Condition

		Distractor R	Distractor L	Attentional weight k	Sensitivity d'_T
A.J.R.	Target R	0.75	0.55	1.12 ± 0.14	1.44 ± 0.11
	Target L	0.38	0.21		
K.E.H.	Target R	0.83	0.67	0.86 ± 0.11	1.59 ± 0.10
	Target L	0.41	0.22		
R.F.M.	Target R	0.86	0.65	0.98 ± 0.08	2.02 ± 0.10
	Target L	0.36	0.14		
S.A.K.	Target R	0.60	0.54	0.95 ± 0.27	0.66 ± 0.09
	Target L	0.44	0.32		
T.F.S.	Target R	0.71	0.59	1.02 ± 0.16	1.18 ± 0.10
	Target L	0.43	0.24		

The first two columns of numbers show the proportion of trials on which the observer responded "right," conditional on the target and distractor signal dots moving left or right. For example, the top left cell shows that observer A.J.R. responded "right" on 75% of the trials on which both the target and the distractor signal dots moved to the right. The third and fourth columns show the attentional weight k and the target sensitivity d'_T calculated from these conditional response probabilities using the methods described in the text. The error values are SEs.

Table 3. Results of Experiment 2, 50L50D Condition

		Distractor R	Distractor L	Attentional weight k	Sensitivity d'_T
A.J.R.	Target R	0.78	0.65	0.62 ± 0.10	1.60 ± 0.09
	Target L	0.29	0.15		
K.E.H.	Target R	0.78	0.70	0.48 ± 0.10	1.48 ± 0.08
	Target L	0.31	0.19		
R.F.M.	Target R	0.87	0.68	0.73 ± 0.08	1.96 ± 0.10
	Target L	0.30	0.15		
S.A.K.	Target R	0.87	0.82	0.22 ± 0.08	1.80 ± 0.08
	Target L	0.26	0.20		
T.F.S.	Target R	0.80	0.66	0.54 ± 0.10	1.42 ± 0.08
	Target L	0.31	0.23		

See notes for [Table 2](#).

50L50D Condition

Table 3 shows each observer's conditional response probabilities and the corresponding estimates of the attentional weight k and target sensitivity d'_T , in the 50L50D condition. The estimates of k ranged from 0.22 to 0.73, and the mean estimate across observers was 0.52 ± 0.09 . Each individual estimate of k , as well as the mean estimate, was significantly greater than zero and significantly less than 1 ($p < .01$ for all comparisons), and significantly less than the corresponding value of k in the 50L50L condition ($p < .05$ for all comparisons). Despite individual differences, all observers had a limited ability to direct attention according to contrast polarity: all were appreciably influenced by opposite-polarity distractors, but not as much as by same-polarity distractors.

The values of attentional weight we measured in this experiment using the probe method were similar to those we measured in Experiment 1 using the sampling noise method, although they tended to be slightly lower. Furthermore, the SEs for k that we obtained in this experiment were at least as small as the SEs obtained in Experiment 1, even though we collected fewer than one third as many trials per observer in this experiment. When we applied the sampling noise method used in Experiment 1 to this experiment's data, calculating the covariance ratio within each of the four clusters of trials (RR, RL, LR, and LL) and averaging the resulting four estimates of k , we found that the SEs were at least twice as large as the SEs obtained with the probe method, and in one half the cases were greater than 1, rendering the estimates of k practically useless. Clearly, the probe method is a more efficient way of measuring the attentional weight that observers assign to distractors.

Relation Between the Two Correlation Methods

The sampling noise method that we used in Experiment 1 measures the influence of small statistical variations in the targets and distractors on the observer's responses, whereas the probe method that we used in this experiment measures the influence of larger target and distractor signals on the observer's responses. The probe method assumes that the target and distractor signals do not change the observer's strategy. In this experiment, we assume that the attentional weight k and the target sensitivity d'_T are the same on all trials, regardless of whether the distractor signal is in the same direction or the opposite direction as the target signal. This assumption is reasonable, but it could be false. It could be that when the target and distractor dots move in opposite directions, the observer perceives two transparent sheets of dots sliding over one another, and that this perceptual segregation helps the observer ignore the distractors. If this were so, the attentional weight would be lower on opposite-direction trials. In contrast, the sampling noise method used in Experiment 1 relies on small statistical variations in the stimulus, making it unlikely that the observer's decision rule changes systematically from one stimulus to the next. To state the problem more generally, the probe method introduces larger variations into the stimulus, and this makes the method faster but also relies on the assumption that these variations do not change the observer's decision rule. This is the reason we present both methods here, even though the more efficient probe method is to be preferred in tasks where its assumptions are met. In our experiments, we obtained similar results with both methods, indicating that the slightly stronger

assumptions of the probe method were at least approximately satisfied.

A Static Task

Our third experiment had two purposes. First, the paradigms we used in the first two experiments were unusual for studies of attention because we analyzed observers' performances in a single condition, rather than comparing performance across two conditions that had identical stimuli but different instructions to the observer. In the next experiment, we show how the methods we have proposed can be used in a more traditional instruction-manipulation paradigm. Second, we were surprised to find that observers had such a limited ability to direct attention according to contrast polarity (although this is consistent with previous reports), and we wished to see whether this finding would generalize to other tasks. In the next experiment, we measured observers' abilities to direct attention according to contrast polarity when judging the global orientation of a static pattern.

The stimulus was a static analog of the random dot cinematograms we used in the first two experiments, consisting of 50 white and 50 black elongated Gaussian blobs (Figure 9). The orientations of the white blobs were normally distributed with mean μ_W , and the orientations of the black blobs were normally distributed with mean μ_B . The SDs of the white and black orientation distributions were the same. On each trial, the mean orientations μ_W and μ_B were randomly and independently set to a small, fixed angle μ clockwise or

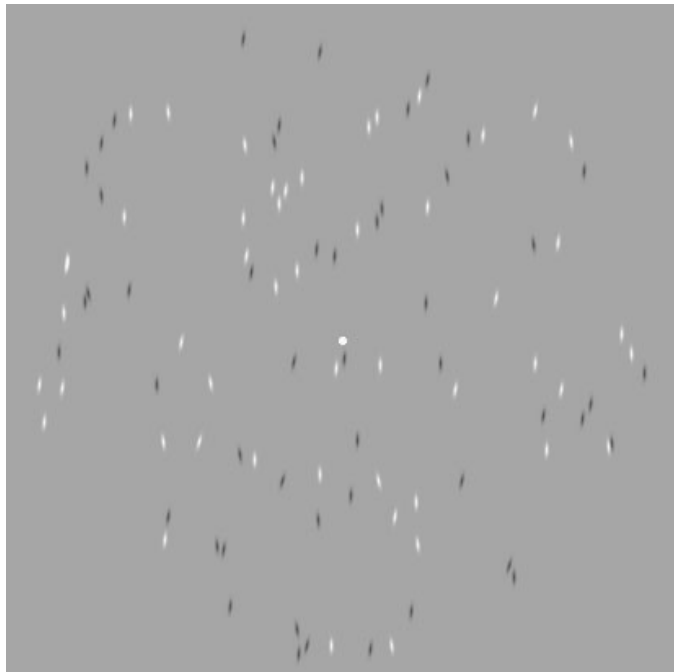


Figure 9. Stimulus in Experiment 3. [Click to view movie](#)

counterclockwise of vertical. In the White condition, observers judged whether the mean orientation of the white blobs was clockwise or counterclockwise of vertical, and in the Black condition, observers judged whether the mean orientation of the black blobs was clockwise or counterclockwise of vertical.

In this task, the distractors contained a nonzero orientation signal, so we used the probe method to measure attentional weight assigned to distractor blobs. Here, though, the distractor signal was fully as strong as the target signal, whereas in Experiment 2, the distractor signal was only half as strong, so we need new expressions for calculating attentional weight from observers' conditional response probabilities.

