

# Does the Superior Colliculus Control Perceptual Sensitivity or Choice Bias during Attention? Evidence from a Multialternative Decision Framework

Devarajan Sridharan,<sup>1</sup> Nicholas A. Steinmetz,<sup>2</sup> Tirin Moore,<sup>2,3</sup> and Eric I. Knudsen<sup>2</sup>

<sup>1</sup>Centre for Neuroscience, Indian Institute of Science, Bengaluru, India, <sup>2</sup>Department of Neurobiology and <sup>3</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305

Distinct networks in the forebrain and the midbrain coordinate to control spatial attention. The critical involvement of the superior colliculus (SC)—the central structure in the midbrain network—in visuospatial attention has been shown by four seminal, published studies in monkeys (*Macaca mulatta*) performing multialternative tasks. However, due to the lack of a mechanistic framework for interpreting behavioral data in such tasks, the nature of the SC's contribution to attention remains unclear. Here we present and validate a novel decision framework for analyzing behavioral data in multialternative attention tasks. We apply this framework to re-examine the behavioral evidence from these published studies. Our model is a multidimensional extension to signal detection theory that distinguishes between two major classes of attentional mechanisms: those that alter the quality of sensory information or “sensitivity,” and those that alter the selective gating of sensory information or “choice bias.” Model-based simulations and model-based analyses of data from these published studies revealed a converging pattern of results that indicated that choice-bias changes, rather than sensitivity changes, were the primary outcome of SC manipulation. Our results suggest that the SC contributes to attentional performance predominantly by generating a spatial choice bias for stimuli at a selected location, and that this bias operates downstream of forebrain mechanisms that enhance sensitivity. The findings lead to a testable mechanistic framework of how the midbrain and forebrain networks interact to control spatial attention.

**Key words:** attention mechanisms; behavior; microstimulation and inactivation; midbrain; multialternative decisions; signal-detection theory

## Significance Statement

Attention involves the selection of the most relevant information for differential sensory processing and decision making. While the mechanisms by which attention alters sensory encoding (sensitivity control) are well studied, the mechanisms by which attention alters decisional weighting of sensory evidence (choice-bias control) are poorly understood. Here, we introduce a model of multialternative decision making that distinguishes bias from sensitivity effects in attention tasks. With our model, we simulate experimental data from four seminal studies that microstimulated or inactivated a key attention-related midbrain structure, the superior colliculus (SC). We demonstrate that the experimental effects of SC manipulation are entirely consistent with the SC controlling attention by changing choice bias, thereby shedding new light on how the brain mediates attention.

## Introduction

Spatial attention is controlled by the coordinated action of two distinct networks in the brain. One of these, the forebrain atten-

tion network, comprises the prefrontal cortex and the posterior parietal cortex (Bisley and Goldberg, 2010; Squire et al., 2013). The contributions of the forebrain network to attention have

Received Oct. 30, 2014; revised July 15, 2016; accepted July 19, 2016.

Author contributions: D.S., T.M., and E.I.K. designed research; D.S. and N.A.S. performed research; D.S. contributed unpublished reagents/analytic tools; D.S. analyzed data; D.S., N.A.S., T.M., and E.I.K. wrote the paper.

This work was supported by a Wellcome Trust Department of Biotechnology–India Alliance Fellowship and a Government of India Department of Biotechnology Ramalingaswami Fellowship (to D.S.), a Stanford Mind, Brain and Computation Integrative Graduate Education and Research Training Fellowship and a National Science Foundation Graduate Research Fellowship (to N.A.S.), National Institutes of Health (NIH) Grant EY014924 (to T.M.) and NIH Grant EY024243 (to E.I.K.). We thank Drs. Richard Krauzlis, Steve Baccus, Surya Ganguli, Roozbeh Kiani, Alireza

Soltani, and Patrick Cavanaugh for comments on preliminary versions of this manuscript. We also thank Drs. Lee Lovejoy, Alexandre Zénon, and Richard Krauzlis for sharing with us data from two published studies of superior colliculus inactivation.

The authors declare no competing financial interests.

Correspondence should be addressed to Devarajan Sridharan, Centre for Neuroscience, Indian Institute of Science, Bangalore, India 560012. E-mail: sridhar@cns.iisc.ernet.in.

DOI:10.1523/JNEUROSCI.4505-14.2017

Copyright © 2017 the authors 0270-6474/17/370480-32\$15.00/0

been studied extensively (Bressler et al., 2008; Corbetta et al., 2008; Corbetta and Shulman, 2011; Noudoost and Moore, 2011; Soltani et al., 2013). The midbrain attention network comprises the superior colliculus (SC) and associated tegmental nuclei (Knudsen, 2011). Much less is known about how the midbrain network mediates its causal role in controlling spatial attention (Krauzlis et al., 2013).

By what mechanisms could the midbrain network contribute to attention control? Signal-detection theory (SDT), a well established Bayesian framework, distinguishes between two major classes of mechanisms by which sensory information is used in making behavioral decisions during attention: (1) those that alter the quality of the sensory information (perceptual sensitivity) and (2) those that alter the selective gating of sensory evidence (choice bias). During spatial attention, brain networks could engage either or both of these classes of mechanisms to differentially enhance perceptual performance at the attended location (Shaw, 1980; Sperling and Doshier, 1986; Eckstein et al., 2009, 2013; Carrasco, 2011).

Four seminal studies have provided converging evidence for a critical role of the SC, the central node in the midbrain network, in controlling spatial attention in primates (Cavanaugh and Wurtz, 2004; Müller et al., 2005; Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012). In each study, manipulations (microstimulation or inactivation) of space-specific loci within the topographic representation in the SC resulted in systematic enhancements or impairments of behavioral performance in diverse attention tasks. However, because of the absence of an adequate mechanistic framework for analyzing and interpreting the behavioral results, none of these studies was able to identify the nature of the SC's contribution to decision making during attention. That is, does the output of the SC improve the processing of sensory information (e.g., increase perceptual sensitivity), does it alter the gating of sensory evidence (e.g., alter spatial choice bias), or both?

Here, we present a multidimensional extension to SDT that enables us to answer this question. We develop a normative framework for distinguishing perceptual sensitivity from choice-bias effects in multialternative decision tasks and demonstrate why conventional behavioral metrics of sensitivity changes are not useful in such tasks. We validate the framework by analyzing avian and primate behavioral data in several multialternative visuospatial tasks, and show that the model not only describes but also accurately predicts behavior in these tasks. We then use this model to re-examine the behavioral evidence from the aforementioned studies that investigated the involvement of the SC in attention (Cavanaugh and Wurtz, 2004; Müller et al., 2005; Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012). We demonstrate that, when examined within the multidimensional decision space of our model, through simulations and model fitting, the reported behavioral effects of SC microstimulation or inactivation can be accounted for by changes in spatial choice bias alone without associated changes in sensitivity. The results provide mechanistic insights into the contribution of the SC to visuospatial selective attention.

## Materials and Methods

### Animals

Experiments were conducted with three adult (>8 months old), female white-leghorn chickens (*Gallus domesticus*) and an adult, male macaque monkey (*Macaca mulatta*). All procedures were in compliance with the guidelines of the National Institutes of Health for the care and use of laboratory animals and were approved by the Institute Animal Care and Use Committee of Stanford University.

## Multialternative detection or change-detection model development

The multidimensional signal-detection model [multialternative detection or change-detection (m-ADC) model] developed here describes behavioral performance in multialternative decision tasks. The mathematical formulation of the extended model is presented below. A verbal description of the model and geometric intuition are provided in the initial section of the Results.

### A model for optimal detection in multialternative tasks

We derive a signal-detection model for m-ADC tasks involving multiple stimulus events and a no-stimulus (“catch”) event. Examples of such tasks are shown in Figure 1A,E. We extend a previous formulation (Sridharan et al., 2014b) to derive an optimal decision rule for m-ADC tasks in which sensitivities for the different stimulus contingencies can be significantly different. Such a formulation is particularly relevant for multialternative attention tasks in which the subject's sensitivity to stimuli presented at a location cued for attention could be substantially higher than that at other locations. In addition, we develop a definition for bias in this signal-detection framework, and demonstrate its analytical relationship with “priors” (prior odds ratio) and “payoffs” (relative values of costs vs benefits).

We develop the model with the example of a multiple ( $m$ )-alternative spatial orientation-change detection task. In this task the subject is required to report the one location, among several, at which an orientation change occurred (Fig. 1E; see Fig. 3A). In each trial, the subject is rewarded for giving a Go response (e.g., saccade) to the location of change, or for giving a NoGo response (e.g., maintain fixation) if no change occurred anywhere. Although we have chosen a specific example, the model is generally applicable to a variety of m-ADC tasks that involve detecting targets (Fig. 1A) or changes in one among several stimulus features (colors, shapes, etc.).

We use the following notation for the stimulus events and response types.  $X$  is a multivariate random variable whose  $i^{\text{th}}$  component,  $X_i$ , denotes the location of the change on each trial:  $X_i = 1$  indicates a change at location  $i$ , and  $X_i = 0$ , otherwise.  $Y$  is a random variable that denotes the type of response on each trial:  $Y = i$  indicates a Go response to location  $i$  and  $Y = 0$  indicates a NoGo response. In each trial of the m-ADC task, a change can occur at one of the many locations ( $\sum_i X_i = 1$ , “change” trial) or not at all ( $\sum_i X_i = 0$ , “catch” trial), and no more than one change event can occur on each trial.

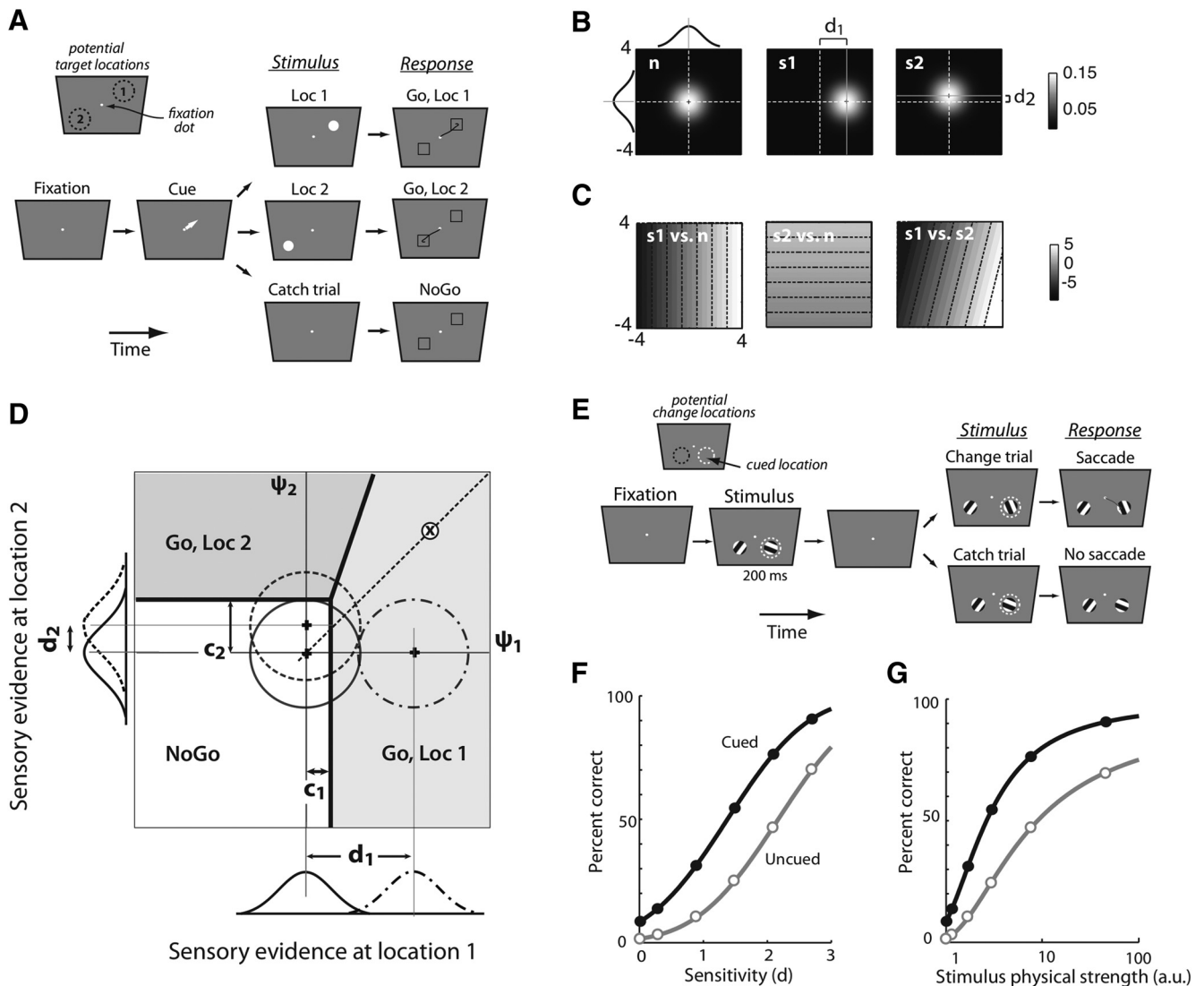
### Model formulation

We developed a latent variable formulation to model behavior in the m-ADC task with unequal sensitivities. The model relates the conditional probability of each type of response for each stimulus event to the perceptual sensitivity ( $d$ ) and choice criterion ( $c$ ) at each location.

We define a multivariate decision variable,  $\Psi$ , whose  $i^{\text{th}}$  component  $\Psi_i$  represents the sensory evidence at location  $i$  ( $i \in \{1 \dots m\}$ ). We posit, as in conventional SDT, that at baseline (no change event), each decision variable component  $\Psi_i$  has a zero mean and unit variance Gaussian (unit normal) distribution, also called the “noise” distribution. The assumption of zero mean for the noise distribution simplifies model development. The model with a nonzero noise mean can be equivalently developed by simply translating the coordinate axes (Sridharan et al., 2014b) and does not alter the results presented here.

A change (target) at location  $i$  ( $X_i = 1$ ) increases the mean of the decision variable distribution at that location by an amount  $d_i$  (i.e.,  $d_i = E(\Psi_i | X_i = 1) - E(\Psi_i | X_i = 0)$ ). We term  $d_i$  an indicator of signal strength, the “perceptual sensitivity” for distinguishing a signal (target or change) from noise (no target or change) at location  $i$ . In general,  $d_i$  is determined by the physical strength of the stimulus event (e.g., contrast or orientation change magnitude) and can be modulated by attention.

In addition, we specify that the components  $\Psi_i$  are independently distributed, so that the covariance matrix of  $\Psi$ , which we denote as  $C$ , is an identity (unit) matrix as follows:  $C = I; C_{ii} = 1, C_{ij} = 0, i, j \in \{1, \dots, m\}, i \neq j$ . This implies that across trials, the noisy fluctuations in the decision variable's value at a location are uncorrelated with the fluctuations in its values at all other locations.