As in the previous experiments, we assume that the observer's decision variable is described by Equations 24 through 26. The decision variable is a weighted sum $s = T^* + kD^*$, and in this task $\Delta\mu_{T^*} = \Delta\mu_{D^*}$ and $\sigma_{T^*}^2 = \sigma_{D^*}^2$. Proceeding exactly as in the introduction to Experiment 2 ("The Probe Method of Measuring Attentional Weight"), we can write the probability of the observer responding clockwise when both the target and the distractor are clockwise of vertical as

$$P_{CW,CW} = P(T^* + kD^* \geq \mu_{T^*} + k\mu_{D^*}) \quad (34)$$

$$= 1 - G(\mu_{T^*} + k\mu_{D^*}, \mu_{T^*} + k\mu_{D^*} + 0.5(1+k)\Delta\mu_{T^*}, \sqrt{1+k^2}\sigma_{T^*}) \quad (35)$$

$$= G\left(\frac{(1+k)d'_T}{2\sqrt{1+k^2}}\right). \quad (36)$$

Similarly, on trials where the target is clockwise and the distractor is counterclockwise, the probability of a clockwise response is

$$P_{CW,CCW} = G\left(\frac{(1-k)d'_T}{2\sqrt{1+k^2}}\right). \quad (37)$$

The same analysis applies to trials where the target distribution is counterclockwise of vertical:

$$P_{CCW,CCW} = 1 - G\left(\frac{(1+k)d'_T}{2\sqrt{1+k^2}}\right) \quad (38)$$

$$P_{CCW,CW} = 1 - G\left(\frac{(1-k)d'_T}{2\sqrt{1+k^2}}\right) \quad (39)$$

As in Experiment 2, we used a simplex search to find the best-fitting values of k and d'_T , given the measured response probabilities.

In Experiment 3, we illustrate the probe method in a more traditional attention paradigm, in that the stimulus is the same in all conditions, and we only change the

instructions to the observer. All stimuli contained equal numbers of oriented white and black blobs, and in different conditions we instructed observers to judge the mean orientation of only the white or the black blobs.

Methods

Participants

Three undergraduate University of Toronto students participated. One observer (S.U.M.) had participated in Experiment 1. All observers were unpracticed at the task and unaware of the hypotheses being tested.

Stimuli

The stimuli showed 100 two-dimensional Gaussian blobs in a circular aperture of radius 5.5 deg (Figure 9). The contrast profile of a vertical blob centered at the origin was $C(x, y) = c_0 g(x, 0, \sigma_W) g(y, 0, \sigma_L)$, where g is the normal probability density function, and the scale constants were $\sigma_W = 0.03$ deg and $\sigma_L = 0.12$ deg. Fifty of the blobs were white, and 50 were black (peak Weber contrast ± 0.40). The orientations of the white blobs were normally distributed with a mean μ_W ° clockwise or counterclockwise of vertical, and a SD of 5°. The orientations of the black blobs were also normally distributed with a mean μ_B ° clockwise or counterclockwise of vertical, and a SD of 5°. The mean orientations of the white and black subsets were randomly and independently set to μ ° clockwise or counterclockwise of vertical on each trial. The mean angle μ was chosen individually for each observer, as explained in "Procedure." The stimuli were shown on a gray background of luminance 40 cd/m². The stimulus duration was 200 ms.

Stimuli were shown on the same monitor as in the first two experiments. Observers viewed the stimuli binocularly from a distance of 1.00 m, and head position was stabilized using a chin-and-forehead rest.

Procedure

Each observer participated in three 1-hr sessions. Each session consisted of six to eight blocks of 100 trials. One half the blocks were White blocks, one half were Black blocks, and the session alternated between the two types of blocks. At the beginning of each White or Black block, the observer was instructed to judge the mean orientation of the white or black blobs, respectively, and to ignore the blobs of the opposite contrast polarity. Each trial began with a 2,200-ms fixation interval, followed by a 200-ms stimulus, followed by a response interval in which the observer pressed one of two keys to indicate whether the mean orientation of the attended blobs was clockwise or counterclockwise of vertical. Auditory feedback indicated whether the observer's response was correct. A small white fixation dot appeared at the center of the screen throughout the White blocks, and a small black fixation dot appeared throughout the Black blocks to remind the observer which contrast polarity to attend to. For each observer, the mean orientation μ from vertical was fixed at a value found during a pilot session to give approximately 70% correct performance. For observers J.A.P. and L.C.S., this was 2.5°, and for S.U.M., it was 2.0°. Over the course of three sessions, each observer ran in approximately 1,100 trials in each condition (White and Black).

Results and Discussion

Tables 4 and 5 show each observer's conditional response probabilities and the corresponding estimates of the attentional weight k and target sensitivity d'_T , in both White and Black conditions. Estimates of k ranged from 0.13 to 0.43, and the average estimate across observers and conditions was 0.27. Each estimate of k was significantly greater than zero and significantly less than 1 ($p < .05$ in all comparisons), and the estimates were not significantly different across the White and Black conditions, although the difference across conditions approached significance for observer J.A.P.

Table 4. Results of Experiment 3, White Condition

		Distractor CW	Distractor CCW	Attentional weight k	Sensitivity d'_T
J.A.P.	Target CW	0.81	0.64	0.43 ± 0.07	1.39 ± 0.09
	Target CCW	0.36	0.17		
L.C.S.	Target CW	0.64	0.58	0.18 ± 0.07	1.25 ± 0.08
	Target CCW	0.21	0.14		
S.U.M.	Target CW	0.79	0.69	0.33 ± 0.09	0.99 ± 0.08
	Target CCW	0.45	0.33		

CW = clockwise; CCW = counterclockwise. See caption of Table 2 for details.

Table 5. Results of Experiment 3, Black Condition

		Distractor CW	Distractor CCW	Attentional weight k	Sensitivity d'_T
J.A.P.	Target CW	0.74	0.67	0.24 ± 0.07	1.28 ± 0.09
	Target CCW	0.31	0.18		
L.C.S.	Target CW	0.65	0.60	0.13 ± 0.06	1.19 ± 0.08
	Target CCW	0.22	0.17		
S.U.M.	Target CW	0.81	0.67	0.28 ± 0.09	0.98 ± 0.09
	Target CCW	0.41	0.37		

CW = clockwise; CCW = counterclockwise. See caption of Table 2 for details.

In this experiment, unlike in the first two experiments, observers were instructed to perform different tasks with the same stimuli. We found that a simple change in instructions led to a large change in the attentional weight assigned to the white and black blobs, and the effects of instructions were mostly symmetric: in the White and Black conditions, performance levels were approximately the same, and an attentional weight on the order of 0.25 was assigned to the distractors. All observers were largely able to direct their attention according to contrast polarity, although the distractors nevertheless had an appreciable effect on observers' responses.

General Discussion

Weighted Sum Models

The idea that observers combine information from different sources in a weighted sum has been used by many authors to describe performance in many different tasks. An early example is Green (1958), who suggested that when observers try to detect an auditory signal with components at two widely spaced frequencies, they monitor two channels centered at the two frequencies, and use a decision variable that is the sum of the outputs of the two channels. More recently, Landy and colleagues (Johnston, Cumming, & Landy, 1994; Landy et al., 1995; Young, Landy, & Maloney, 1993) have shown that depth estimates obtained from binocular disparity, texture gradients, and motion parallax are combined in a weighted average to yield a single estimate of an object's depth. Landy and Kojima (2001) have developed a similar model for edge localization. In a series of studies of visual attention, Kinchla (1969; 1974; 1977; 1980; 1995; Kinchla & Collyer, 1974) suggested that in visual search tasks, observers base their responses on a weighted sum of decision variables corresponding to possible target locations. Also relevant to our study is Blaser, Sperling, and Lu's (1999) account of selective attention to color, in which signals from attended stimuli are amplified before

being combined with signals from unattended stimuli. All these accounts propose that observers combine information from two or more sources in a weighted sum. In most cases, the authors present the weighted sum as a simple, plausible hypothesis as to how information is combined from several sources, but give little theoretical motivation for this form of decision rule. The main exceptions are in the cue combination literature, where it is often noted that the optimal way of combining several noisy estimates of a single quantity is with a weighted average (e.g., Landy et al., 1995). Our derivation of attentional weight in the "Introduction" shows that in selective attention tasks, there are also good reasons why observers might combine information this way.