**Figure 1.** Multidimensional signal-detection (m-ADC) model to dissociate bias from sensitivity effects in multialternative attention tasks. **A**, Schematic of a spatial two-alternative, cued detection task. The subject initiates a trial by fixating on a dot at the center of the screen. A cue (white line segment) indicates the likely location of subsequent target stimulus. A target (positive contrast dot) is briefly presented at one of two potential target locations. The subject is rewarded for indicating the location of the target with a saccade (line) to the corresponding response box (squares). On catch trials, no target is presented for a prolonged period following fixation. On these trials, the subject is rewarded for maintaining fixation on the zeroing dot even after the response boxes appear. **B**, Distributions of the bivariate decision variable ( $\Psi$ ) in a two-dimensional decision space for no-stimulus (catch; n), stimulus at location 1 (s1), and stimulus at location 2 (s2). Dashed white lines indicate orthogonal perceptual dimensions for locations 1 and 2, respectively. Units: SD of the decision variable distribution. Black curves, One-dimensional Gaussians, marginal distributions of each decision variable component. The s1 and s2 distributions correspond to the n distribution translated along the respective perceptual dimension by an amount equal to the perceptual sensitivity for detecting the stimulus at that location ( $d_1$  or  $d_2$ , respectively). Perceptual sensitivity at location 1 is illustrated as greater than at location 2 ( $d_1 > d_2$ ), as could result from cueing location 1 for attention. **C**, Log posterior odds ratio ( $\Lambda$ ) and the family of optimal decision surfaces (hyperplanes) of constant  $\Lambda$  (dashed lines) for distinguishing a stimulus at location 1 versus noise (s1 vs n), a stimulus at location 2 versus noise (s2 vs n); and a stimulus at location 1 versus at location 2 (s1 vs s2). **D**, Decision variable ( $\Psi$ ) distributions (circles) and putative optimal decision surfaces (thick black lines) for the 2-ADC model with unequal perceptual sensitivities ( $d_1 \neq d_2$ ) and choice criteria ( $c_1 \neq c_2$ ) at the two locations. Solid circle, Contour of the noise distribution; dot-dashed and dashed circles, contours of the signal distribution for a stimulus at location 1 and 2, respectively. The decision boundaries partition the decision space into nonoverlapping domains for each potential response: NoGo (unshaded region), Go response to location 1 (light shading), or Go response to location 2 (dark shading). One-dimensional Gaussians along each axis: marginal distributions of each decision variable component. Dashed diagonal line with X symbol: line of equal sensory evidence ( $\Psi_1 = \Psi_2$ ). **E**, A cued two-alternative orientation-change detection paradigm. Following fixation, the subject is presented with two oriented gratings that flashed on and off for variable durations. At a random time, the orientation of one of the gratings changes, and the subject is rewarded for making a saccadic eye movement to the location of change (change trials) or for maintaining fixation on trials when no change occurred (catch trials). Attention is cued to a location across a block of trials by increasing the probability of change occurrence at that location (dashed white circle). The task is conceptually similar to one in which an explicit cue is presented on each trial (**A**). **F**, Simulated percentage correct as a function of sensitivity ( $d$ ) for detecting orientation changes at cued (closed circles) and uncued (open circles) locations. **G**, Simulated psychometric functions, percentage correct as a function of stimulus physical strength (e.g., orientation change magnitude or contrast) at cued (closed circles) and uncued (open circles) locations.

Thus, the decision variable is distributed as follows (Eq. 1):

$$\Psi(\mathbf{X}) \sim \mathcal{N}_m(\mu = \mathbf{d} \odot \mathbf{X}, \mathbf{C} = \mathbf{I}) = A e^{-\sum_{k=1}^m \frac{1}{2} (\Psi_k - d_k X_k)^2}$$

where  $\mathcal{N}_m$  is the multidimensional Gaussian density function,  $d$  is a vector whose  $i^{th}$  component is  $d_i$  (perceptual sensitivity at loca-

tion  $i$ ),  $\odot$  denotes the element-wise multiplication operation, and  $A$  is a normalization constant in order for  $\mathcal{N}_m$  to be a probability density ( $A = 1/(2\pi)^{m/2}$ ).

Because of the independence of the  $\Psi_i$ , Equation 1 can be equivalently written in terms of the marginals (DeCarlo, 2012), as follows (Eq. 2):

$$\Psi_i = d_i X_i + \varepsilon_i; \quad \varepsilon_i \sim \mathcal{N}(0, 1)$$

where  $\varepsilon_i$  represents the noise distribution (unit normal) at location  $i$  and  $\mathcal{N}$  the one-dimensional Gaussian density function.

In our model, the components of  $d$  are not necessarily equal ( $d_i \not\equiv d_j, \forall i \neq j$ ) so that the sensitivities to detect a change can be different across locations. Figure 1B,D illustrates the decision variable  $\Psi$  distribution for a two-alternative detection (2-ADC) model.

### Optimal decision surfaces

In any behavioral task, it is reasonable to suppose that the subject seeks to optimize some quantity of interest. Here, we derive the family of optimal decision surfaces for the m-ADC model that maximize average utility or minimize average risk; utility (or risk) is defined as the benefit (or cost) associated with choosing a particular response when a particular stimulus occurs, and is assumed to be uniquely specified for each stimulus–response contingency. Here, we seek to identify the family of optimal decision surfaces for the general case of unequal  $d_i$ .

In deriving these optimal decision surfaces, we assume that the cost of making an error for each change event is the same regardless of the response type. For example, when a change happens at some location  $k$ , we specify that the cost of making a Go response to any location other than location  $k$  or a NoGo response are all the same. This is a reasonable assumption for common multialternative tasks in which the typical penalty for an error is withheld reward (same cost) regardless of the specific error. Notice, though, that we allow for these costs to be different across the different change contingencies. For example, an erroneous response when a change occurred at one location can be more or less costly (e.g., more or less severely penalized) than the same erroneous response when a change occurred at another location, or during catch trials.

We define the following relations (to be used in our derivations) as follows (Eqs. 3, 4):

$$\mathcal{L}_{i0}(\Psi) = \frac{\mathcal{N}_m(\Psi|X_i=1)}{\mathcal{N}_m(\Psi|X_k=0 \forall k)}$$

$$\Lambda_{i0}(\Psi) = \frac{p_i}{p_0} \mathcal{L}_{i0}$$

where  $\mathcal{L}_{i0}(\Psi)$  is the likelihood ratio corresponding to a change at location  $i$  relative to no change, and  $\Lambda_{i0}(\Psi)$  is the posterior odds ratio given by scaling the likelihood ratio by the prior odds ratio of a change at location  $i$  relative to no change:  $p_i/p_0 = p(X_i = 1)/p(X_k = 0 \forall k)$ .

Optimal decision surfaces in the m-ADC model are hyperplanes of the constant log posterior odds ratio as follows (Eqs. 5 and 6):  $\Lambda_{kl}(\Psi) = \beta_{kl}$  and  $\Lambda_{k0}(\Psi) = \beta_{k0}$ , where  $\Lambda_{kl}(\Psi) = \frac{\Lambda_{k0}(\Psi)}{\Lambda_{l0}(\Psi)} = \frac{p_k}{p_l} \frac{\mathcal{N}_m(\Psi|X_k=1)}{\mathcal{N}_m(\Psi|X_l=1)}$  is the posterior odds ratio of a change at location  $k$  relative to a change at location  $l$ , given the sensory evidence ( $\Psi$ ) and  $\beta_{kl} = (C_l^k - C_k^k)/(C_l^k - C_k^k)$ , where  $C_l^k$  represents the cost of responding to location  $k$  when a change occurred at location  $l$  ( $k, l \in \{0, 1, 2, \dots, m\}$ ),  $C_k^0$  represents the cost of giving a NoGo response, when a change occurred at location  $k$  and  $C_0^k$  represents the cost of responding to location  $k$ , when no change occurred (catch trial). This result has been demonstrated previously in Sridharan et al. (2014b) for the case of equal sensitivities, and holds true here in the case of unequal sensitivities as well.

We next identify the exact form of these optimal decision surfaces for the m-ADC model in which sensitivity values can be widely different at the two locations (Fig. 1B,D).

In an m-ADC task, the decision must be made among  $m$  locations of change, and one no-change (or noise) contingency. Because the distributions of the decision variables are Gaussian, calculating the logarithm of the posterior odds ratio provides simpler mathematical expressions; because of the monotonicity of the log function, isosurfaces of the posterior odds ratio are also isosurfaces of the log posterior odds ratio.

We first calculate the log-likelihood ratio for a change at location  $j$  versus no-change (noise). From Equations 3 and 1 this is given by the following (Eqs. 7 and 8):

$$\begin{aligned} \log \mathcal{L}_{j0} &= \log \frac{Ae^{-\frac{1}{2}(\Psi_j - d_j)^2} - \sum_{k=1, k \neq j}^m \frac{1}{2} \Psi_k^2}{Ae^{-\sum_{k=1}^m \frac{1}{2} \Psi_k^2}} \\ &= \Psi_j d_j - d_j^2 / 2 \end{aligned}$$

where  $d_j$  represents the perceptual sensitivity for a change at location  $j$ .

The log posterior odds ratio is given as follows (from Eq. 4):

$$\log \Lambda_{j0} = \log \frac{p_j}{p_0} + \Psi_j d_j - \frac{d_j^2}{2}$$

Optimal decision surfaces for distinguishing a change at location  $j$  from no change (isosurfaces of  $\log \Lambda_{j0}$ ) are of the following form (Eqs. 10 and 11; from Eq. 6):

$$\log \frac{p_j}{p_0} + \Psi_j d_j - \frac{d_j^2}{2} = \log \beta_{j0}$$

$$\Psi_j = \frac{1}{d_j} \left( \log \beta_{j0} - \log \frac{p_j}{p_0} + \frac{d_j^2}{2} \right)$$

$$\Psi_j = f(\beta_{j0}, p_j/p_0, d_j)$$

Thus, these optimal surfaces are functions of the priors, costs (or payoffs), and sensitivity. For a given set of these parameters, we term the value of this function as the choice criterion,  $c_j$ , so that optimal decision surfaces are hyperplanes of the form  $\Psi_j = c_j$ . The subject decides between no-change and a change at location  $j$  based on whether the decision variable falls above or below the cutoff value specified by the choice criterion at location  $j$  ( $\Psi_j \geq c_j$ ).

Next, we calculate the log-likelihood ratio for a change at location  $i$  versus a change at location  $j$ . This is given by the following (Eq. 12):

$$\begin{aligned} \log \mathcal{L}_{ij} &= \log \frac{\Psi(X_i|X_j=1)}{\Psi(X_j|X_i=1)} \\ &= \log \frac{Ae^{-\frac{1}{2}(\Psi_i - d_i)^2 - \sum_{k=1, k \neq i, j}^m \frac{1}{2} \Psi_k^2}}{Ae^{-\frac{1}{2}(\Psi_j - d_j)^2 - \sum_{k=1, k \neq i, j}^m \frac{1}{2} \Psi_k^2}} \\ &= \frac{1}{2} (2\Psi_i d_i - d_i^2 - 2\Psi_j d_j + d_j^2) \end{aligned}$$

where  $d_i$  and  $d_j$  represent sensitivities to changes at locations  $i$  and  $j$ , respectively.

The log posterior odds ratio is given by the following (Eq. 13):

$$\log \Lambda_{ij} = \Psi_i d_i - \frac{d_i^2}{2} - \Psi_j d_j + \frac{d_j^2}{2} + \log \frac{p_i}{p_j}$$

Optimal decision surfaces for distinguishing a change at location  $i$  from a change at location  $j$  (isosurfaces of  $\log \Lambda_{ij}$ ) are of the following form (Eqs. 14–16):

$$\Psi_i d_i - \frac{d_i^2}{2} - \Psi_j d_j + \frac{d_j^2}{2} + \log \frac{p_i}{p_j} = \log \beta_{ij}$$

$$\Psi_i d_i - \Psi_j d_j = \log \beta_{ij} - \log \frac{p_i}{p_j} + \frac{d_i^2 - d_j^2}{2}$$

$$\Psi_i d_i - \Psi_j d_j = f(\beta_{ij}, p_i/p_j, d_i, d_j)$$

Thus, as before, the optimal decision surfaces are a function of the priors, costs (or payoffs), and sensitivities. We note the value of right-hand



side as  $B_{ij}$  so that the optimal decision surfaces are hyperplanes of the form  $\Psi_i d_i - \Psi_j d_j = B_{ij}$ .

Figure 1C illustrates this family of optimal decision surfaces for the 2-ADC case.

To determine the value of the constant  $B_{ij}$ , we employ the following result: optimal decision surfaces defined by Equations 11 and 16 intersect at a point (proved formally in Sridharan et al., 2014b). It is apparent that if the decision surfaces do not intersect at a point (Fig. 1D, thick black lines), then the decision space could contain domains where the optimal decision is not uniquely specified.

Given this, each of the decision surfaces defined by Equation 16 must pass through the point of intersection of the optimal decision surfaces defined in Equation 11, given by  $(\Psi_i, \Psi_j) = (c_j, c_i)$ . Hence, the constant  $B_{ij} = c_i d_i - c_j d_j$  and the optimal decision hyperplane are given by  $\Psi_i d_i - \Psi_j d_j = c_i d_i - c_j d_j$  or  $d_i(\Psi_i - c_i) = d_j(\Psi_j - c_j)$ .

Thus, optimal decision surfaces for the m-ADC model are given by the following pair of equations (Eqs. 17 and 18):  $\Psi_k = c_k \forall k$  and  $d_i(\Psi_i - c_i) = d_j(\Psi_j - c_j) \forall i, j \ i \neq j$ .

### The m-ADC decision rule

Based on the optimal decision surfaces derived above, the decision rule for the m-ADC model may be defined as follows: the subject decides between a change at location  $i$  versus no change based on whether the component of the decision variable corresponding to that location  $\Psi_i$  fell above or below the choice criterion  $c_i$ . In addition, the subject decides between a change at location  $i$  and a change at location  $j$  based on the relative magnitudes of  $d_i(\Psi_i - c_i) \cong d_j(\Psi_j - c_j)$ .

We formulate this decision rule analytically as follows (Eq. 19):

$$Y = i \text{ if } \Psi_i > c_i \cap \bigcap_{k=1, k \neq i}^m d_i(\Psi_i - c_i) > d_k(\Psi_k - c_k)$$

$$Y = 0 \text{ if } \bigcap_{k=1}^m \Psi_k \leq c_k$$

where  $\cap$  represents the intersection of the various constraints. According to this rule, the subject's ultimate choice of response location (say,  $r$ ) is the one at which the decision variable component exceeds the criterion ( $\Psi_r > c_r$ ) and the one at which the magnitude of  $d_r(\Psi_r - c_r)$  is largest (among all locations). If no decision variable component exceeds criterion at any location ( $\Psi_k \leq c_k \forall k$ ), then the subject gives a NoGo response ( $Y = 0$ ).

Figure 1D illustrates this decision rule for the 2-ADC case.

Note that when  $d_i = d_j$ , this decision rule reduces to that for an m-ADC model with equal sensitivities (Sridharan et al., 2014b), in which the decision rule is independent of the sensitivity at each location. Thus, the key difference in this case of unequal sensitivities is that the decision rule also incorporates the relative values of sensitivity at each pair of locations. Using a 2-ADC model, the difference may be characterized geometrically as follows: for the 2-ADC case with equal sensitivities, the oblique decision boundary in Figure 1D is a line tilted at 45° relative to each decision axis, whereas for the 2-ADC case with unequal sensitivities, the oblique decision boundary is a line whose tilt depends on the relative values of the sensitivities at each location.