It is worth noting that the weighted sum model describes performance well even in tasks where it is not the optimal method of combining information from different sources. In Kinchla's (1995) experiments, for example, observers tried to detect a target that could appear at only one of four cued or uncued locations on each trial. In this task, the decision variables corresponding to the four locations were not independent, and the optimal Bayesian decision rule is a nonlinear function of the four decision variables. In Kinchla's earlier studies (Kinchla, 1969, 1974, 1977; Kinchla & Collyer, 1974), observers also made decisions based on several statistically dependent information sources. Nevertheless, in all these studies, the weighted sum hypothesis gave a good description of many aspects of observers' performances (although Kinchla did not directly compare the weighted sum model with the optimal decision rule). It may be that the visual system has a limited repertoire of decision strategies, and uses simple, easily computable strategies, such as weighted sum of internal responses, even in tasks where they are not optimal.

Why Measure Attentional Weight?

We believe that attentional weight has several advantages over some other proposed measures of selective attention.

First, one shortcoming of some studies of selective attention is that they pose a yes-no question of the form, "Can observers direct attention according to X?" and have no natural way of giving a quantitative answer between yes and no. With attentional weight, we can describe the role of selective attention with a single continuous parameter that describes the relative influence of targets and distractors on the observer's responses.

Second, attentional weight is designed to be invariant across experimental paradigms, and hence it attempts to measure a characteristic of the observer, and not just to quantify the effects of attention in a particular task. For example, Equations 30 through 33 for attentional weight in Experiment 2 assume that there are an equal number of target and distractor dots in the stimulus; but, if we were to modify the task so that there were 50 target dots and only 25 distractor dots, we could derive new expressions for k suited to this new task. So long as our model of the observer's decision variable is correct, the measured value of k will be the same in both experiments. This is not true of many other measures of selective attention, such as the difference in reaction times between trials where the response suggested by distractors is consistent or inconsistent with the response suggested by targets (e.g., Garner & Felfoldy, 1970), or the difference between direction discrimination thresholds when distractors have the same polarity or the opposite polarity as the targets (e.g., Edwards & Badcock, 1994). In this sense, attentional weight is like the signal detection theory measure of sensitivity, d' : it attempts to measure a property that is invariant across experimental designs and perceptual tasks, and hence is truly a characteristic of the observer. Consequently, we need to use different expressions for measuring attentional weight in different tasks, just as we use different expressions for d' (Macmillan & Creelman, 1991). The advantage, however, is that we can meaningfully compare the efficacy of selective attention across different tasks.

Finally, attentional weight has a straightforward interpretation in terms of how the observer uses available information to perform a task: it measures the observer's relative weighting of the evidence, in the sense of log likelihood ratios, provided by attended and unattended parts of the stimulus. We can contrast this with approaches in which attention is measured by a parameter in a mathematical model that adequately describes observers' performances, but that has little theoretical motivation (e.g., Kinchla & Collyer, 1974), or in which attention is measured by a parameter in a specific computational model of visual processing (e.g., Blaser, Sperling, & Lu, 1999). Certainly, both the latter approaches are useful, and we do not mean to suggest

that a more abstract Bayesian measure is always to be preferred. Rather, we believe that in addition to these approaches, it is useful to have a measure that is theoretically motivated, and yet is sufficiently abstract to be independent of specific models of visual processing. If we measured visual selective attention with a parameter in a particular model of visual processing, and we measured auditory selective attention with a parameter in a very different model of auditory processing, it could be difficult to compare the effects of selective attention across these conditions. On the other hand, just as ideal observer analysis allows us to compare performance on an absolute scale (i.e., efficiency) across very different tasks, a Bayesian measure such as attentional weight allows us to meaningfully compare the efficacy of selective attention across different tasks.

Why Use Correlation Methods?

The correlation methods that we have presented allow us to measure selective attention by analyzing an observer's performance in a single task, and this is perhaps the most significant difference between these methods and traditional methods. It is worth stating the advantages of this approach once more, and to contrast it with other approaches. In one common type of selective attention task, we instruct the observer to attend to a set of targets, and we compare performance across conditions with and without distractors, or across conditions with different types of distractors (e.g., Edwards & Badcock, 1994; Garner & Felfoldy, 1970). In another common type of task, we hold the stimulus constant, and compare performance across conditions in which the observer is instructed to attend to different aspects of the stimulus (e.g., Stroop, 1935). In both paradigms, we compare performance across two conditions, and there is always the possibility that performance differs across the two conditions for reasons that have nothing to do with attention. With correlation methods, on the other hand, we instruct the observer to attend to the targets, and we measure the effect of both targets and distractors on the observer's responses in a single task. If the distractors are correlated with the observer's responses, this shows that the observer cannot base his responses on only the targets. Indeed, it is difficult to imagine any more direct evidence than this, or to see how this result could ever be due to a confound.

For specific examples of difficulties in comparing performance across conditions, consider Edwards and Badcock's (1994) study that investigated whether observers can direct attention according to contrast polarity when judging global direction of motion. Edwards and Badcock compared direction discrimination thresholds in a condition where targets and distractors had the same contrast polarity, as in our 50L50L condition, with thresholds in a condition where targets and distractors had opposite contrast polarities, as in our

50L50D condition. The logic of this approach is clear: if observers can selectively attend to a single contrast polarity, then performance in the presence of opposite-polarity distractors should be better than performance in the presence of same-polarity distractors, and if observers cannot selectively attend to a single contrast polarity, then performance should be the same in the two conditions.

However, there are several ways in which opposite-polarity distractors might worsen performance even if observers can attend to a single contrast polarity. First, there is evidence that judgments of global motion in random dot cinematograms are largely based on low spatial frequency components (Barton, Rizzo, Nawrot, & Simpson, 1996), at least when the step size is larger than 0.20 deg, as it is in our stimuli. This poses a problem for the approach of comparing performance across conditions, because if a 50L50D cinematogram is low-pass filtered, the black and white dots blur onto one another, and positive and negative contrast polarities partially cancel out. This reduction in effective contrast may offset any improvement in performance resulting from selectively attending to the target dots. Second, and more difficult to quantify, is the confound that in the 50L50L condition observers simply judge the mean direction of the entire cinematogram, whereas in the 50L50D condition, they restrict their attention to the white dots, and judge the mean direction of only this part of the cinematogram. The mental effort required to perform the second, more complex task may offset any improvement resulting from attending to the target dots. Third, the observer may be able to selectively attend to the white dots, but at the expense of using a less efficient strategy. For instance, if the observer used only a small number of white dots in the 50L50D condition, his threshold might be the same as in the 50L50L condition, even though he used only white dots to perform the task.⁵

These are just three examples of difficulties in comparing performance across two or more conditions. Furthermore, there are other difficulties that may not arise with the present tasks, but that in general could make it difficult to compare different conditions: the distractors could change the observer's sensitivity to the target stimuli, the distractors could increase the level of internal noise, and so on. Perhaps control experiments could rule out these and other confounds, but the point of these examples is that when we compare performance across different conditions, we always face the possibility of confounds. Our correlation methods avoid all problems of this kind, because as we have emphasized, they do not compare performance across different tasks. Instead, they measure the correlation between selected parts of the stimulus and the observer's responses in a single task, and thereby reveal the influence of the distractors on the observer's responses. This eliminates a whole range of possible confounds.

For instance, all three of the confounds that we just attributed to earlier studies of global direction

discrimination are due to differences between the 50L50L and 50L50D conditions. In our experiments, we estimated attentional weight by analyzing performance in only the 50L50D condition, and we included the 50L50L condition simply to validate the correlation methods in a task where we knew the correct value of attentional weight. Obviously, then, our experiments avoid all problems relating to confounds between the 50L50D and 50L50L conditions, because the 50L50L condition plays no essential role in our measurement of the attentional weight assigned to opposite-polarity distractors.

Range of These Methods

It may be helpful at this point to consider when we can and cannot use the methods we have presented.