### Relating response probabilities to sensitivities and criteria

The probability of a particular response for a stimulus event can be expressed as the product of the prior probability of a stimulus event with the conditional probability of that response for that stimulus event. The prior probability of each stimulus event is governed by the task specification. Here, we derive the conditional probability of a response to location  $i$  for each stimulus event  $X$ , based on the decision rule as follows (Eq. 20):

$$p(Y = i | \mathbf{X}) = p(\Psi_i > c_i \cap \bigcap_{k, k \neq i} [d_i(\Psi_i - c_i) > d_k(\Psi_k - c_k)])$$

Substituting for  $\Psi_p$ ,  $\Psi_k$  based on the latent variable model from (Eq. 2) yields (Eq. 21):

$$p(Y = i | \mathbf{X}) = p(\varepsilon_i > c_i - d_i X_i \cap \bigcap_{k, k \neq i} [d_i(\varepsilon_i + d_i X_i - c_i) > d_k(\varepsilon_k + d_k X_k - c_k)])$$

Conditioning the above probability on a given value of  $\varepsilon_i = e_i$  yields the following (Eq. 22):

$$p(Y = i | \mathbf{X}, e) = p(e_i > c_i - d_i X_i \cap \bigcap_{k, k \neq i} [d_i(e_i + d_i X_i - c_i) > d_k(\varepsilon_k + d_k X_k - c_k)])$$

$$= \mathcal{H}(e_i - c_i + d_i X_i) \cap \bigcap_{k, k \neq i} p(\varepsilon_k < \frac{d_i}{d_k}(e_i + d_i X_i - c_i) - d_k X_k + c_k)$$

$$= \mathcal{H}(e_i - c_i + d_i X_i) \prod_{k, k \neq i} \mathcal{F}\left(\frac{d_i}{d_k}(e_i + d_i X_i - c_i) - d_k X_k + c_k\right)$$

where  $\mathcal{H}$  is the Heaviside function,  $\mathcal{F}$  is the cumulative distribution function of the unit normal distribution  $\mathcal{N}$ , and we have factored the joint probability into the product of the marginal probabilities, as the  $\varepsilon_i$ 's are independently distributed.

The conditional probability of a response at location  $i$  is then expressed as follows (Eq. 23):

$$p(Y = i | \mathbf{X}) = \int_{c_i - d_i X_i}^{\infty} \prod_{k, k \neq i} \mathcal{F}\left(\frac{d_i}{d_k}(e_i + d_i X_i - c_i) - d_k X_k + c_k\right) \mathcal{N}(e_i) de_i$$

This conditional probability represents an integration over the noise distribution ( $e_i$ ).

The probability of a NoGo response,  $p(Y = 0 | X)$ , for each stimulus event,  $X$ , can be obtained from the probabilities of the other responses as follows:  $p(Y = 0 | X) = 1 - \sum_{i=1}^m p(Y = i | X)$ . Formally, this can be shown to be equal to the following (Eq. 24):

$$p(Y = 0 | \mathbf{X}) = p(\bigcap_k \varepsilon_k \leq c_k - d_k X_k)$$

$$= \prod_k p(\varepsilon_k \leq c_k - d_k X_k)$$

$$= \prod_k \mathcal{F}(c_k - d_k X_k)$$

where, as before, we have factored the joint probability of the independent  $\varepsilon_k$ 's into the product of the marginal probabilities.

For the 2-ADC case, the NoGo response rate during catch trials (i.e., the correct rejection rate) is given by  $p(Y = 0 | X_1 = 0, X_2 = 0) = \mathcal{F}(c_1) \mathcal{F}(c_2)$ . The false-alarm rate during catch trials is, hence,  $p_{fa} = 1 - \mathcal{F}(c_1) \mathcal{F}(c_2)$ . Thus, given the false-alarm rate during catch trials, and one of the criteria (say  $c_1$ ), the other criterion ( $c_2$ ) is determined by the following relation (Eq. 25):

$$c_2 = \mathcal{F}^{-1}\left(\frac{1 - p_{fa}}{\mathcal{F}(c_1)}\right)$$

### Relationship between choice criterion, choice bias, priors, and payoffs

Based on the results derived above, we formalize the relationship between the choice criterion, choice bias, priors, and payoffs in the m-ADC model.

We reproduce Equation 11 here as follows (Eq. 26):

$$c_j = \frac{1}{d_j} \left( \log \beta_{j0} - \log \frac{p_j}{p_0} + \frac{d_j^2}{2} \right)$$

$$c_j = \frac{1}{d_j} \log \left( \frac{\beta_{j0}}{p_j/p_0} \right) + \frac{d_j}{2}$$

Let us examine each term in turn.

First, the optimal choice criterion at each location  $j$ ,  $c_j$ , depends on the logarithm of the relative cost or benefit of choosing each alternative ( $\beta_{j0}$ ): the greater the benefit (or the smaller the cost) of a correct Go response to location  $j$  ("hit"), relative to a NoGo response ("miss";  $C_j^1 - C_j^0$ ), the lower the value of  $\beta_{j0}$ , and the lower the optimal  $c_j$ . This defines the analytical relationship between the "payoff" at location  $j$  and the optimal

choice criterion to report changes at that location: the greater the payoff, the lower the optimal choice criterion value.

Second, the optimal choice criterion  $c_j$  depends on the logarithm of the prior odds ratio ( $p_j/p_0$ ): the greater the value of the prior probability for a change event at location  $j$  (the “prior”), the larger the prior odds ratio, and the smaller the value of the optimal  $c_j$ . This defines the analytical relationship between the “prior” at location  $j$  and the optimal choice criterion to report changes at that location: the greater the prior, the lower the optimal choice criterion value.

Finally, the optimal choice criterion is also a (nonlinear) function of the perceptual sensitivity to the change event at location  $j$ ,  $d_j$ .

When sensitivities to events at different locations are equal, the relative values of the choice criteria suffice to quantify choice bias. A lower value of criterion at a location (relative to another location) represents a tendency to afford greater weight to the sensory evidence at that location and a greater choice bias toward that location (Sridharan et al., 2014b). On the other hand, when sensitivities to different locations are different, as in attention tasks, this definition of bias no longer suffices, because this definition of choice bias is a function of sensitivity.

We define choice bias for this more general scenario. First, we note that in the absence of differences in prior probabilities ( $p_j = p_0$ ) and relative costs ( $\beta_{j0} = 1$ ), the optimal criterion is given by  $c_j = d_j/2$  (from Eq. 26; i.e., the optimal criterion is located at the midpoint of the signal and noise distributions). We refer to this criterion as  $c_j^*$ . The difference  $c_j - c_j^*$  represents the “constant criterion” measure of bias, analogous to the measure defined in Macmillan and Creelman (2005) for the 2-AFC task.

Here, we define choice bias, based on the likelihood ratio measure, as follows (Eq. 27):

$$\begin{aligned} \log b_j &= -\log \mathcal{L}_{j0}(\Psi_j = c_j) \\ &= -d_j(c_j - c_j^*) \\ &= \log \left( \frac{p_j/p_0}{\beta_{j0}} \right) \end{aligned}$$

The last relationship is obtained by rearranging Equation 26. Thus, for optimal decisions, choice bias to give a Go response to location  $j$  (vs a NoGo response) in the m-ADC model is given by the following (Eq. 28):

$$\begin{aligned} b_j &= e^{-d_j(c_j - c_j^*)} \\ &= \frac{p_j/p_0}{\beta_{j0}} \end{aligned}$$

$b_j$  is analytically related to the prior odds ratio and to the relative payoffs (for the 2-AFC model, compare with Macmillan and Creelman, 2005, their Eq. 2.11). Even when targets of different strengths are sampled at a location (e.g., as when measuring the psychometric function), as long as the priors and payoffs do not change with target strength, the optimal bias measure remains invariant to sensitivity values. The relative values of  $b$  represent the relative biases to give a Go response to different locations.

For suboptimal decisions, other sources of bias can be captured within this framework, with a multiplicative term, as follows (Eq. 29):

$$b_j = \left( \frac{p_j/p_0}{\beta_{j0}} \right) b_j^s$$

where  $b_j^s$  represents a component of bias that leads to suboptimal decisions. This could include contributions from motoric response biases, irrational biases, or biases that originate from sources other than priors and payoffs. Because the model estimates  $d_j$  and  $c_j$ , choice bias  $b_j$  can be quantified with Equation 28. If the values of the priors ( $p_j/p_0$ ) and payoffs ( $\beta_{j0}$ ) are precisely known, the value of the suboptimal bias,  $b_j^s$ , can be quantified.

*Optimal two-alternative detection model for 100% valid cues is a two-alternative forced-choice (Yes-No) model.* We first consider the two-alternative detection (2-ADC) model where the change can happen at one of two locations, or not at all. For this model, response probabilities

are related to the sensitivities and criteria by the following set of equations (Eqs. 30 and 31):

$$\begin{aligned} p(Y = 1|\mathbf{X}) &= \int_{c_1 - d_1 X_1}^{\infty} \mathcal{F} \left( \frac{d_1}{d_2} (e_1 + d_1 X_1 - c_1) - d_2 X_2 + c_2 \right) \mathcal{N}(e_1) de_1 \\ p(Y = 2|\mathbf{X}) &= \int_{c_2 - d_2 X_2}^{\infty} \mathcal{F} \left( \frac{d_2}{d_1} (e_2 + d_2 X_2 - c_2) - d_1 X_1 + c_1 \right) \mathcal{N}(e_2) de_2 \end{aligned}$$

These are the same as Equation 23 but adapted for the two-alternative case.

The optimal choice criterion for distinguishing a change at location 2 from no change (Eq. 11) is as follows (Eq. 32):

$$\Psi_2^* = c_2 = \frac{1}{d_2} \left( \log \beta_{20} - \log \frac{p_2}{p_0} + \frac{d_2^2}{2} \right)$$

Now, let one of the two locations, say 1, be cued with a 100% valid cue so that changes can happen at this location or not at all; these two events are mutually exclusive and exhaustive. Thus, the prior probability of a change at location 2,  $p_2 = 0$ .

In this case, from Equation 32 the optimal criterion at location 2 tends to infinity (i.e.,  $\lim_{p_2 \rightarrow 0} c_2 \rightarrow +\infty$ ). Substituting this limit into Equations 30 and 31 results in the following (Eqs. 33 and 34):

$$\begin{aligned} p(Y = 1|\mathbf{X}) &= \int_{c_1 - d_1 X_1}^{\infty} \lim_{s \rightarrow +\infty} \mathcal{F}(s) \mathcal{N}(e_1) de_1 = \mathcal{F}(d_1 X_1 - c_1) \\ p(Y = 2|\mathbf{X}) &= \int_{-\infty}^{\infty} \lim_{s \rightarrow -\infty} \mathcal{F}(s) \mathcal{N}(e_2) de_2 = 0 \end{aligned}$$

where we have assumed that the other three parameters,  $d_1$ ,  $d_2$  and  $c_1$ , are finite, and have employed the following identities:  $\lim_{s \rightarrow +\infty} \mathcal{F}(s) = 1$ ;  $\lim_{s \rightarrow -\infty} \mathcal{F}(s) = 0$  and  $\int_q^{\infty} \mathcal{N}(e) de = \mathcal{F}(-q)$ .

These equations highlight three key results for the optimal 2-ADC model with a 100% valid cue: (1) the probability of a response to location 2 is vanishingly small (rarely or never occurs), (2) the hit rate at location 1,  $h_1$ , corresponding to the probability of a response to location 1 when a change occurred at location 1,  $p(Y = 1|X_1 = 1, X_2 = 0)$  is given by  $\mathcal{F}(d_1 - c_1)$ , and (3) the false-alarm rate at location 1,  $f_1$ , corresponding to the probability of a response to location 1 when no change occurred,  $p(Y = 1|X_1 = 0, X_2 = 0)$  is given by  $\mathcal{F}(-c_1)$ .

Rearranging 2 and 3, we get the following (Eq. 35):

$$d_1 = \mathcal{F}^{-1}(h_1) - \mathcal{F}^{-1}(f_1) \quad c_1 = \mathcal{F}^{-1}(f_1)$$

These equations are identical to the two-alternative forced-choice (2-AFC) Yes/No model (Macmillan and Creelman, 2005) in which a change can occur at location 1 or not at all. Thus, the optimal two-alternative detection model reduces to a 2-AFC (Yes/No) model when cues are 100% valid.

### Extension of the model to perceptual discrimination tasks

The m-ADC model described above was formulated for detection and change-detection tasks. The model can be readily extended to perceptual discrimination tasks that, optionally, incorporate a NoGo response (see Fig. 7A). In this task the subject must discriminate between one of several features (e.g., direction of dot motion) of the target stimulus. We describe the model for this two-alternative discrimination task (see Fig. 7A, distinguishing between leftward vs rightward motion) in which the target can occur at one of two locations.

For this model, the decision variable distribution is specified as follows (Eq. 36):

$$\Psi_i = d_i X_i s_i + \varepsilon_i; \quad \varepsilon_i \sim \mathcal{N}(0, 1) \quad i \in \{1, 2\}$$

Note that this is the same as the decision variable distribution for the m-ADC model from Equation 2 with the exception of the multiplicative term  $s_i$ . We specify that  $s_i$  takes the value 1 when the stimulus is moving in a rightward direction at location  $i$  and takes a value  $-1$  when the stimulus is moving leftward at location  $i$ . In general, the convention for the sign of  $s_i$  at each location  $i$  can be arbitrary, as long as signal strength in one direction is represented along the positive axis and in the other direction is represented along the negative axis of the decision variable for that location ( $\Psi_i$ ). With this specification, the sensitivity to discriminate between leftward and rightward motion at location  $i$  is given by  $2d_i$ .

We hypothesize the subject employs choice criteria  $c_1$  and  $-c_1$  at location 1, and  $c_2$  and  $-c_2$  at location 2 when making a decision about the perceived direction of motion, as shown in Figure 7B,C. The decision rule is formulated as follows (Eq. 37):

$$Y = R \text{ if } (\Psi_1 > c_1 \cap d_1(\Psi_1 - c_1) > d_2(\Psi_2 - c_2)) \cup (\Psi_1 < -c_1 \cap \Psi_2 < -c_2 \cap d_1(\Psi_1 + c_1) > d_2(\Psi_2 + c_2))$$

$$Y = L \text{ if } (\Psi_2 > c_2 \cap d_1(\Psi_1 - c_1) < d_2(\Psi_2 - c_2)) \cup (\Psi_1 < -c_1 \cap \Psi_2 < c_2 \cap d_1(\Psi_1 + c_1) < d_2(\Psi_2 + c_2))$$

$$Y = N \text{ if } -c_1 \leq \Psi_1 \leq c_1 \cap -c_2 \leq \Psi_2 \leq c_2$$

where  $Y = R$ ,  $Y = L$ , and  $Y = N$  indicate rightward motion, leftward motion, and NoGo responses, respectively. The decision is based on a family of six decision boundaries. Four of these boundaries represent lines parallel to the decision axes corresponding to  $\Psi_1 = c_1$ ,  $\Psi_1 = -c_1$ ,  $\Psi_2 = c_2$ , and  $\Psi_2 = -c_2$ . The angle of the oblique decision boundaries are determined by the relative values of  $d_1$  and  $d_2$  as with the m-ADC model decision rule (Eq. 19), and their equations are given by  $d_1(\Psi_1 - c_1) = d_2(\Psi_2 - c_2)$  and  $d_1(\Psi_1 + c_1) = d_2(\Psi_2 + c_2)$ .