First, what kinds of attentional effects can we measure with attentional weight? Attentional weight measures the relative influence of targets and distractors on an observer's responses, so it is useful primarily as a measure of *selective* attention (i.e., an observer's ability to judge selected stimuli in a visual scene and to ignore others). As Kinchla (1974; Kinchla & Collyer, 1974) has shown, attentional weight is also a useful measure in tasks involving distributed attention, such as cued detection tasks and visual search. On the other hand, we can see no obvious way of using attentional weight, as we have defined it, to study some other tasks normally thought to involve attention, such as dual tasks. Furthermore, as we have emphasized, attentional weight is an appropriate measure only if attending to or away from a stimulus does not qualitatively change how an observer processes the stimulus. For instance, if observers had very different directional selectivities for attended and unattended dots in random dot cinematograms, as well as weighting attended and unattended stimuli differently, then a scalar measure would be inadequate. We would need a more flexible approach (e.g., we could measure the directional selectivity for attended and unattended stimuli using the HCA method demonstrated in Experiment 1). In this respect, our account of selective attention is similar to recent accounts of visual search (Eckstein, 1998; Palmer, 1994; Shaw, 1984), in that it implies that selective attention does not change the observer's representation of a stimulus.

Second, when are correlation methods useful? Again, because correlation methods compare the influence of targets and distractors on an observer's responses, they are most useful for studying tasks involving selective attention, where we are interested in whether an observer can make judgments based only on targets, and ignore distractors. Furthermore, the correlation methods we have presented are most useful when we believe that observers combine internal responses to attended and unattended stimuli in a weighted sum, because in this case they allow us to measure the relative weight assigned to targets and distractors. Nevertheless, as we pointed out earlier, correlation methods can be useful even when we

have little idea how the observer performs the task. In Experiment 1, for instance, we found that statistical fluctuations in the black distractors had almost as large an effect on observers' responses as fluctuations in the white targets. If our weighted sum model of performance in this task is wrong, then our estimate of the precise value of attentional weight might be mistaken, but it is nevertheless difficult to see any way around the conclusion that observers cannot restrict their attention to the white targets.

An Extension to Arbitrary Targets and Distractors

In our experiments, the targets and distractors were similar stimuli, but the framework we have presented can easily be adapted to tasks where the targets and distractors are qualitatively different. In the Stroop task, for example, an observer reports the color of ink in which a color name is written, or reads a color name written in colored ink (Stroop, 1935). Here, the targets and distractors are either word identity or ink color. Using the probe method of Experiments 2 and 3, we could measure the attentional weight assigned to distractors even in this task.

Consider how we might do this. First, we could choose a set of stimuli where the targets and distractors were equally discriminable (e.g., the words RED and GREEN written in red and green ink, with contrast and chromaticity chosen so that color word identification was 75% accurate when ink color was held constant, and ink color naming was 75% accurate when word identity was held constant). With these equally discriminable stimuli, the internal responses to the target and distractor stimuli, T^* and D^* , would have the same sensitivity d' to the target and distractor discrimination tasks. In a Stroop task where both word identity and ink color are randomly chosen on each trial, and the observer reports, say, the word identity, the weighted sum hypothesis holds that the observer's decision variable is $s = T^* + kD^*$. The problem of measuring attentional weight in this Stroop task is exactly the same as the problem we faced in Experiment 3: the targets and distractors are equally discriminable, and the observer uses a decision variable of form 24. Hence, we could simply use Equations 36 through 39 that we used in Experiment 3 to measure the attentional weight assigned to distractors in this appropriately constructed Stroop task.⁶

Melara and Mounts (1993) showed that the relative discriminability of colors and words has a large effect on performance in Stroop tasks. In fact, they found that the well-known asymmetry wherein color word identity interferes with ink color naming, but not vice versa, was largely or entirely due to the greater discriminability of color words. One advantage of using the probe method to measure selective attention is that it explicitly takes account of the relative discriminability of targets and distractors, so that measurements of attention are not

confounded with low-level differences between targets and distractors. Furthermore, the probe method is not restricted to tasks where the targets and distractors are equally discriminable. The distractors may be less discriminable than the targets (as in Experiment 2 where the mean horizontal displacement of the distractors was one half the displacement of the targets), equally discriminable (as in Experiment 3 where the orientation difference of the clockwise and counterclockwise distractors was as large as the orientation difference of the targets), or even more discriminable, and so long as we measure the discriminability of targets and distractors, we can use the probe method to estimate the attentional weight assigned to distractors.

A More General Model: Three Ways That Selective Attention May Fail

Selective attention is often said to have "failed" in a task if we measure an observer's performance with and without distractors, and find that performance is worse in the presence of distractors (e.g., Garner & Felfoldy, 1970). If the observer's responses are based on a decision variable $s = T^*$ in the no-distractor condition and a decision variable $s = T^* + kD^*$ in the distractor condition, as in Equations 24 through 26, then the only ways that selective attention can fail are for the observer to assign a nonzero attentional weight to the distractors, or for the target component T^* of the decision variable to be computed less efficiently in the distractor condition than in the no-distractor condition. Furthermore, the only ways that T^* may be computed less efficiently in the distractor condition are either for the selectivity function f to be less efficient or for the internal noise Z_T to be higher. Thus we obtain a simple taxonomy of the ways selective attention may fail: attentional weight may be assigned to distractors, target selectivity may be impaired by distractors, and internal noise may be increased by distractors.

Throughout this work we have argued that if an observer truly cannot direct attention away from the distractors, then small variations in the distractors should influence the observer's responses. We have considered it an advantage of the correlation methods we have presented that they measure the influence of distractors on an observer's responses in a single condition, and are not confounded by performance differences across distractor and no-distractor conditions that arise for other reasons, such as an increase in internal noise. In some cases, though, we may want to know the *complete* effect that distractors have on performance, including both the correlation they have with the observer's responses, and also any performance decrements they cause by other means.

The methods we have used in these experiments, and closely related methods, can easily be adapted to quantify all three types of failure of selective attention. First, we

have already shown how to measure the attentional weight assigned to distractors. Second, we have shown how the HCA method can be used to compare the computations performed on attended and unattended stimuli. In exactly the same way, we could use HCA to compare the computations performed on targets alone and targets in the presence of distractors, to see whether the presence of distractors impairs the observer's processing of the targets. Third, we could use the two-pass method developed by Green (1960) and Burgess and Colborne (1988) to measure the internal noise that limits observers' performances, and to see whether the presence of distractors makes an observer's decision variable noisier. With these methods, we could arrive at a fairly complete characterization of how distractors affect an observer's performance of a task. Similar methods have been used successfully to study the effects of attentional set (Lu & Doshier, 1998) and perceptual learning (Doshier & Lu, 1999; Gold, Bennett, & Sekuler, 1999).

Summary

Drawing on earlier studies of how observers combine information from two or more sources, we have defined a measure of selective attention, *attentional weight*, that measures the relative influence of targets and distractors on an observer's responses. We have presented two methods for estimating attentional weight by measuring the influence that targets and distractors have on an observer's responses. In three experiments, we showed that these methods give a description of observers' abilities to direct attention according to contrast polarity that is consistent both with our prior expectations, as in the same-polarity condition (50L50L) where we found an attentional weight of $k = 1$, and with previous empirical results, as in the opposite-polarity conditions (50L50D, White, and Black) where we found that observers had only a limited ability to direct attention according to contrast polarity. Furthermore, we found that selectivity was the same for attended and unattended stimuli in the direction discrimination tasks, as it must be if attentional weight is to be a valid measure of selective attention. Finally, we have shown that the weighted sum framework can be used to study selective attention in a broad range of tasks, and we have suggested how it could be extended to give a more thorough characterization of how selective attention affects performance.

Appendix A

Our main goal in this appendix is to show that the sampling noise method of measuring attentional weight, given in Equations 22 and 23 is valid for the broad class of observers described by Equations 24 through 26. However, to make our results as general as possible, we will derive a few simple statistical properties of these

observers' decision variables, and we will show that the method works for *any* observer whose decision variable has these properties. In particular, it should be clear that this method is not restricted to tasks involving random dot cinematograms, but can be used to study how an observer combines information from any two sources using a decision rule with the stated properties.

The Model

We assume that the cinematogram has N_T target dots and N_D distractor dots, that n_T target signal dots move directly left or right, that n_D distractor signal dots move directly left or right, and that the remainder of the dots move in random directions. (The stimuli in Experiment 1 had an equal number of target and distractor dots, $N_T = N_D$, and had no distractor signal dots, $n_D = 0$.) We will represent the cinematogram by a collection of multivariate random variables t_i and d_i that represent individual target and distractor dot displacements, respectively. Each t_i and d_i encodes any properties of the dot displacements that are relevant to the observer's responses. For instance, we could represent each dot displacement by a triplet (θ, x, y) that reports its direction θ and its position (x, y) . For convenience, we assume that the indices are ordered so that t_1, \dots, t_{n_T} represent the n_T target signal dots, and d_1, \dots, d_{n_D} represent the n_D distractor signal dots. We assume that $t_{n_T+1}, \dots, t_{N_T}$ and $d_{n_D+1}, \dots, d_{N_D}$, corresponding to the noise dots, are identically distributed. Properties of the target dots may be correlated ($\text{cov}[t_i, t_j] \neq 0$), as may properties of the distractor dots ($\text{cov}[d_i, d_j] \neq 0$), but we assume that the target and distractor dot distributions are independent ($\text{cov}[t_i, d_j] = 0$).

Let $g(t_i)$ and $g(d_i)$ be the horizontal component of dot displacements t_i and d_i , respectively. Then, the total horizontal target and distractor displacements are

$$T = \sum_{i=1}^{N_T} g(t_i) \quad (\text{A1})$$

$$D = \sum_{i=1}^{N_D} g(d_i). \quad (\text{A2})$$

The expected value of the total horizontal target displacement is $\mu_T = \pm n_T \Delta d$, and the expected value of the total horizontal distractor displacement is $\mu_D = \pm n_D \Delta d$, where Δd is the size of a single dot step and the signs depend on whether the signal dots move left or right.

Consider an observer who judges the mean direction (left or right) of a random dot cinematogram, using a decision variable described by Equations 24 through and 26. The decision variable has the following properties. First, as assumed in 24, the decision variable is a weighted sum of two quantities, T^* and D^* :

$$s = T^* + kD^* \quad (\text{A3})$$

Second, because T^* is calculated from the target dots and D^* is calculated from the distractor dots, T^* is uncorrelated with D , and D^* is uncorrelated with T :

$$\text{cov}[T^*, D] = \text{cov}[D^*, T] = 0 \quad (\text{A4})$$

Third, because T^* and D^* are obtained from the target and distractor dots using the same calculation, their covariances are approximately related by

$$(N_D - n_D) \text{cov}[T^*, T] = (N_T - n_T) \text{cov}[D^*, D]. \quad (\text{A5})$$

For later convenience, we will rewrite A5 as

$$\text{cov}[T^*, T] = (N_T - n_T)c \quad (\text{A6})$$

$$\text{cov}[D^*, D] = (N_D - n_D)c \quad (\text{A7})$$

where c is a constant. To prove Equation A6 and A7, we simply evaluate the covariances:

$$\text{cov}[T^*, T] = \text{cov} \left[\sum_{i=1}^{N_T} f(t_i) + Z_T, \sum_{j=1}^{N_T} g(t_j) \right] \quad (\text{A8})$$

$$= \sum_{ij=1}^{N_T} \text{cov}[f(t_i), g(t_j)] \quad (\text{A9})$$

$$\text{cov}[D^*, D] = \text{cov} \left[\sum_{i=1}^{N_D} f(d_i) + Z_D, \sum_{j=1}^{N_D} g(d_j) \right] \quad (\text{A10})$$

$$= \sum_{ij=1}^{N_D} \text{cov}[f(d_i), g(d_j)] \quad (\text{A11})$$

We can drop the unvarying signal dots from the sum, and write the covariances as

$$\text{cov}[T^*, T] = \sum_{ij=n_T+1}^{N_T} \text{cov}[f(t_i), g(t_j)] \quad (\text{A12})$$

$$= (N_T - n_T)c_1 + 0.5(N_T - n_T)(N_T - n_T - 1)c_0 \quad (\text{A13})$$

$$\text{cov}[D^*, D] = \sum_{ij=n_D+1}^{N_D} \text{cov}[f(d_i), g(d_j)] \quad (\text{A14})$$

$$= (N_D - n_D)c_1 + 0.5(N_D - n_D)(N_D - n_D - 1)c_0. \quad (\text{A15})$$

Here we have defined

$c_1 = \text{cov}[f(t_i), g(t_i)] = \text{cov}[f(d_i), g(d_i)]$ and $c_0 = \text{cov}[f(t_i), g(t_j)] = \text{cov}[f(d_i), g(d_j)]$, where $i \neq j$.

The c_1 term is the dominant term in Equations A13 and A15, as c_1 measures how strongly the effect of a single randomly chosen noise dot displacement on T is correlated with the effect of the same dot displacement on T^* , and this correlation may be large. On the other hand, c_0 measures how strongly the effect of a single randomly chosen noise dot displacement on T is

correlated with the effect of a *different* randomly chosen noise dot displacement on T^* , and for any reasonably large cinematogram, this correlation is negligible. The correlation need not be zero, because some properties of different noise dots displacements may be correlated (e.g., in our cinematograms, the lifetime of each dot was eight frames, so the positions of two randomly chosen dots on different frames were weakly correlated). (In fact, because the directions of individual dots are chosen independently, any of a number of reasonable assumptions about direction selectivity imply that the correlation is zero, for example, that the selectivity function f has equal and opposite responses to dots moving in opposite directions. However, we do not wish to introduce ad hoc assumptions at this point.) Consequently, we will neglect the c_0 term, and approximate the covariances as in A6 and A7.

Alternatively, in a task where we suspect that the correlation c_0 is appreciable, we can construct our stimuli so that $N_T - n_T = N_D - n_D$, which according to A13 and A15 implies that $\text{cov}[T, T^*] = \text{cov}[D, D^*]$ for any values of c_0 and c_1 , and A6 and A7 follow trivially.

Finally, the central limit theorem ensures that T and T^* are approximately jointly normal, and also that D and D^* are approximately jointly normal.

We will show that Equations 22 and 23 give unbiased measures of attentional weight for any observer whose decision variable has properties A3, A4, and A5, and who has internal responses T^* and D^* , such that T and T^* are jointly normal, and D and D^* are jointly normal.

Proof

Let C be a random variable that equals +1 or -1 on trials where the target signal dots move right or left, respectively, and let R be a random variable that equals +1 or -1 on trials where the observer responds “right” or “left,” respectively. Consider the trials on which the target and distractor signal dots move to the right. The expected value of T over all such trials where the observer responds “right” is $E[T | C = 1 \text{ and } R = 1] = E[T | T^* + kD^* \geq a]$, and the expected value of T over all such trials where the observer responds “left” is $E[T | C = 1 \text{ and } R = -1] = E[T | T^* + kD^* < a]$, where a is the observer’s response criterion. These expressions denote the expected value of the normal random variable T , conditional on the correlated normal random variable $T^* + kD^*$ falling above or below a criterion. In Appendix C, we show that if (X, Y) are jointly normal random variables with covariance c_{XY} , then

$$E[X | Y \geq a] = \mu_X + \frac{c_{XY}g(z)}{\sigma_Y G(-z)},$$

and

$$E[X | Y < a] = \mu_X - \frac{c_{XY}g(z)}{\sigma_Y G(z)},$$

where $\mu_Y = E[Y]$, $\sigma_Y^2 = \text{var}[Y]$, and $z = (a - \mu_Y) / \sigma_Y$. Here g is the standard normal probability density function, and G is the standard normal cumulative distribution function. Hence if we define $\mu_s = E[s]$, $\sigma_s^2 = \text{var}[s]$, and $z = (a - \mu_s) / \sigma_s$, then the conditional expected values in question are

$$E[T | C = 1 \text{ and } R = 1] = \mu_T + \frac{(N_T - n_T)c}{\sigma_s} \cdot \frac{g(z)}{G(-z)} \quad (\text{A16})$$

$$E[T | C = 1 \text{ and } R = -1] = \mu_T - \frac{(N_T - n_T)c}{\sigma_s} \cdot \frac{g(z)}{G(z)}. \quad (\text{A17})$$

Here c is the constant introduced in [Equations A6](#) and [A7](#). Similarly, the conditional mean distractor displacements are

$$E[D | C = 1 \text{ and } R = 1] = \mu_D + \frac{k(N_D - n_D)c}{\sigma_s} \cdot \frac{g(z)}{G(-z)} \quad (\text{A18})$$

$$E[D | C = 1 \text{ and } R = -1] = \mu_D - \frac{k(N_D - n_D)c}{\sigma_s} \cdot \frac{g(z)}{G(z)}. \quad (\text{A19})$$

The following more general form of [Equation 22](#) can be confirmed by direct substitution of [A16](#) through [A19](#):

$$k = \frac{(N_T - n_T) [E[D | C = 1, R = 1] - E[D | C = 1, R = -1]]}{(N_D - n_D) [E[T | C = 1, R = 1] - E[T | C = 1, R = -1]]} \quad (\text{A20})$$

If we set $N_T = N_D = N$ and $n_D = 0$, we obtain [Equation 22](#) as a special case. This is the equation that we used to calculate attentional weight in Experiment 1, using the target and distractor dot displacements in [Table 1](#).