This decision rule partitions the decision space into three domains: the first corresponding to a Go response indicating rightward motion (see Fig. 7B,C, lower right), the second corresponding to a Go response indicating leftward motion (see Fig. 7B,C, upper left), and the third corresponding to a NoGo response (see Fig. 7B,C, gray-shaded region). On each trial, the subject indicates rightward or leftward motion or gives a NoGo response, depending on the region of the decision space into which the decision variable falls on that trial. Thus, the probability of each type of response is given by the following (Eq. 38):

$$p(Y = L|\mathbf{X}) = \int_{-\infty}^{-c_1} \left( \int_{-\infty}^{\frac{d_1}{d_2}(\Psi_1 + c_1) - c_2} \Psi(\mathbf{X}; d_1, d_2) d\Psi_2 \right) d\Psi_1$$

$$+ \int_{-c_1}^{c_1} \left( \int_{-\infty}^{-c_2} \Psi(\mathbf{X}; d_1, d_2) d\Psi_2 \right) d\Psi_1$$

$$+ \int_{c_1}^{\infty} \left( \int_{-c_2}^{\frac{d_1}{d_2}(\Psi_1 - c_1) + c_2} \Psi(\mathbf{X}; d_1, d_2) d\Psi_2 \right) d\Psi_1$$

$$p(Y = R|\mathbf{X}) = \int_{-\infty}^{-c_1} \left( \int_{\frac{d_1}{d_2}(\Psi_1 + c_1) - c_2}^{\infty} \Psi(\mathbf{X}; d_1, d_2) d\Psi_2 \right) d\Psi_1$$

$$+ \int_{-c_1}^{c_1} \left( \int_{c_2}^{\infty} \Psi(\mathbf{X}; d_1, d_2) d\Psi_2 \right) d\Psi_1$$

$$+ \int_{c_1}^{\infty} \left( \int_{\frac{d_1}{d_2}(\Psi_1 - c_1) + c_2}^{\infty} \Psi(\mathbf{X}; d_1, d_2) d\Psi_2 \right) d\Psi_1$$

$$p(Y = N|\mathbf{X}) = \int_{-c_1}^{c_1} \int_{-c_2}^{c_2} \Psi(\mathbf{X}; d_1, d_2) d\Psi_2 d\Psi_1$$

where  $p(Y = L|\mathbf{X})$  represents the conditional probability of indicating leftward motion ( $Go, L$ ),  $p(Y = R|\mathbf{X})$  represents the conditional probability of indicating rightward motion ( $Go, R$ ),  $p(Y = N|\mathbf{X})$  represents the conditional probability of a NoGo response and, as before,  $\mathbf{X}$  is a multivariate random variable whose  $i^{\text{th}}$  component,  $X_i$ , denotes the location of the stimulus on each trial:  $X_i = 1$  indicates a stimulus at location  $i$ , and  $X_i = 0$ , otherwise.

In tasks in which the subject is not permitted to make a NoGo response, and if the decision variable falls within the NoGo response region on a given trial, we specify that the subject indicates either direction of motion with equal (50%) probability, with no response bias toward either direction. In this case, the probability of leftward and rightward Go responses is given by the following (Eq. 39):

$$p(Go, L|\mathbf{X}) = p(Y = L|\mathbf{X}) + \frac{1}{2}p(Y = N|\mathbf{X})$$

$$p(Go, R|\mathbf{X}) = p(Y = R|\mathbf{X}) + \frac{1}{2}p(Y = N|\mathbf{X})$$

These equations are the basis for the simulations shown in Figure 7 (see Materials and Methods, Motion direction discrimination task with SC microstimulation).

### Extension of the model to “filtering” tasks involving perceptual discrimination

In the previous section, we extended the m-ADC model to discrimination tasks in which the target stimulus occurs alone. Here, we demonstrate how a similar formulation can be applied to tasks in which the target occurs concurrently with one or more foil (distracter) stimuli (see Fig. 8). Such tasks are termed “filtering” tasks in literature (Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2010). We take the specific case of the filtering discrimination task employed in Lovejoy and Krauzlis (2010). The task is described in Figure 8A, and involves reporting the direction of motion in the target patch while ignoring motion in the foil patch(es).

In this model, we specify that the sensory evidence at the target and foil locations are represented along mutually orthogonal directions (along the  $x$ -axis and  $y$ -axis, respectively; see Fig. 8B). In the Lovejoy and Krauzlis (2010) task, the target-patch motion could occur in one of four diagonal directions: ( $+45^\circ$ ,  $-45^\circ$ ,  $+135^\circ$ ,  $-135^\circ$ ). The foil motion occurred in one of the remaining three directions. Here, for simplicity of illustrating the model, we posit that modeling any one pair of target and foil motion directions should give results not different from modeling any other pair of directions. We chose the direction of target motion to be up and to the right ( $+45^\circ$ ) and the direction of foil motion to be down and to the left ( $-135^\circ$ ). This simplification nevertheless captures the essential results of Lovejoy and Krauzlis (2010; see Fig. 8D,E). The task employed in the Lovejoy and Krauzlis study (2010), including the two other potential directions of motion ( $-45^\circ$ ,  $+135^\circ$ ), can be modeled in its entirety by adding two more dimensions orthogonal to the target and foil dimensions shown in Figure 8B.

For our model, the decision variable distribution is defined as follows (Eq. 40):

$$\Psi = \mathbf{d} + \varepsilon; \quad \varepsilon \sim \mathcal{N}(0, 1)$$

$$\Psi = [\Psi_t \ \Psi_f]^T; \quad \mathbf{d} = [d_t \ d_f]^T$$

where  $\Psi_t$  and  $\Psi_f$  are the marginal distributions of the decision variables for the target and foil locations, respectively, and the target and foil motion strengths are denoted  $d_t$  and  $d_f$ , respectively. The key difference with the previous latent variable formulation is that, in this case, because multiple stimuli (target, foil) are concurrently presented on each trial, the mean of the decision variable distribution ( $d_t, d_f$ ) does not lie along either coordinate axis but rather within one of the four quadrants (see Fig. 8B, dashed gray circle).

On the other hand, the decision rule is identical with the previous model, except that we posit that the subject employs choice criteria  $c_t$  and  $-c_t$  at the target's location, and  $c_f$  and  $-c_f$  at the foil's location when making a decision, as shown in Figure 8B. The decision rule, for this example, is as follows (Eq. 41):

$$Y = +45^\circ \text{ if } (\Psi_t > c_t \cap d_t(\Psi_t - c_t) > d_f(\Psi_f - c_f)) \cup (\Psi_t < -c_t \cap \Psi_f < -c_f \cap d_t(\Psi_t + c_t) > d_f(\Psi_f + c_f))$$

$$Y = -135^\circ \text{ if } (\Psi_f > c_f \cap d_t(\Psi_t - c_t) < d_f(\Psi_f - c_f)) \cup (\Psi_t < -c_t \cap \Psi_f < c_f \cap d_t(\Psi_t + c_t) < d_f(\Psi_f + c_f))$$

$$Y = N \text{ if } -c_t \leq \Psi_t \leq c_t \cap -c_f \leq \Psi_f \leq c_f$$

In this case,  $Y = +45^\circ$  represents a correct answer indicating the target direction,  $Y = -135^\circ$  represents an incorrect answer indicating the foil direction, and  $Y = N$  represents an incorrect answer indicating “neither” target nor foil direction (one of the other directions:  $-45^\circ$  or  $+135^\circ$ ). Alternative decision rules are also possible for this task, such as one in which the decision boundary is a line passing through the origin whose tilt indicates the bias for target versus foil, or one in which the boundaries of the “neither” response are oriented diagonally (rather than parallel to the axes). We employed the decision rule in Equation 41 here, as it was essentially identical with that of the previous model (Eq. 37) and captured key aspects of the Lovejoy and Krauzlis (2010) results. Formal model comparisons are necessary to select the best among these different decision rules.

We posit that the probability of a correct (target-based) or incorrect (foil-based or neither) response will remain the same regardless of the specific choice of target or foil motion. Hence, we define the probability of a response based on target motion,  $p(Y = 45^\circ | X)$ , as the probability of a correct response,  $p(Y = C | X)$ , and the probability of a response based on foil motion,  $p(Y = -135^\circ | X)$ , as the probability of an incorrect (foil-based) response,  $p(Y = F | X)$ . Thus, the probability of each type of response is given by the following (Eq. 42):

$$p(Y = C | X) = \int_{-\infty}^{-c_t} \left( \int_{-\infty}^{d_t(\Psi_t + c_t) - c_f} \Psi(\mathbf{X}; d_t, d_f d\Psi_f) d\Psi_f \right) d\Psi_t$$

$$+ \int_{c_t}^{\infty} \left( \int_{-\infty}^{-c_f} \Psi(\mathbf{X}; d_t, d_f d\Psi_f) d\Psi_f \right) d\Psi_t$$

$$+ \int_{c_t}^{\infty} \left( \int_{-\infty}^{d_t(\Psi_t - c_t) + c_f} \Psi(\mathbf{X}; d_t, d_f d\Psi_f) d\Psi_f \right) d\Psi_t$$

$$p(Y = F | X) = \int_{-\infty}^{-c_t} \left( \int_{d_t(\Psi_t + c_t) - c_f}^{\infty} \Psi(\mathbf{X}; d_t, d_f d\Psi_f) d\Psi_f \right) d\Psi_t$$

$$+ \int_{-c_t}^{c_t} \left( \int_{c_f}^{\infty} \Psi(\mathbf{X}; d_t, d_f d\Psi_f) d\Psi_f \right) d\Psi_t$$

$$+ \int_{c_t}^{\infty} \left( \int_{d_t(\Psi_t - c_t) + c_f}^{\infty} \Psi(\mathbf{X}; d_t, d_f d\Psi_f) d\Psi_f \right) d\Psi_t$$

$$p(Y = N | X) = \int_{-c_t}^{c_t} \int_{-c_f}^{c_f} \Psi(\mathbf{X}; d_t, d_f) d\Psi_f d\Psi_t$$

These equations are the basis for the simulations shown in Figure 8 (see Materials and Methods, Motion direction discrimination (filtering) task with SC inactivation).

## Behavioral tasks

### Four-alternative spatial target detection task

Chickens ( $n = 3$ ) were trained on a four-alternative target detection task based on luminance (Fig. 2A), reported previously (Sridharan et al., 2013). A subset of the data from the previous study was reanalyzed here within the framework of the m-ADC model to distinguish bias from sensitivity effects. The behavioral paradigm is summarized below. For details regarding the behavioral apparatus and training protocol, refer to Sridharan et al. (2013).

A brief (50 ms) dot stimulus,  $3^\circ$  in diameter (target), was presented at one of four potential locations in the bird's visual field. Target locations and dimensions were defined relative to the midsagittal plane and the horizontal plane that contains the optical axes of the eyes and the tip of the beak. To begin a trial, a small ( $\pm 5$  mm) cross appeared at the center of the computer screen (Fig. 2A). To peck on the cross, the chicken had to position its body between two vertical posts, forcing the body into a standard position relative to the screen. The chicken had to peck accurately on the cross (within 10 mm) to initiate a trial. Immediately after the peck on the cross, the target appeared on the screen for 50 ms. The bird was rewarded (brief access to food) for pecking at the location of the target on stimulus trials, and for pecking twice on the cross (NoGo) on catch trials. During training, NoGo responses to Go trials (misses) were not rewarded. However, Go responses to NoGo trials (false alarms) were penalized to minimize the rate of false alarms.

The target appeared at one of four possible locations:  $60^\circ$  in azimuth and  $60^\circ$  in elevation from the zeroing cross in each of the four visual field quadrants. Targets appeared with equal probability at any of the four locations, and the sequence of target locations was pseudorandomized. Luminance threshold for each target location was measured by stair-casing target contrast (Sridharan et al., 2013). The analyses reported here were performed after excluding the first several stimulus trials following session initiation ( $>10\%$  contrast targets) to allow the animal's behavior to stabilize, although the results were similar with these data included.

### Four-alternative orientation-change detection task

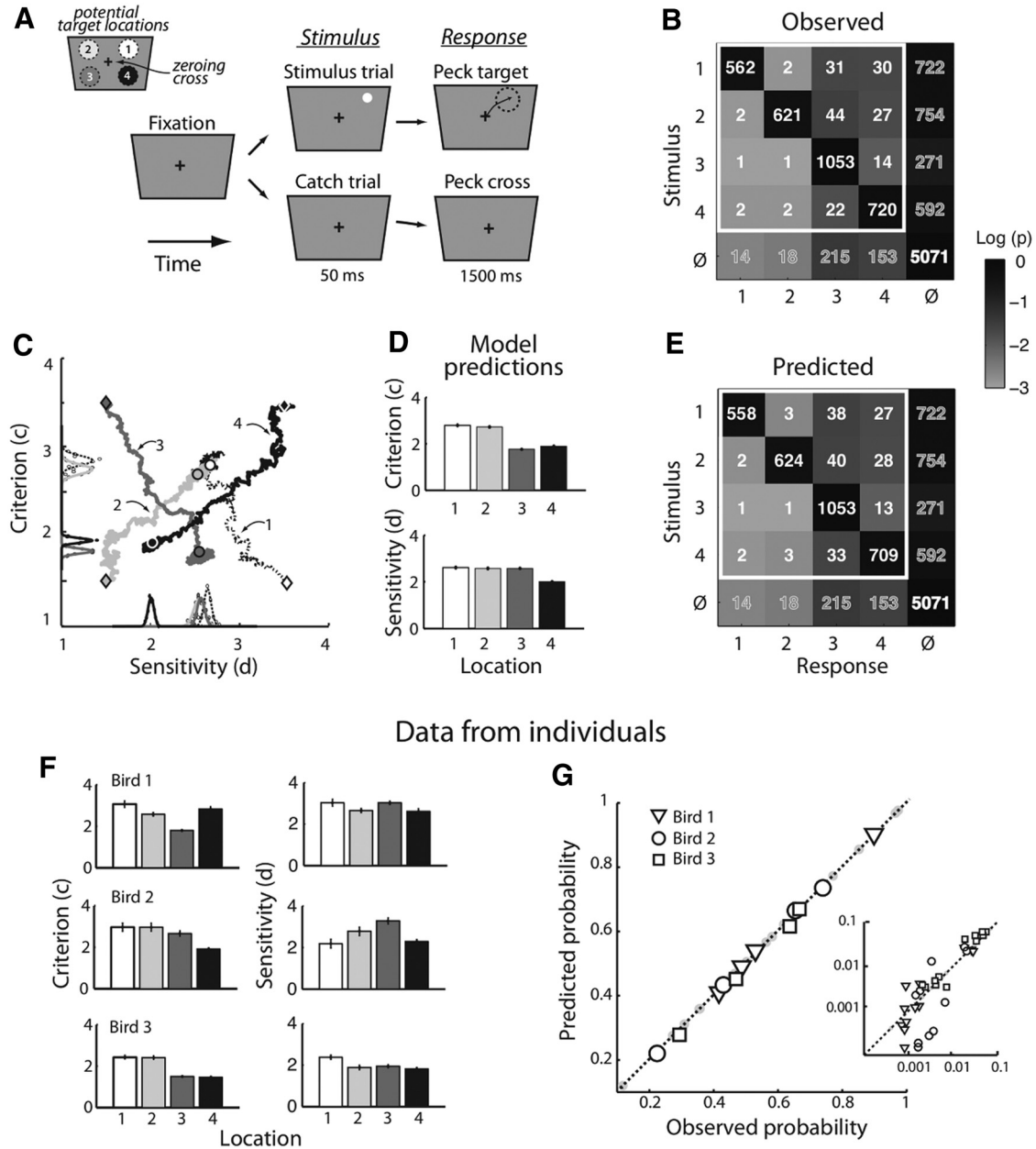
A monkey was trained on an orientation-change detection task with an antisaccade component. The task sequence was as follows (Fig. 3A). A small white dot appeared on the screen, and the monkey initiated a trial by fixating it. Within 100 ms, four peripheral stimuli appeared (described below). After a period of 800–2700 ms, the four stimuli disappeared for a brief interval ( $<270$  ms; blank period), and then reappeared. Upon reappearance, one of the four stimuli had changed its orientation (i.e., had been rotated in place) on 50% of trials. On these trials (change trials), the monkey could earn a reward by executing a saccadic eye movement within 800 ms to the stimulus located diagonally opposite to the changed stimulus (antisaccade). On the other 50% of trials (catch trials), all four stimuli reappeared with orientations identical to the initial orientations. In this case, the monkey was rewarded for maintaining fixation on the central dot for 800 ms (NoGo response).