We can also confirm that the more intuitive ratio of covariances in [Equation 23](#) correctly measures attentional weight for this broader class of observers. Over all trials where the correct answer is “right,” the covariance of the target dot displacement with the observer’s responses is

$$\begin{aligned} \text{cov}[T, R | C = 1] &= E[TR | C = 1] - E[T | C = 1]E[R | C = 1] \quad (\text{A21}) \\ &= E[T | C = 1, R = 1]P(R = 1 | C = 1) \\ &\quad - E[T | C = 1, R = -1]P(R = -1 | C = 1) \\ &\quad - E[T | C = 1]E[R | C = 1]. \end{aligned} \quad (\text{A22})$$

Substituting [Equations A16](#) and [A17](#), and using the fact that the probability of a “right” response is $G(-z)$, this covariance evaluates to

$$= 2(N_T - n_T)g(z) / \sigma_s. \quad (\text{A23})$$

Similarly, the covariance of the distractor displacement with the observer’s responses is

$$\text{cov}[D, R | C = 1] = 2k(N_D - n_D)g(z) / \sigma_s. \quad (\text{A24})$$

Taking the ratio of [A23](#) and [A24](#), we find

$$k = \frac{(N_T - n_T)}{(N_D - n_D)} \cdot \frac{\text{cov}[D, R | C = 1]}{\text{cov}[T, R | C = 1]}. \quad (\text{A25})$$

With $N_T = N_D = N$ and $n_D = 0$, we obtain [Equation 23](#) as a special case. This establishes that if an observer performs the same computation on target and distractor dots, and if selective attention uniformly reduces the influences of the distractor dots, then we can measure the attentional weight using [Equation 23](#) even if the computation yielding the decision variable is unknown and possibly stochastic.

Note that we have made no essential use of the fact that T and D are the total horizontal displacements of the targets and distractors. As we noted at the beginning of the proof, all that matters is that (a) T is uncorrelated with D^* , and D is uncorrelated with T^* , as in [A4](#), (b) $T \cdot \mu_T$ and $D \cdot \mu_D$ are noisy estimates of $T^* \cdot \mu_{T^*}$ and $D^* \cdot \mu_{D^*}$ to within a scale factor, as in [A5](#), and (c) the pairs T and T^* , and D and D^* , are jointly normal. In effect, we have chosen two measurable stimulus properties T and D as estimates of the unobservable internal responses T^* and D^* , and in this appendix, we have shown that we can use the relative influence of the observable variables on the subject’s responses to measure the relative influence of the unobservable variables on the subject’s responses, so long as T and D mirror T^* and D^* in these two respects.

Nevertheless, the better the estimates that T and D give of T^* and D^* , the more reliable our measurements of k will be. As we have pointed out, the sampling noise method can be seen as measuring the slope of the line connecting points M_L and M_R in [Figure 2](#), which are the mean target and distractor displacements over trials where the observer responds “left” and “right,” respectively. Our estimates of these points are noisy, simply because we can collect only a finite number of trials. This sampling error matters less when the distance between M_L and M_R is large. [Equations A16](#) through [A19](#) show that the distance between M_L and M_R grows with $\text{cov}[T, T^*]$ and $\text{cov}[D, D^*]$, so if we choose properties T and D that give good estimates of T^* and D^* (i.e., if T and D are strongly correlated with T^* and D^*), then the distance between the two points M_R and M_L will be large, and our estimates of k will be less variable.

Appendix B

We will represent a random dot cinematogram with n dot displacements as a collection of n random variables d_i , each assuming a value between $-\pi$ and π to indicate the direction of the corresponding dot displacement, with an angle of 0 indicating a dot moving directly to the right. In a noisy linear model of direction discrimination, we represent the observer’s decision variable as the sum of the responses that the dot displacements d_i evoke in a

filter, with an internal additive noise Z added as well. The observer responds “right” when the decision variable exceeds a criterion a . If we describe the directional selectivity of the filter with a function $f(\theta)$, we can write the decision variable as

$$s = \sum_i f(d_i) + Z . \quad (\text{B1})$$

To measure the directional selectivity $f(\theta)$, we will examine how the direction of a single dot affects the observer’s responses. We define $p_{\theta R}$ as the probability of the observer responding “right” when a particular dot d_k moves in direction θ , $p_{\theta R} = P(\text{‘right’} | d_k = \theta)$. Then,

$$p_{\theta R} = P\left(f(\theta) + \sum_{i \neq k} f(d_i) + Z \geq a\right) \quad (\text{B2})$$

$$= G\left(\frac{f(\theta) - \mu - a}{\sigma}\right) . \quad (\text{B3})$$

Here μ and σ are the mean and standard deviation, respectively, of

$$\sum_{i \neq k} f(d_i) + Z ,$$

and G is the standard normal cumulative distribution function. We can solve B3 for $f(\theta)$:

$$f(\theta) = \sigma G^{-1}(p_{\theta R}) + \mu + a \quad (\text{B4})$$

If the range of $p_{\theta R}$ is small, which is to say that the single dot d_k has only a small effect on the observer’s responses, then we can approximate the inverse cumulative normal G^{-1} with the first two terms of a Taylor series. We define p_R as the unconditional probability of the observer responding “right.” Then, Equation B4 becomes

$$\approx \sigma G^{-1}(p_R) + \sigma G^{-1 \prime}(p_R)(p_{\theta R} - p_R) + \mu + a \quad (\text{B5})$$

$$= \sigma G^{-1 \prime}(p_R)p_{\theta R} + \sigma G^{-1 \prime}(p_R)(1 - p_R) + \mu + a . \quad (\text{B6})$$

That is, if we plot $p_{\theta R}$ as a function of the dot direction d_k , we recover an affine transformation of the directional selectivity function, $uf(\theta)+v$.

A single dot displacement has only a small effect on the observer’s responses, so the conditional probability $p_{\theta R}$ varies only slightly as a function of θ (e.g., between 0.49 and 0.51), and even with a large number of trials, the Bernoulli variability in probability estimates makes it difficult to measure $f(\theta)$ accurately. However, we can perform this analysis for each dot displacement d_k , and average the resulting conditional probabilities: an average of functions of the form $uf(\theta)+v$ is itself a function of this form, so we can recover the directional selectivity function $f(\theta)$ equally well from the much less noisy average of all the conditional probabilities.

This method is a special case of Chubb’s (1999) histogram contrast analysis (HCA), which measures the influence of stimulus elements on an observer’s judgments of arbitrary stimulus properties.

Appendix C

In signal detection theory, we often model an observer’s responses as being based on a decision variable Y that is imperfectly correlated with a stimulus property X . It is sometimes useful to know the statistics of X , conditional on Y falling above or below a criterion a .

Theorem. Let $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ be two normal random variables with covariance c_{XY} . Let a be the observer’s criterion, and let z be the normal deviate of a with respect to Y , i.e., $z = (a - \mu_Y) / \sigma_Y$. Then,

$$(a) E[X | Y \geq a] = \mu_X + \frac{g(z)c_{XY}}{G(-z)\sigma_Y} \quad (\text{C1})$$

$$(b) E[X | Y < a] = \mu_X - \frac{g(z)c_{XY}}{G(z)\sigma_Y} . \quad (\text{C2})$$

Here $g(x, \mu, \sigma)$ is the normal probability density function, and $G(x, \mu, \sigma)$ is the normal cumulative distribution function. When we omit μ and σ , they default to zero and one, respectively.

Proof. (a) We can consider Y to be the sum of a term kX that is proportional to X , and a term W that is independent of X , i.e., $Y = kX + W$, where $\text{cov}[X, W] = 0$. Then, $\mu_Y = k\mu_X + \mu_W$, $\sigma_Y^2 = k^2\sigma_X^2 + \sigma_W^2$, and $c_{XY} = k\sigma_X^2$. First, consider the case where $\mu_X = 0$.

$$E[X | Y \geq a] = E[X | kX + W \geq a] \quad (\text{C3})$$

$$= \int_{-\infty}^{\infty} xP(X = x | kX + W \geq a)dx \quad (\text{C4})$$

$$= \int_{-\infty}^{\infty} x \frac{P(kX + W \geq a | X = x)P(X = x)}{P(kX + W \geq a)} dx \quad (\text{C5})$$

$$= \frac{1}{P(kX + W \geq a)} \int_{-\infty}^{\infty} xP(W \geq a - kx)P(X = x)dx \quad (\text{C6})$$

$$= \frac{1}{1 - G(a, \mu_Y, \sigma_Y)} \int_{-\infty}^{\infty} x(1 - G(a - kx, \mu_W, \sigma_W))g(x, 0, \sigma_X)dx \quad (\text{C7})$$

$$= \frac{1}{G(-(a - \mu_Y) / \sigma_Y)} \quad (\text{C8})$$

$$\times \int_{-\infty}^{\infty} xg(x, 0, \sigma_X)G(x, (a - \mu_W) / k, \sigma_W / k)dx$$

Integrating by parts, this becomes

$$= \frac{1}{G(-z)} \left(\left(-\sigma_X^2 g(x, 0, \sigma_X) G\left(x, \frac{a - \mu_Y}{k}, \frac{\sigma_Y}{k}\right) \right) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\sigma_X^2 g(x, 0, \sigma_X) g\left(x, \frac{a - \mu_W}{k}, \frac{\sigma_W}{k}\right) dx \right) \quad (C9)$$

$$= \frac{\sigma_X^2}{G(-z)} \int_{-\infty}^{\infty} g\left(\frac{\mu_W - a}{k}, 0, \sqrt{\sigma_X^2 + \sigma_W^2 / k^2}\right) g(x, \mu^*, \sigma^*) dx \quad (C10)$$

Here we have used the fact that the pointwise product of two normal density functions is a scaled normal density function (specifically,

$$g(x, \mu_1, \sigma_1) g(x, \mu_2, \sigma_2) = g(\mu_1 - \mu_2, 0, \sqrt{\sigma_1^2 + \sigma_2^2}) g(x, \mu, \sigma),$$

where $\mu = (\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2) / (\sigma_1^2 + \sigma_2^2)$ and

$$\sigma^2 = \sigma_1^2 \sigma_2^2 / (\sigma_1^2 + \sigma_2^2),$$

and we have defined

$$\mu^* = \sigma_X^2 (a - \mu_W) / (k^2 \sigma_X^2 + \sigma_W^2) \quad \text{and}$$

$$\sigma^{*2} = \sigma_X^2 \sigma_W^2 / (k^2 \sigma_X^2 + \sigma_W^2). \quad \text{The density function with}$$

parameters μ^* and σ^* integrates to 1, and we are left with

$$= \frac{\sigma_X^2}{G(-z)} g((\mu_W - a) / k, 0, \sqrt{\sigma_X^2 + \sigma_W^2 / k^2}) \quad (C11)$$

$$= \frac{k \sigma_X^2}{G(-z)} g(\mu_W - a, 0, \sqrt{k^2 \sigma_X^2 + \sigma_W^2}) \quad (C12)$$

$$= \frac{c_{XY}}{G(-z)} g(\mu_Y - a, 0, \sigma_Y) \quad (C13)$$

$$= \frac{c_{XY}}{G(-z) \sigma_Y} g(-(a - \mu_Y) / \sigma_Y) \quad (C14)$$

$$= \frac{g(z) c_{XY}}{G(-z) \sigma_Y} \quad (C15)$$

When $\mu_X \neq 0$, we can reduce the problem to the zero-mean case by defining $X' = X - \mu_X$.

$$E[X | Y \geq a] = E[(X' + \mu_X) | Y \geq a] \quad (C16)$$

$$= E[X' | Y \geq a] + \mu_X \quad (C17)$$

This conditional mean can be evaluated using C15.

$$= \mu_X + \frac{g(z) c_{XY}}{G(-z) \sigma_Y} \quad (C18)$$

(b) This problem reduces to case (a):

$$E[X | Y < a] = -E[-X | Y < a] = -E[-X | -Y > -a].$$

According to C18, this evaluates to

$$E[X | Y < a] = \mu_X - \frac{g(z) c_{XY}}{G(z) \sigma_Y} \quad (C19)$$

Acknowledgments

We would like to thank Jason Gold, George Najemnik, and Christopher Taylor for helpful comments on an early draft of this manuscript. We would also like to thank our *Journal of Vision* section editor and two anonymous reviewers for their suggestions. This research was supported by National Science and Engineering Research Council Grants OGP0105494 and OGP0042133. Commercial Relationships: None.

Footnotes

¹We introduced Equation 8 as a constraint, on the grounds that the effect of the modulating function f should not depend on how we conceptually divide a stimulus into independently varying elements.

Alternatively, we can view Equation 8 as claiming that selective attention *uniformly* attenuates the influence of distractors. An example of a nonuniform transformation of likelihoods is a robust statistical calculation, which limits the influence of highly unlikely events (e.g., in a typical robust statistical calculation a single event of probability 10^{-5} has less influence than five events of probability 10^{-1} : $f(10^{-5}) < f(10^{-1})^5$).

Equation 8 states that selective attention does not incorporate a robustness transformation, or a transformation of the opposite type that emphasizes highly unlikely events, but instead uniformly reduces the influence of all distractors. The only functions that satisfy this constraint are the power functions, suggesting that selective attention takes the form of a single weighting factor, as in Equation 12.

²The question of whether selective attention affects the likelihood ratios corresponding to targets, distractors, or both, superficially recalls the question of whether selective attention operates by amplifying attended stimuli, inhibiting unattended stimuli, or both (e.g., James, 1890/1950; Tipper & Driver, 1988). However, any mechanism that affects the influence of a stimulus element on an observer's responses can be seen as adjusting the likelihood ratios of a Bayesian decision-maker, so our account is agnostic as to whether evidence is reweighted by amplification, by inhibition, or by some other mechanism.

³For Edwards and Badock's (1994) cinematograms, and for the cinematograms in our experiments, total horizontal displacement is *not* the ideal decision variable: only the number of dots that move directly to the left or right is informative as to the correct response (see "Methods" section of Experiment 1 for details), whereas a decision variable based on total horizontal displacement allows dots that move at oblique angles to influence the observer's responses. However, as explained earlier, just because we use a Bayesian framework, we need not assume that observers use an

ideal decision rule. Furthermore, channels for perception of global motion are quite broadly tuned (see results of Experiment 1, as well as Williams et al., 1991), so we will use total horizontal displacement as a plausible decision variable. As we have said, though, we make this assumption only to make the exposition more concrete, and we will show that our results do not depend on this assumption.

⁴In principle, observers could partly distinguish between target dots and distractor dots by counting the number of steps each dot took directly to the left or to the right; dots with more steps directly to the left or to the right would be more likely to be target dots, and so their horizontal displacements could be weighted more heavily into the decision variable. However, this seems an implausibly complex strategy, and in any case, our results indicate that observers do weight the distractor dots as heavily as the target dots when they have the same contrast polarity.

⁵Instead of considering these three types of interference as confounds, we could say that observers subject to such interference are simply unable to attend to positive-contrast target dots. However, this would obscure an important distinction: in all three types of interference, the directions of distractors do not influence observers' left-right responses, and so it is only in a very weak sense that such observers could be said to be attending to the distractors. Later ("A More General Model") we discuss the question of how to determine whether the distractors interfere with performance, without actually being correlated with observers' responses.