The target stimuli were static, black and white square-wave gratings in a circular Gaussian aperture. The gratings were  $4^\circ$  in diameter, 1 cycle/ $^\circ$ , at full contrast for the monitor. The four gratings were presented at equal eccentricities (between  $5$  and  $8^\circ$  across sessions) and were spaced evenly around a circle. The screen background was dark gray. The orientations of the initial gratings took one of seven possible values, evenly spaced from  $0$  to  $157.5^\circ$  in  $22.5^\circ$  intervals. The initial grating orientations were chosen independently for each of the four stimuli and for each trial. Rotation magnitude ranged from  $10$  to  $90^\circ$  in  $5^\circ$  intervals. Rotation magnitude and direction were independently chosen for each trial.

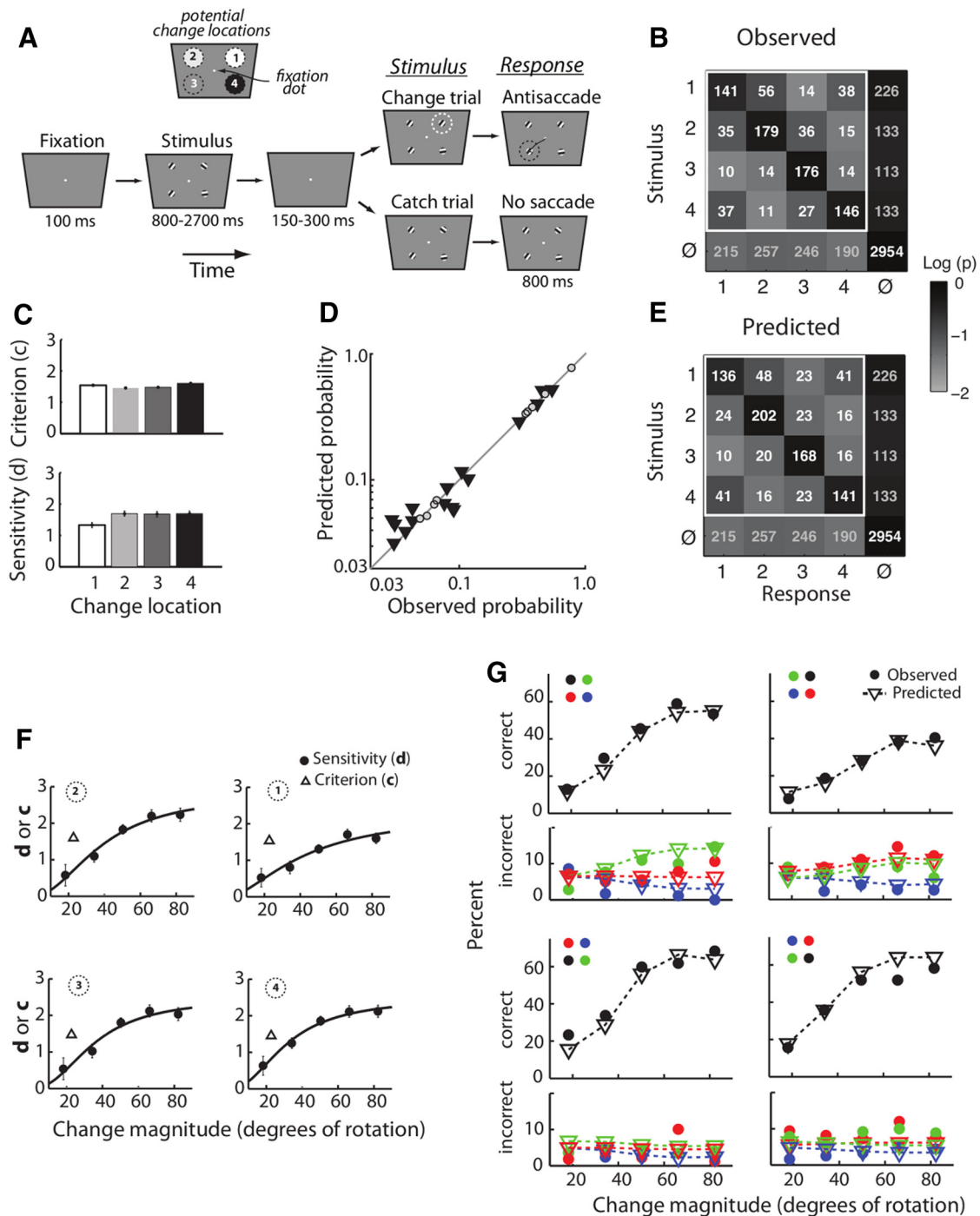
### Four-alternative, cued orientation-change detection task

The cued orientation-change detection task was identical to the previous task, except that change detection was made easier for the monkey by the presentation of a quasisymbolic cue indicating which location would contain the change. The cue consisted of a white line segment, less than half a degree in length and one pixel in width, originating at the fixation dot and extending in the direction of one of the four potential target locations. The cue appeared after a brief delay (300–500 ms) following presentation of the initial stimulus array, and indicated with 93% validity





**Figure 2.** Analysis of a four-alternative target detection task with the m-ADC model. **A**, Spatial, four-alternative target detection task based on target luminance. The bird (chicken) initiated a trial by pecking on the zeroing cross in the center of a touch-sensitive screen. Following trial initiation, a target (positive contrast, 3° dot) was presented briefly (50 ms) at one of four potential locations, one in each visual quadrant (stimulus trials). On interleaved trials, no target was presented (catch trials). Birds were rewarded for pecking accurately at target’s location when it occurred and for pecking at the zeroing cross (NoGo) on trials when no stimulus was presented. Stimulus and catch trials were equally probable (50%). Targets appeared with equal probability (25%) at any one of the four locations, pseudorandomly interleaved. Top left graphic, Numerical convention for identifying the target’s quadrant (1–4) on the screen. **B**, Observed response counts (numbers within cells) and conditional response probabilities (shading of cells). Data pooled across 108 experiments in three birds. Rows, Target locations (as defined in **A**); last row (∅), no-stimulus (catch). Columns, Response locations (as defined in **A**); last column (∅), NoGo responses. Cells within white square, Hits (diagonal cells) and misidentifications (off-diagonal cells) used for model predictions (**E**). **C**, Bayesian parameter estimation of the sensitivity (*x*-axis) and criterion (*y*-axis) at each target location (as defined in **A**) with the MCMC (Metropolis sampling) algorithm. Diamonds, Initial guess for each sensitivity and criterion; circles, final value for each sensitivity and criterion after convergence (medians of the posterior distributions); trajectories, Markov chains for convergence from the initial guesses to the final values. Curves along axes, Gaussian fits to the posterior distributions of the sensitivity and criterion values at each location. **D**, ML estimates of criteria (above) and sensitivities (below) for each location obtained by fitting the m-ADC to the behavioral data from **B**. Bar shading, Numerical convention for identifying the target’s quadrant, as defined in **A**. Error bars, ML standard error. **E**, Model predictions of response counts (numbers within cells) and probabilities (shading of cells) for the highlighted 4 × 4 cells of the contingency table, based on model parameters estimated from only the last row and the last column of the contingency table (misses and false alarms); conventions as in **B**. **F–G**, Analyses for individual birds. **F**, ML estimates of sensitivities (left) and criteria (right) for data analyzed from each bird individually with a 4-ADC model. Bar shading, Numerical convention for identifying the target’s quadrant, as defined in **A**. Error bars, ML standard error. **G**, Response probabilities predicted from the model plotted against experimentally observed response probabilities. Open symbols, Probabilities not used for fitting the model (predictions). Shaded dots, Probabilities used to fit the model. Symbol type, Individual birds. Inset, Smaller response probabilities in the contingency table plotted on a logarithmic scale.



**Figure 3.** Analysis of a four-alternative orientation-change detection task with the m-ADC model. **A**, Spatial four-alternative, orientation-change detection task. The monkey initiated a trial by fixating on a zeroing dot in the center of the screen. Following trial initiation, four stimuli (oriented gratings) were presented, one in each visual quadrant. After an unpredictable interval (800–2700 ms), the screen went blank (150–300 ms). Then, the four stimuli reappeared. On 50% of trials, one of the four gratings had changed in orientation by 10–90° (change trials), and on the remaining trials none of the stimuli had changed (catch trials). The animal was rewarded for making an eye saccade diagonally opposite to the location of the changed stimulus (antisaccade) when a change had occurred, and was rewarded for maintaining fixation on the zeroing dot (NoGo) when no change had occurred (see Materials and Methods). Catch trials constituted 50% of all trials, and during change trials, the orientation change could happen with equal probability (25%) at any one of the four locations. **B**, Observed response counts (numbers within cells) and conditional response probabilities (shading of cells). Data pooled across orientation change values in  $n = 22$  experiments. Rows, Change locations (as defined in **A**); last row ( $\emptyset$ ), no-stimulus (catch). Columns, Response locations (as defined in **A**); last column ( $\emptyset$ ), NoGo responses. Cells outside of white square, Cells representing misses and false-alarm rates used to derive model parameters. Other conventions are the same as in Figure 2B. **C**, ML estimates of sensitivities and criteria for each location estimated from data in **B**. Bar shading, Numerical convention for identifying the target’s quadrant, as in **A**. Error bars, ML standard error. **D**, Response probabilities predicted from the model plotted against experimentally observed response probabilities. Triangles, Probabilities not used for fitting the model (predictions). Shaded dots, Probabilities used to fit the model. **E**, Model predictions for the central  $4 \times 4$  cells of the contingency table based on model parameters estimated from only the last row and the last column of the contingency table (misses and false alarms); conventions as in **B**. **F**, Criteria ( $c$ , open triangles) and psychophysical functions of sensitivity ( $d'$ , circles) for various orientation change values estimated for each quadrant (defined in **A**) from only 8 of 25 contingencies (**B**, cells outside the highlighted box). Values of  $d'$  and  $c$  were estimated by binning the orientation change values into 16° bins. Black line, Naka–Rushton fit to the psychophysical function. Error bars, ML standard error. **G**, Psychometric functions of the proportion of observed responses (filled circles) and model predictions (open triangles) based on the sensitivity and criteria estimates from **F**. The eight plots are arranged as four matched pairs (top and bottom). The top (Figure legend continues.)

**Table 1. Model fits and parameters for 4-ADC target detection task: ML and Bayesian estimates of sensitivities and criteria (pooled data)**

Parameter	MLE (mean ± SE)	Bayesian (mean ± SE)	95% CI
Sensitivity	$d_1 = 2.62 \pm 0.08$	$d_1 = 2.66 \pm 0.06$	$d_1: 2.50 - 2.77$
	$d_2 = 2.58 \pm 0.08$	$d_2 = 2.57 \pm 0.09$	$d_2: 2.44 - 2.70$
	$d_3 = 2.58 \pm 0.05$	$d_3 = 2.59 \pm 0.06$	$d_3: 2.52 - 2.67$
	$d_4 = 2.00 \pm 0.05$	$d_4 = 2.00 \pm 0.05$	$d_4: 1.94 - 2.07$
Criterion	$c_1 = 2.80 \pm 0.07$	$c_1 = 2.83 \pm 0.08$	$c_1: 2.70 - 2.93$
	$c_2 = 2.73 \pm 0.07$	$c_2 = 2.73 \pm 0.08$	$c_2: 2.63 - 2.84$
	$c_3 = 1.77 \pm 0.03$	$c_3 = 1.77 \pm 0.04$	$c_3: 1.72 - 1.83$
	$c_4 = 1.90 \pm 0.03$	$c_4 = 1.90 \pm 0.04$	$c_4: 1.85 - 1.95$

Location 1, Upper right; location 2, upper left; location 3, lower left; location 4, lower right.

which of the four stimuli (if any) would change orientation. The stimulus array disappeared after 600–2200 ms, while the cue remained on the screen. Upon reappearance of the stimulus array, one of the four gratings had changed its orientation on 50% of the trials, and the monkey was rewarded for reporting the change location with an antisaccade to the diagonally opposite target location, or for maintaining fixation on catch trials. For this task, the amount of rotation was one of three values: 45, 67, or 90° (interleaved randomly). Performance during invalidly cued trials, in which the change occurred at a location other than that indicated by the cue, was measured only for the 45° change; these data were analyzed for this task. Data from this task have been published previously (Steinmetz and Moore, 2014). A subset of these data (first 12 recording sessions) was reanalyzed here within the framework of the m-ADC model to quantify bias and sensitivity effects induced by cueing.

### Model fitting

Model fitting was performed with unconstrained optimization techniques: maximum likelihood estimation (MLE; line search) and Bayesian [Markov-chain Monte Carlo (MCMC)]. These techniques have been described in detail in a previous study (Sridharan et al., 2014b). Both estimation approaches produced virtually identical estimates for the parameter values and the associated SEs (Table 1).

### Evaluating goodness of fit

We evaluated how well each model was able to account for the data with a randomization test (based on the  $\chi^2$  statistic). We selected the randomization test over the more conservative, Pearson's  $\chi^2$  goodness-of-fit test because of low counts (<5) for some stimulus–response contingencies. However, in the model comparison section, we also report  $p$  values with the Pearson's  $\chi^2$  test (see Tables 9, 10). The randomization test was conducted as follows: expected response probabilities were constructed based on the best-fit m-ADC model parameters (MLE). We simulated multinomial distributions of responses with random sampling from these expected probabilities after matching the overall number of trials for each stimulus event. This randomization procedure was repeated 100,000 times. The  $\chi^2$  statistic, based on observed and simulated response counts, was computed for each of these replications to generate a null distribution. Next, we computed the  $\chi^2_{\text{best-fit}}$  based on the observed responses and best-fit responses computed with the ML parameter estimates. The  $p_r$  values reported correspond to the proportion of values of the null distribution that exceeded  $\chi^2_{\text{best-fit}}$ . Model fits were deemed acceptable if  $p_r$  was >0.05 (typical  $p_r$  values were in the range of 0.1–0.9; see Tables 3, 8).