⁶In this analysis, we have glossed over a problem concerning the relative scale of T^* and D^* . To derive Equations 36 through 39 in Experiment 3, we used not only the fact that T^* and D^* had equal sensitivity d' , but also that T^* and D^* had equal means and standard deviations. This was justified in Experiment 3, because the target and distractor stimuli were identical except for their contrast polarity, and we explicitly assumed that the observer performed the same computation on targets and distractors. In general, though, it is difficult to assign an absolute scale to decision variables, and for convenience, we often assume that a decision variable has standard deviation 1 (e.g., Green & Swets, 1974). For instance, if the color words and ink colors in our Stroop task are equally discriminable, we know that the corresponding decision variables have equal sensitivity $d' = \mu/\sigma$, but it is difficult to rule out the possibility that the color word decision variable has mean μ and standard deviation σ , and that the ink color decision variable has mean 2μ and standard deviation 2σ . For this reason, when describing the Stroop task decision variables, we implicitly used the common modeling assumption that $\sigma=1$. With this assumption, the equal discriminability of targets and distractors implies that T^* and D^* have equal means and standard deviations.

References

- Ahumada, A., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, *49*, 1751-1756.
- Anderson, N. H. (1974). Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2, pp. 215-298). New York: Academic Press.
- Ball, K., & Sekuler, R. (1981). Cues reduce direction uncertainty and enhance motion detection. *Perception and Psychophysics*, *30*, 119-128. [PubMed]
- Barlow, H. B. (1956). Retinal noise and absolute threshold. *Journal of the Optical Society of America A*, *46*, 634-639.
- Barton, J. J., Rizzo, M., Nawrot, M., & Simpson, T. (1996). Optical blur and the perception of global coherent motion in random dot cinematograms. *Vision Research*, *36*, 3051-3059. [PubMed]
- Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception and Psychophysics*, *51*, 145-162. [PubMed]
- Beard, B. L., & Ahumada, A., Jr. (1998). A technique to extract relevant image features for visual tasks. In B. E. Rogowitz & T. N. Pappas (Eds.), *SPIE proceedings: vol. 3299. Human vision and electronic imaging III* (pp. 79-85). Bellingham, WA: SPIE.
- Blaser, E., Sperling, G., & Lu, Z. L. (1999). Measuring the amplification of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 11681-11686. [PubMed]
- Brawn, P., & Snowden, R. J. (1999). Can one pay attention to a particular color? *Perception and Psychophysics*, *61*, 860-873. [PubMed]
- Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America A*, *5*, 617-62. [PubMed]
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, *12*, 199-213. [PubMed]
- Chubb, C. (1999). Texture-based methods for analyzing elementary visual substances. *Journal of Mathematical Psychology*, *43*, 539-567. [PubMed]
- Chubb, C., Econopouly, J., & Landy, M. S. (1994). Histogram contrast analysis and the visual segregation of IID textures. *Journal of the Optical Society of America A*, *11*, 2350-2374. [PubMed]
- Croner, L. J., & Albright, T. D. (1997). Image segmentation enhances discrimination of motion in visual noise. *Vision Research*, *37*, 1415-1427. [PubMed]

- Davis, E. T., & Graham, N. (1981). Spatial frequency uncertainty effects in the detection of sinusoidal gratings. *Vision Research*, *21*, 705-712. [PubMed]
- Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*, 3197-3221. [PubMed]
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501-517. [PubMed]
- Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, *9*, 111-118.
- Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2001). The footsteps of attention in the Posner paradigm revealed by classification images [Abstract]. *Journal of Vision*, *1*, 83a, <http://journalofvision.org/1/3/83/>, DOI 10.1167/1.3.83.[Abstract]
- Edwards, M., & Badcock, D. R. (1994). Global motion perception: Interaction of the ON and OFF pathways. *Vision Research*, *34*, 2849-2858. [PubMed]
- Edwards, M., Badcock, D. R., & Nishida, S. (1996). Contrast sensitivity of the motion system. *Vision Research*, *36*, 2411-2421. [PubMed]
- Egley, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161-177. [PubMed]
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69-73. [PubMed]
- Falmagne, J. -C. (1985). *Elements of psychophysical theory*. Oxford, UK: Oxford University Press.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225-241.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267-314. [PubMed]
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, *402*, 176-178. [PubMed]
- Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, *10*, 663-666.
- Green, D. M. (1958). Detection of multiple component signals in noise. *Journal of the Acoustical Society of America*, *30*, 904-911.
- Green, D. M. (1960). Consistency of auditory detection judgments. *Psychological Review*, *71*, 392-407.
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: R. E. Krieger.
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621-3629. [PubMed]
- James, W. (1890/1950). *The principles of psychology*. New York: Holt; New York: Dover.
- Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, *34*, 2259-2275. [PubMed]
- Kinchla, R. A. (1969). Temporal and channel uncertainty in detection: A multiple observation analysis. *Perception and Psychophysics*, *5*, 129-136.
- Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception and Psychophysics*, *15*, 149-158.
- Kinchla, R. A. (1977). The role of structural redundancy in the perception of visual targets. *Perception and Psychophysics*, *22*, 19-30.
- Kinchla, R. A. (1980). The measurement of attention. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 213-238). Hillsdale, NJ: Erlbaum.
- Kinchla, R. A., Chen Z., Evert, D. (1995). Precue effects in visual search: Data or resource limited? *Perception and Psychophysics*, *57*, 441-450. [PubMed]
- Kinchla, R. A., & Collyer, C. E. (1974). Detecting a target letter in briefly presented arrays: A confidence rating analysis in terms of a weighted additive effects model. *Perception and Psychophysics*, *16*, 117-122.
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, *18*, 2307-2320. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389-412. [PubMed]
- Li, H. -C. O., & Kingdom, F. A. A. (2001). Segregation by color/luminance does not necessarily facilitate motion discrimination in the presence of motion distractors. *Perception and Psychophysics*, *63*, 660-675. [PubMed]
- Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*, 1183-1198. [PubMed]
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

- Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory and Cognition*, *21*, 627-645.
- Murray, R. F., Sekuler, A. B., Bennett, P. J., & Sekuler, R. (1998). Attending to component directions in random dot cinematograms [Abstract]. *Investigative Ophthalmology and Visual Science*, *39*, S1078.
- Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, *401*, 695-698. [PubMed]
- Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, *8*, 2201-2211. [PubMed]
- Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*, 1703-1721. [PubMed]
- Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3-24). Cambridge, UK: Cambridge University Press.
- Pinter, R. B., & Nabet, B. (1992). *Nonlinear vision: Determination of neural receptive fields, function, and networks*. Boca Raton, FL: CRC Press.
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160-174. [PubMed]
- Shaw, M. L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 106-121). Hillsdale, NJ: Erlbaum.
- Shimozaki, S. S., Abbey, C. K., & Eckstein, M. P. (2001). Cue validity effects in the Posner task without enhanced processing or limited resources: An ideal observer analysis [Abstract]. *Investigative Ophthalmology and Visual Science*, *42*, S867.
- Snowden, R. J., & Edmunds, R. (1999). Color and polarity contributions to global motion perception. *Vision Research*, *39*, 1813-1822. [PubMed]
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.
- Tipper, S. P., & Driver, J. (1988). Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli. *Memory and Cognition*, *16*, 64-70.
- van der Smagt, M. J., & van de Grind, W. A. (1999). Integration and segregation of local motion signals: The role of contrast polarity. *Vision Research*, *39*, 811-822. [PubMed]
- Watamaniuk, S. N. (1993). Ideal observer for discrimination of the global direction of dynamic random-dot stimuli. *Journal of the Optical Society of America A*, *10*, 16-28. [PubMed]
- Watamaniuk, S. N., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research*, *29*, 47-59. [PubMed]
- Watson, A. B. (1987). The ideal observer concept as a modelling tool. In *Frontiers in Visual Science: Proceedings of the 1985 symposium* (pp. 32-37). Washington, DC: National Academy Press.
- Williams, D., Tweten, S., & Sekuler, R. (1991). Using metamers to explore motion perception. *Vision Research*, *31*, 275-286. [PubMed]
- Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*, 2685-2696. [PubMed]
- Zohary, E., Scase, M. O., & Braddick, O. J. (1996). Integration across directions in dynamic random dot displays: Vector summation or winner take all? *Vision Research*, *36*, 2321-2331. [PubMed]