←

(Figure legend continued.) subplot of each pair (black symbols) shows the percentage correct (hits) as a function of change magnitude (psychometric function of accuracy). The bottom subplot of each pair (colored symbols) shows the percentage errors (incorrect/misidentifications) as a function of change magnitude. The four pairs are spatially aligned with the four corresponding locations of change occurrence (**A**). In each plot, colors indicate the locations of the responses relative to the location of change. Black, Response to the location of change; blue, response to the diagonally opposite location; red, responses to the location in the same vertical hemifield; green, responses to the location in the opposite hemifield.

### Evaluating predictive power

The predictive power of the model was measured as follows: consider a multialternative detection task with  $m$  potential target stimulus locations and catch trials (in total,  $m + 1$  stimulus events), and in which the subject can give a Go response to one of the  $m$  locations (where the target was detected) or can give a NoGo response (in total,  $m + 1$  response types). Thus, this task has an  $(m + 1) \times (m + 1)$  stimulus–response contingency table (Fig. 2B).

The m-ADC model fits the conditional probabilities of each response type for each stimulus event (Materials and Methods, Eqs. 23 and 24). Hence, of the  $(m + 1)^2$  observations,  $m^2 + m$  are independent (as conditional probabilities for all response types for a given stimulus event have to sum to 1). The m-ADC model is parametrized with only  $2m$  free parameters, the sensitivity ( $d$ ) and the criterion ( $c$ ) at each of the  $m$  locations, far fewer than the number of independent observations (for  $m \geq 2$ ). Therefore, we estimated model parameters based on a subset of the experimentally measured stimulus–response contingencies, specifically the misses and false alarms (last row and column of each contingency table; see below for why these contingencies were chosen). With parameter values thus estimated, we measured the ability of the model to predict the animals' responses to all of the remaining stimulus–response contingencies in the table (Fig. 2E, highlighted square). The predicted responses were then compared with the experimentally observed responses with goodness-of-fit tests, as described above.

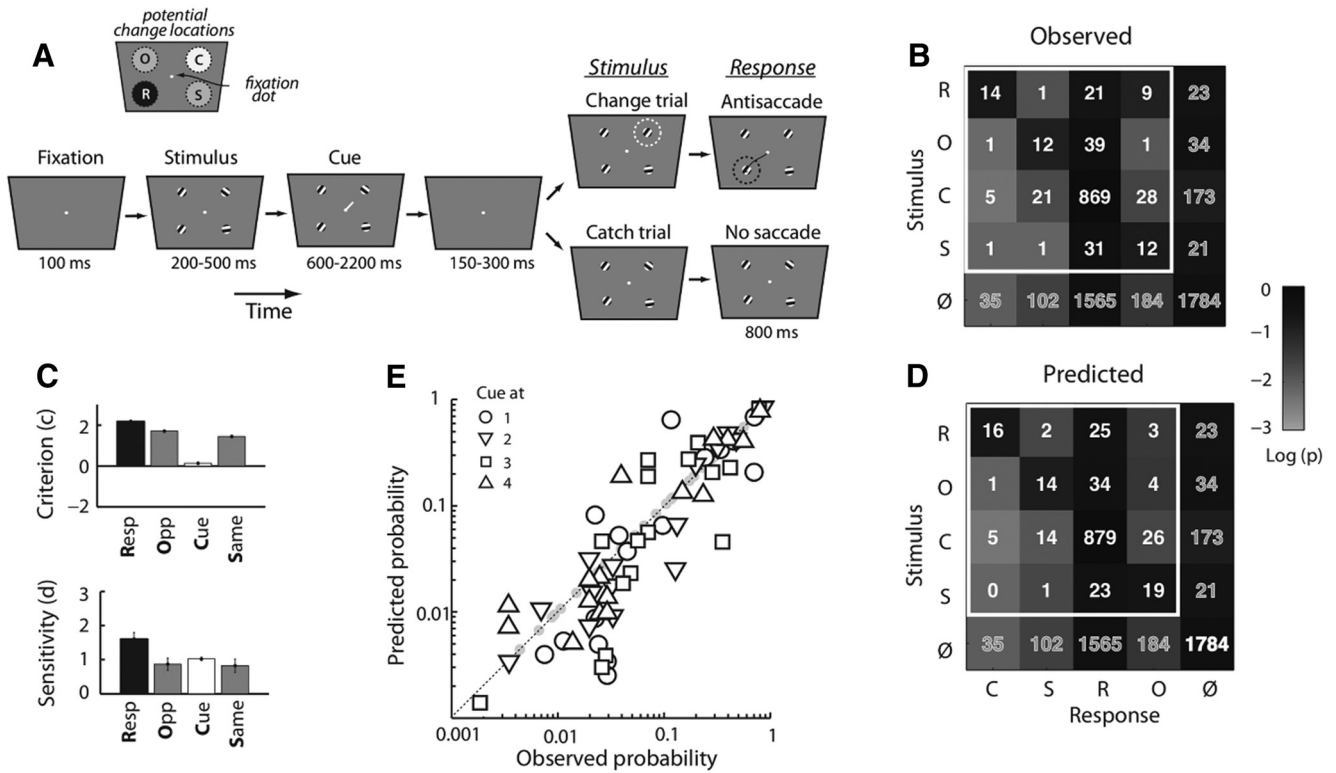
For each four-alternative task (Figs. 2A, 3A, 4A), we estimated parameters (sensitivities, criteria) based on data from eight stimulus–response contingencies, and tested the ability of the model to predict responses for the remaining 17 contingencies. At a minimum, eight independent observations were necessary to estimate the eight model parameters. For these predictions, we chose the four false-alarm and four miss rates (last row and column of the contingency table, respectively). The model required at least one independent observation from each row and column of the table: a sufficiently large number of responses for the miss (row) and false-alarm (column) contingencies ensured that parameter estimation was not sensitive to small variations in these response counts. We performed MLE with the false alarms and misses by using the exact response counts for these contingencies, and pooling across the response counts (and probabilities) for other contingencies (hits and misidentifications), i.e., by pooling across each row of the highlighted  $4 \times 4$  subset of the contingency table (Figs. 2B, 3B, 4B). Thus, the model did not have access to the distribution of responses in the  $4 \times 4$  table, and had to predict these based on the sensitivity and criterion estimated from only the aforementioned eight data points. The 17th contingency, the correct-rejection rate, was not used in parameter estimation, but was trivially predicted as  $1 - f_{\text{or}}$ , where  $f_{\text{or}}$  is the overall false-alarm rate on catch trials (sum of first four columns of the last row).

We tested for goodness of fit of the predicted to the observed responses with a randomization test (described above).  $\chi^2$  values for model fits were inflated when several cells in the contingency table contained low (<5) response counts, indicating apparently poor predictions (small  $p$  values; see Tables 3, 8, asterisks). For these cases,  $\chi^2$  statistics and  $p$  values were recomputed with counts aggregated across the off-diagonal entries (incorrectly reported Go responses) for each stimulus event. Acceptable fits with these aggregated counts (see Tables 3, 8, parenthetical values) indicated that the model successfully predicted percentage correct performance and the overall percentage of incorrect responses.

### Analysis of the four-alternative orientation-change detection task (monkey)

**Two-stage model, incorporating lapse rates.** The attentional demands imposed by the change-blindness paradigm and the complex rule of reporting a change by making an antisaccade response resulted in less-than-perfect asymptotic performance. Although the monkey rarely made saccades toward the change location, saccades to locations adjoining the antisaccade response location were occasionally made, even for the largest orientation changes (70–90°) and shortest blank durations (150–200 ms) tested.

Data from these trials, in which performance typically had asymptoted (Fig. 3G), were used to construct an asymptotic-performance (AP) ma-



**Figure 4.** Analysis of a cued orientation-change detection task with the m-ADC model. **A**, Spatial four-alternative, cued orientation-change detection task. A monkey was trained on a four-alternative, cued orientation-change detection task. The task was identical to the task shown in Figure 3A, except that on each trial a cue (directed line segment close to fixation spot), indicating the location of the upcoming change with >90% validity, appeared shortly after stimulus onset. **B**, Contingency table as in Figure 3B, except that reported responses correspond to a single magnitude of orientation change: 45°. **C**, ML parameter estimates, as in Figure 3C. **D**, Model predictions, as in Figure 3E. **E**, Response probabilities predicted from the model plotted against experimentally observed response probabilities for data analyzed individually for each cue location. Open symbols, Probabilities not used for fitting the model (predictions). Shaded dots, Probabilities used to fit the model. Symbol type, Individual cued locations.

trix (see Table 4; complement of a lapse-rate matrix). This matrix indicated the monkey’s best performance in executing a correct (antisaccade) response, following a detected change. Hence, for this task, we employed the following two-stage model to predict the animal’s actual response probabilities. Response probabilities were predicted with the m-ADC model using ML parameter estimates based on only miss and false-alarm rates (as described above). We then convolved these response probabilities, which corresponded to the animal’s perception of the change, with the AP matrix, which corresponded to the animal’s ability to execute the appropriate response. The data used for the construction of the AP matrix were excluded from further analysis. A single AP matrix was sufficient to accurately predict responses across the range of rotation magnitudes, and did not have to be tailored to fit responses for each rotation value.

When the change was preceded by a 100% valid cue (data acquired in trials interleaved with uncued trials), AP was near perfect (hit rates in the AP matrix were generally ~90%; see Table 5). The high validity of the cue provided the animal with reliable information regarding the response location before the upcoming change, resulting in the animal making far fewer mistakes when executing the subsequent response. Thus, the AP matrix was not necessary and did not significantly improve model fits in the cued detection task (Fig. 4).

**Estimation of psychometric functions.** Psychometric functions of change magnitude were constructed as follows: detection performance across orientation change magnitudes (10–90°) was binned in five evenly spaced (16°) bins (clockwise and counterclockwise rotations pooled). Miss rates for each bin of rotation magnitudes, and overall false-alarm rates during catch trials (which, by definition, did not vary with rotation magnitudes) were then entered into the MLE procedure to produce psychophysical functions of sensitivity across rotation magnitudes (Fig. 3F, filled circles) and corresponding criteria (Fig. 3F, triangle) for each

change location. Psychometric functions at each location were estimated for each rotation magnitude and were convolved with the AP matrix. The resulting predicted and observed psychometric functions are shown in Figure 3G (inverted, open triangles).

*Analysis of the cued orientation change detection task: pooling of data across cue positions*

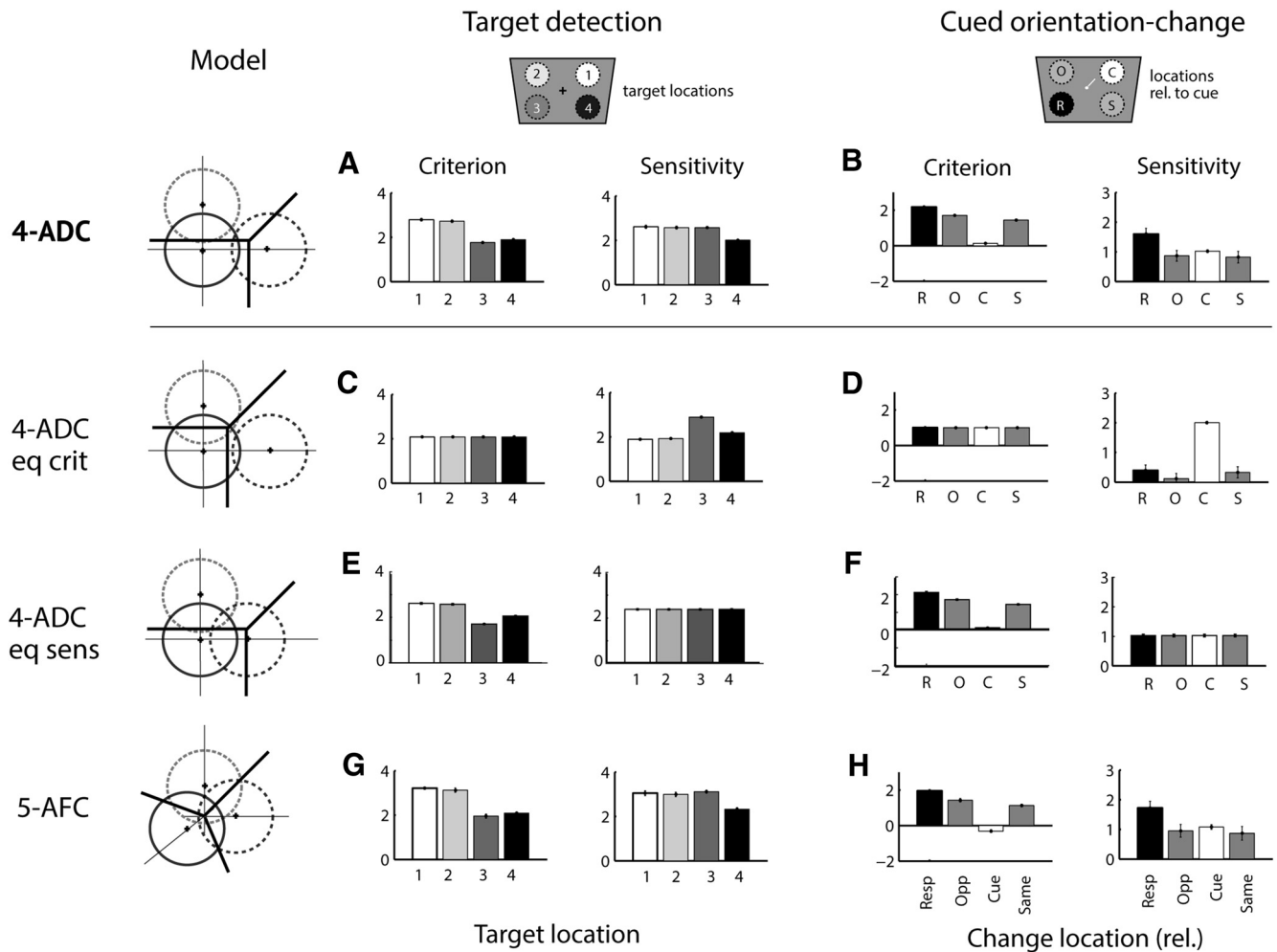
The four potential cue locations (one in each quadrant) yielded four 5 × 5 stimulus–response contingency tables (one for each cue location). Because invalid cues constituted only ~7% of all trials, and responses to these were distributed among 15 cells (3 cue-invalid locations × 5 potential responses to each), several stimulus–response contingencies contained very low response counts. To fit the model robustly, we combined response counts across cue locations by pooling responses to locations relative to the cue location as either same-as-cue (C), diagonally opposite (R), same hemifield (S), or mirror-symmetric opposite hemifield (O; Fig. 4A, top). This pooling created a single 5 × 5 contingency table (Fig. 4B) of the monkey’s responses to validly cued (Fig. 4B, third row) and invalidly cued (Fig. 4B, first, second, and fourth rows) locations. Responses for the change trials (Fig. 4B, first four rows) were acquired at a single magnitude of target orientation change: 45°.

Other analysis procedures were identical to those for the uncued orientation change detection task, except, as described above, the AP matrix (correction for lapse rates) was not used for this task, as it did not significantly improve fits.

*Model comparison*

We compared three candidate models against the 4-ADC model: (1) 4-ADC equal criteria (4-ADC eq crit), (2) 4-ADC equal sensitivity (4-ADC eq sens), and (3) 5-AFC.





**Figure 5.** Comparison of m-ADC model with other candidate models. Leftmost column, Schematics of three models compared with the 4-ADC model (top). Compared with the 4-ADC model, model 4-ADC eq crit and 4-ADC eq sens have three fewer parameters, whereas model 5-AFC has one additional parameter. **A, C, E, G**, ML estimates of criterion and sensitivity from each model for the four-alternative target detection task (chickens). **B, D, F, H**, Corresponding model estimates for the four-alternative, cued orientation-change detection task (monkey). Other conventions are as in Figure 2D. For details regarding models and model fitting, see Materials and Methods.

**The 4-ADC eq crit model.** The 4-ADC eq crit model is a signal-detection model that incorporates a symmetric decision rule with the same criteria for all alternatives (Fig. 5C,D; Macmillan and Creelman, 2005). The model specified a single, uniform choice criterion to stimuli at all locations, resulting in three fewer parameters (four sensitivity parameters and one criterion) compared with the 4-ADC model. This model is a simplification of the 4-ADC model in that all aspects were identical with the 4-ADC model except that the choice criteria for all locations were assumed to be equal.

**The 4-ADC eq sens model.** The 4-ADC eq sens model is a signal-detection model that assumes equal sensitivities at all locations (Fig. 5E,F; Macmillan and Creelman, 2005). The model specified a single, uniform  $d'$  to stimuli at all locations, resulting in three fewer parameters (four criterion parameters and one sensitivity) compared with the 4-ADC model. This model is a simplification of the 4-ADC model in that all aspects were identical with the 4-ADC model except that the sensitivities for all locations were assumed to be equal.

**The 5-AFC model.** The 5-AFC model is a signal-detection model that treats NoGo responses as an extra alternative in a forced-choice design (Fig. 5G,H). The data were analyzed with a model with one additional dimension, as if the task were a five-alternative forced-choice (5-AFC) design, based on a recent multialternative forced-choice model (DeCarlo, 2012). For fitting a  $5 \times 5$  design, this model requires estimating nine parameters (five sensitivity and four criterion parameters), one more than the m-ADC model.

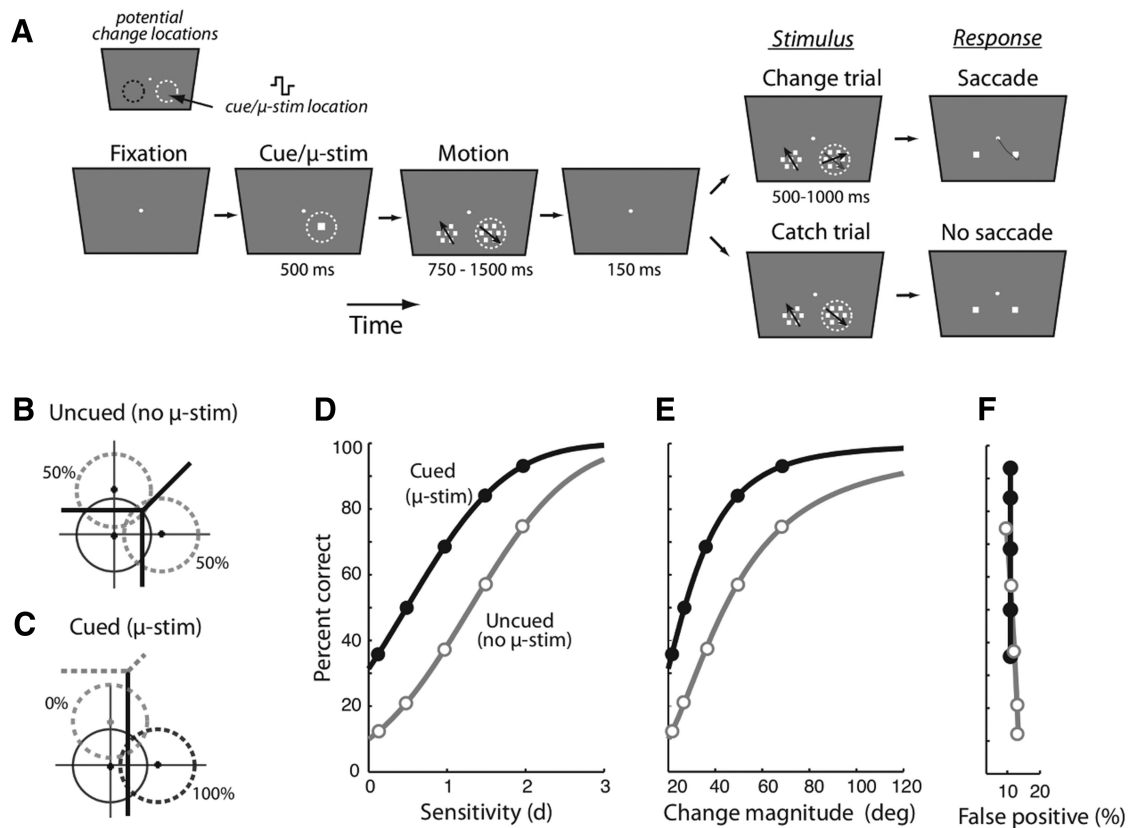
These candidate models and the number of free parameters for each model are given in Tables 9 and 10. Parameters were estimated with the MLE approach. For the 5-AFC model, all criteria were quantified relative to the criterion for the NoGo response as reference. With the log-likelihood function from these ML model fits, we computed the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC; see Tables 9, 10; Burnham and Anderson, 2002).

Matlab code for fitting the m-ADC models (ML and MCMC) and evaluating goodness of fit (randomization test) is available upon request.

#### Model simulations and ideal observer analysis

The 2-ADC model was used to simulate data from attention tasks that involve spatial cueing (Fig. 1E,F) or microstimulation or inactivation of the SC (see Figs. 7–9). Specific values for sensitivity and choice criterion parameters were chosen based on ideal observer analysis (described below) or with the aim of matching observed behavioral performance in published studies (Cavanaugh and Wurtz, 2004; Müller et al., 2005; Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012). In all of these simulations, sensitivities are either assumed to be equal at all locations or identical across conditions (e.g., control and inactivation). Hence, altering the criterion value differentially across locations simulates a change of bias.

**Simulation of the cued orientation-change detection task.** For the simulations shown in Figure 1E–G,  $d$  values across the two locations were



**Figure 6.** Bias effects in a multialternative detection task with cueing or with microstimulation of the SC (Cavanaugh and Wurtz, 2004). **A**, A two-alternative motion direction-change detection paradigm with interleaved cueing or microstimulation. Following fixation, two patches of random dot motion appeared. After a variable duration, the patches briefly disappeared ( $\sim 150$  ms) and reappeared. On some trials (change trials), the direction of motion had changed in one of the patches (target patch; dashed circle) upon reappearance; in other trials (catch trials) no change in direction of motion occurred in any of the patches. The monkey was rewarded for making an eye saccade to the target patch on change trials and for maintaining fixation (NoGo) on catch trials. In interleaved trials, a spatial cue indicated the location of the target patch (location of putative motion change) with 100% validity. In other experiments, in place of spatial cues, focal electrical microstimulation was delivered to the spatial representation of the target patch in the SC. **B**, 2-ADC model for uncued trials (or trials without microstimulation). Dashed gray circles,  $\Psi$  Distribution for changes at each location on uncued trials. Solid black circle,  $\Psi$  Distribution for no change trials. **C**, Model for cued trials (or trials with microstimulation). Dashed black and gray circles,  $\Psi$  Distribution for changes at cued and uncued locations, respectively, on cued trials. Dashed thick gray line, The criterion (decision boundary) at the uncued location moves to the edge of decision space (see text and Materials and Methods). **D**, Simulated percentage correct as a function of sensitivity for detecting motion changes ( $d$ ) during cued/microstimulated (closed circles) and uncued (open circles) trials. **E**, Simulated psychometric function, percentage correct as a function of motion direction change magnitude during cued/microstimulated and uncued trials. Other conventions are the same as in **D**. **F**, Simulated percentage correct versus percentage of false positives for detecting motion changes during cued/microstimulated and uncued trials. Other conventions are the same as in **D**.

assumed to be identical, and identically distributed (uniform distribution) across a range of 0 to 3.0 (measured in units of noise SD). We performed ideal observer analysis by identifying the pair of criteria (one at each location) that maximized the overall percentage correct for specified values or distributions of sensitivity ( $d$ ), including performance on all trial types (cued or uncued, change or catch; see below). These optimal criteria were identified by an exhaustive search on a  $400 \times 400$  grid over a range of criteria spanning the interval  $[-5.0, 5.0]$ .

The percentage of correct and incorrect responses predicted by the 2-ADC model are a function of four parameters: the criterion and the sensitivity at each of the two locations. We sought to reduce the number of free parameters in the analysis. Training protocols for multialternative detection tasks seek to reduce the tendency of the subject to guess when no stimulus or change occurs (catch trials); this is achieved by rewarding the subject for not guessing during catch trials (Cohen and Maunsell, 2009; Zénon and Krauzlis, 2012; Steinmetz and Moore, 2014), increasing the proportion of catch trials (Cavanaugh and Wurtz, 2004), and/or penalizing (e.g., with time outs) Go responses during catch trials (false alarms; Sridharan et al., 2013). As a result of this reinforcement strategy, it is reasonable to infer that the subject would maintain a constant and low proportion of false alarms (Go responses) during catch trials. The implication of this assumption for 2-ADC simulations is that a decrease in criterion at one location would be accompanied by an increase in

criterion at the other location, thereby quantitatively amplifying the behavioral effect in our simulations. With this assumption, the number of free parameters in our analysis fell to just one criterion; the other was automatically determined by the proportion of false alarms during catch trials.

For this task, the percentage correct was computed as the percentage of responses to the location of change in trials in which a change occurred at any location. Psychophysical functions describing the variation of sensitivity ( $d$ ) with change magnitude (either orientation or motion direction) were modeled with a Naka–Rushton (hyperbolic ratio) function of the following form:  $d = d_{\max} (\theta^n) / (\theta^n + \theta_{50}^n)$ , where  $\theta$  is the change magnitude,  $d_{\max}$  is the asymptotic sensitivity,  $n$  is the exponent (slope) of the psychophysical function, and  $\theta_{50}$  is the orientation change magnitude at which the sensitivity reaches half of its asymptotic value. For these simulations, we used the following values of these parameters:  $d_{\max} = 4.0$ ,  $n = 1.0$ ,  $\theta_{50} = 40^\circ$  (for the orientation-change detection task), and  $32^\circ$  (for the motion-change detection task).

*Simulation of the motion direction-change detection task with cueing or SC microstimulation.* For these simulations (Fig. 6),  $d$  values across the two locations and parameters of the psychophysical function (Naka–Rushton) were taken to be same as those for the previous simulations. For the cued trials in this task, the optimal 2-ADC model effectively reduces to a 2-AFC model. This can be understood intuitively

tively as follows: because a highly trained monkey knows that on cued trials any change will occur only at the cued location, it can ignore uncued locations (no matter how many there are) and treat such trials as if from a simple, binary choice 2-AFC or Yes/No task: did a change happen at the cued location, or not? Since the type of trial (cued vs uncued) is revealed to the animal well before the actual change occurs, it is reasonable to posit that the animal adopts different choice criteria for cued and uncued trials. Analytically, this can be demonstrated with optimal decision theory: the prior odds ratio (and, hence, the posterior odds ratio) of a change at an uncued location is zero. Hence, the optimal criterion at the uncued location moves to the edge of decision space (Fig. 6C), resulting in a large bias against responding to uncued locations. In this case, the 2-ADC model is identical to a 2-AFC model [see Materials and Methods, Optimal two-alternative detection model for 100% valid cues is a two-alternative forced choice (Yes-No) model] for deciding between two possibilities: change versus no change (catch trials) at the cued location.

For this task, the percentage of false positives (Fig. 6F) were computed as the percentage of incorrect responses to a location when no change had occurred at the location, regardless of whether a change had occurred at another location, or nowhere (catch trials). This is identical to the percentage of incorrect Go responses on all trials, and differs from (but is closely related to) our definition of false-alarm rates (Figs. 2–4), the conditional probability of false-alarm (Go) responses during catch trials alone. When matching the percentage of false positives between uncued and cued trials (Fig. 6F), false alarms on cued trials were modeled exclusively as responses toward the cued location when no change occurred anywhere.

We note a difference between the experimental task in the original study (Cavanaugh and Wurtz, 2004) and our simulated task: in the original study, the change in the direction of motion could occur at one of three patch locations. To simulate this, a 3-ADC model is required. However, the results of a 3-ADC model are not easily represented on paper. To simplify the representation of the results, and to highlight the essential conceptual advance, we simulated a two-alternative version of the task (Fig. 6A), in which changes in motion direction could occur at one of only two patch locations, with a 2-ADC model. The behavioral effects simulated with the 2-ADC task are a lower bound of the effects expected with a task with multiple ( $\geq 3$ ) alternatives. This can be intuitively understood as follows: the greater the number of alternatives in a task, the greater the reduction in uncertainty (or improvement in certainty) about the location of the upcoming target with a 100% valid cue (or with microstimulation). Hence, the expected behavioral improvement arising from a change in the decision rule (Fig. 6B, C) would be greater for a task with a greater number of alternatives.

**Simulation of the motion direction discrimination task with SC microstimulation.** For the perceptual discrimination task in Figure 7, directions of motion of the target patch were equally distributed (50%) between leftward and rightward motion, and across the two locations. Psychophysical functions describing the variation in sensitivity ( $d$ ) with the proportion of coherent dots were modeled with a Naka–Rushton (hyperbolic ratio) function of the following form:  $d = d_{\max} (\rho^n) / (\rho^n + \rho_{50}^n)$ , where  $\rho$  the proportion of coherent dots was sampled in uniform, logarithmically spaced bins from 10 to 100%,  $d_{\max}$  is the asymptotic sensitivity,  $n$  is the exponent (slope) of the psychophysical function, and  $\rho_{50}$  is the proportion of coherent dots at half-maximum sensitivity. For these simulations, we used the following parameters:  $d_{\max} = 4.0$ ,  $n = 2.0$  and  $\rho_{50} = 25\%$ , although the results were fairly robust to the specific values of these parameters.

The discrimination model for this task is elaborated in Materials and Methods, Extension of the model to perceptual discrimination tasks. At baseline, there is no spatial choice bias toward any location, so that  $c_1 = c_2 = 1.5$ . Upon microstimulation, the choice criterion at the microstimulated location reduces symmetrically, by 20%, for both directions of motion (Fig. 7C;  $c_1 = 1.2$ ,  $c_2$  is unchanged). This magnitude of change in criterion (20%) yielded a performance improvement (in percentage correct) of  $\sim 6$ –8% for intermediate values of motion coherence ( $\sim 20$ –50% coherence), which mimicked the performance improvement obtained with SC microstimulation by Müller et al. (2005) over a similar

range of target motion coherence ( $\sim 20$ –50%; Müller et al., 2005, their Fig. 4b). For this task, percentage correct was computed as the proportion of correct motion direction reports. The model was simulated with these parameters to produce the psychometric functions shown in Figure 7D, E.

**Simulation of the cued motion direction discrimination (filtering) task with SC inactivation.** For this task (Lovejoy and Krauzlis, 2010), target and foil stimuli were assumed to occur at one of two spatial locations (Fig. 8A, top right, bottom left) with equal probability. The direction of the motion of the cued (target) patch was equally distributed among one of four potential directions of motion ( $\pm 45$  and  $\pm 135^\circ$  from the horizontal). The foil-patch motion could occur in any one of these four directions, as long as it did not match the direction of target-patch motion on that trial; in Figure 8, an exemplar trial is shown in which the target-patch motion is at  $+45^\circ$  and the foil-patch motion is at  $-135^\circ$ .

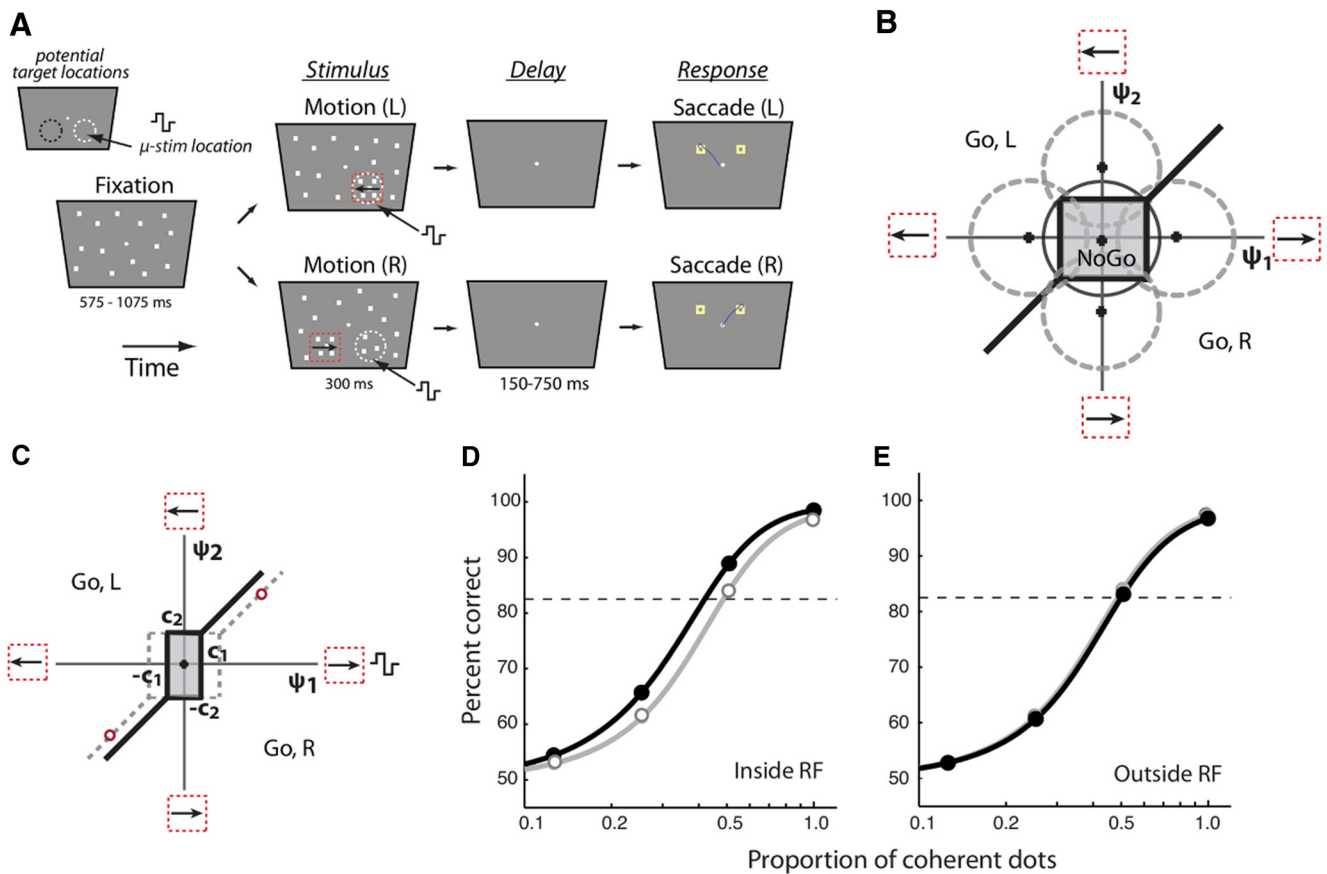
The discrimination model for this task is elaborated in Materials and Methods, Extension of the model to filtering tasks involving perceptual discrimination. In this model, we posit that sensory evidence for the target-patch motion and that for the foil-patch motion are evaluated on orthogonal axes. A key difference from the model for the previous task is that, when the decision variable value falls below the criteria at both locations (central gray zone), the animal indicates a location that is different from both the target and the foil location (“neither”).

For these simulations, the sensitivity ( $d$ ) to the target patch (cued) was taken to be 1.35 and that for the foil patch was taken to be 1.05. To model the animal’s bias to base its decisions upon the target’s sensory evidence, we posited that the choice criterion for the target ( $c_t = 0.2$ ) was substantially lower than the choice criterion for foil ( $c_f = 1.4$ ; Fig. 8B). Upon inactivation of the SC representation at the target’s location, we posited that the values of these criteria reverse ( $c_t = 1.4$  and  $c_f = 0.2$ ; Fig. 8C). Upon inactivation of the SC representation of the foil’s location, we posited that the criteria further skews, relative to baseline, in favor of the target ( $c_t = 0.14$  and  $c_f = 6.0$ ). The simulated proportions of the animal’s choices regarding the direction of the target and foil patches (Fig. 8D, E, squares) were derived from m-ADC model simulations based on these parameter values. To model the experimentally observed variation in the proportion of these choices (Lovejoy and Krauzlis, 2010), we repeated these simulations 30 times. In each simulated run, the criteria and sensitivities were selected randomly within a range  $\pm 50$  or  $\pm 33\%$  of their mean values (given previously), respectively. The proportions of choices from each simulated run are shown in Figure 8D, E.

**Simulation of the cued motion direction change detection (filtering) task with SC inactivation.** For the “filtering” task in Figure 9, sensitivities ( $d$  values) for motion change detection were assumed to be 2.0 at the cued location and 1.0 at the uncued location. In this task, the animal reported a binary Yes/No response (button press) without indicating the specific location of perceived change. To capture this task design, we adopted a slight variant of the 2-ADC decision rule: the animal was hypothesized to provide a “Yes” response if the decision variable exceeded criterion at either location, and a “No” response otherwise (Fig. 9A).

For this task, percentage correct (Fig. 9D–G) was computed as the percentage of “Yes” responses for trials in which the change occurred at the cued location, and error rates (percentage incorrect) were computed as the percentage of “Yes” responses for trials in which the change occurred at the uncued location. To quantitatively reproduce the experimentally observed effect of SC inactivation (Zénon and Krauzlis, 2012), we posited an increase in criterion ( $\Delta c_{\text{inact}} = 1.5$ ) at the inactivated location. The criterion at the noninactivated location was estimated from the criterion at the inactivated location based on the false-alarm rate (“Yes” responses) during catch trials. The effect of inactivation on the choice criterion was assumed to be identical across cued and uncued trials, so the same  $\Delta c_{\text{inact}}$  was applied regardless of whether the inactivated location was cued or uncued. Thus, a change in a single parameter,  $\Delta c_{\text{inact}}$  was sufficient to capture the key behavioral effects of SC inactivation (Zénon and Krauzlis, 2012).

**8-ADC model fitting of the cued motion direction discrimination (filtering) task with SC inactivation.** This task (Lovejoy and Krauzlis, 2010) involved the simultaneous presentation of target motion and foil motion patches, each moving in one of four directions, producing



**Figure 7.** Bias effects in a motion discrimination task upon microstimulation of the SC (Müller et al., 2005). **A**, A 2-ADC paradigm involving a direction of motion discrimination judgment. Following fixation, a wide-field array of flickering dots was presented. After a variable duration (575–1075 ms), a small patch of dots began to move coherently in one of two directions (left vs right). The motion patch was located, on interleaved trials, either within the receptive field of the SC neuron or in a mirror symmetric location about the vertical meridian. After a brief delay (150–750 ms), two response boxes appeared. The monkey was rewarded for indicating the direction of motion of the patch by saccading to the appropriate response box. The proportion of coherently moving dots was varied among four values (12.5–100%) to measure the psychometric function. On interleaved trials, focal microstimulation was delivered at the SC receptive field location. **B**, 2-ADC model for a perceptual discrimination with NoGo responses. *x*-axis, Sensory evidence at location 1 (microstimulated location). Positive values indicate evidence in favor of rightward motion. Negative values indicate evidence in favor of leftward motion. *y*-axis, Sensory evidence at location 2 (nonmicrostimulated location). Positive values indicate evidence in favor of leftward motion. Negative values indicate evidence in favor of rightward motion. Thick lines, Decision boundaries parameterized by two choice criteria ( $c_1$  and  $c_2$ ; see **C**). Central gray zone, Region of decision space corresponding to NoGo response. A Go response indicating rightward motion is made when the sensory evidence falls in the lower-right half-plane and outside the NoGo decision region. Similarly, a Go response indicating leftward motion is made when the sensory evidence falls in the upper-left half-plane and outside the NoGo region. Dashed circles, Potential decision variable distributions. **C**, Same as in **B**, but following microstimulation of location 1. Thick lines, Decision boundary following microstimulation of location 1, such that  $c_1$  decreases and  $c_2$  remains unchanged. The (default) decision boundary with no microstimulation is also shown for reference (dashed gray lines). Red circles lie along the lines of equal strength of (conflicting) sensory evidence at the two locations. **D**, Simulated psychometric function, which indicates percentage correct as a function of the proportion of coherent dots, with (black) or without (gray) SC microstimulation. The microstimulation site was aligned with the location of the coherent motion. **E**, Same as in **D**, but with the microstimulation site away from the location of coherent motion. Other conventions are the same as in **D**.

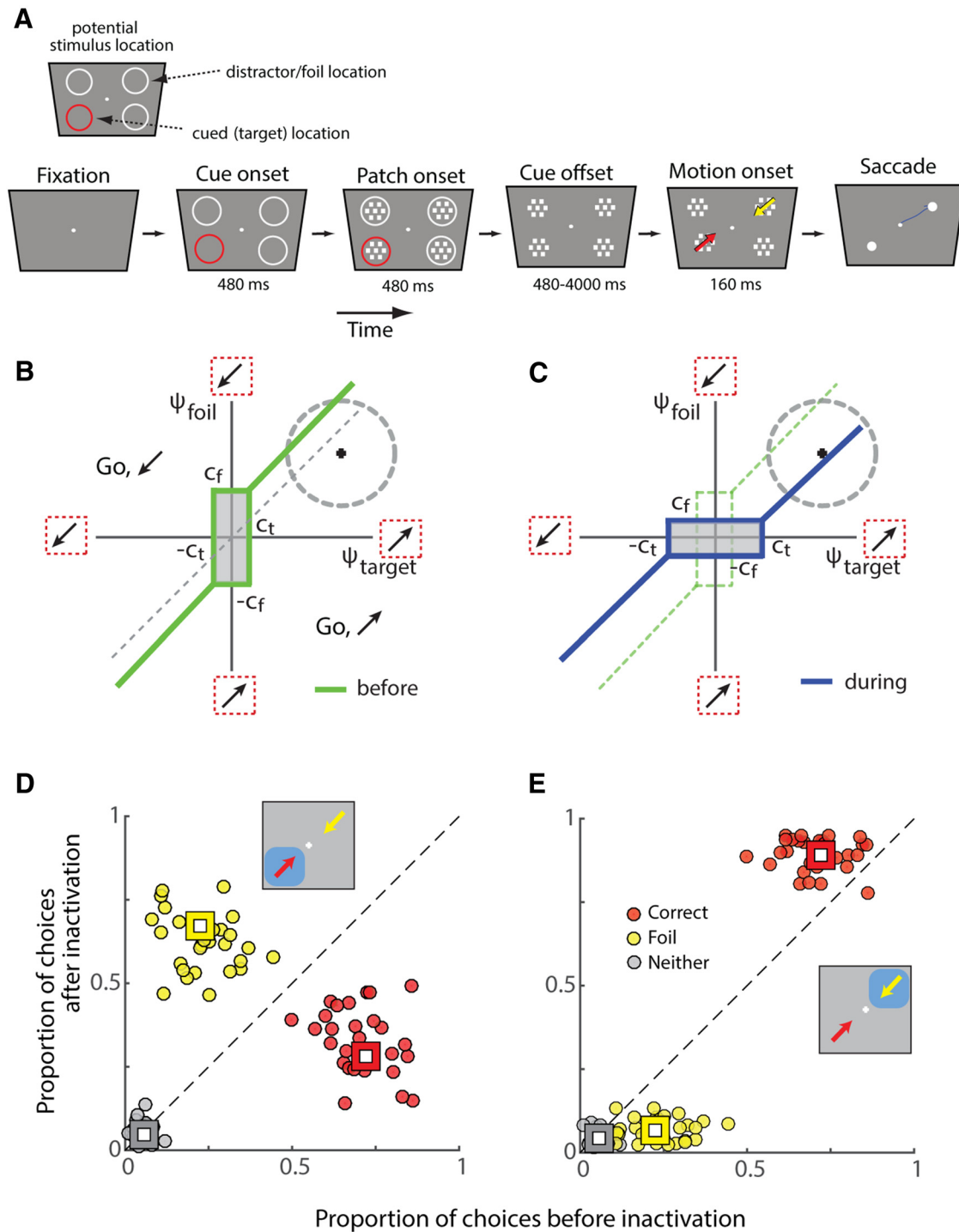
16 ( $4 \times 4$ ) potential stimulus combinations. However, on no trial did the target and foil move in the same direction. Thus, only 12 target/foil motion combinations were tested. Animals responded by indicating the direction of motion of the target patch with a saccade or button press. This produced a  $12 \times 4$  contingency table. Because the animal made very few responses to directions other than those corresponding to target or foil motion, we pooled these responses. Thus, we analyzed a  $12 \times 3$  contingency table corresponding to 12 stimulus events and three response types: response indicating the direction of target patch, the foil patch, or neither. We analyzed only the subset of sessions in which target and foil stimuli were both presented, and only those trials in which the stimuli contained a motion coherence of 0.1875 for subject F and 0.25 for subject M; in the original study, these values were chosen to maintain performance at 65–70% for both monkeys (Lovejoy and Krauzlis, 2010, their Fig. 4).

We developed an 8-ADC model for fitting these data. In this model, sensory evidence for each of the four directions of motion of the target and the foil were represented by independent decision variables in an eight-dimensional decision space. In principle, sensitivity ( $d$ ) and criterion ( $c$ ) parameters could be estimated for each of the eight decision

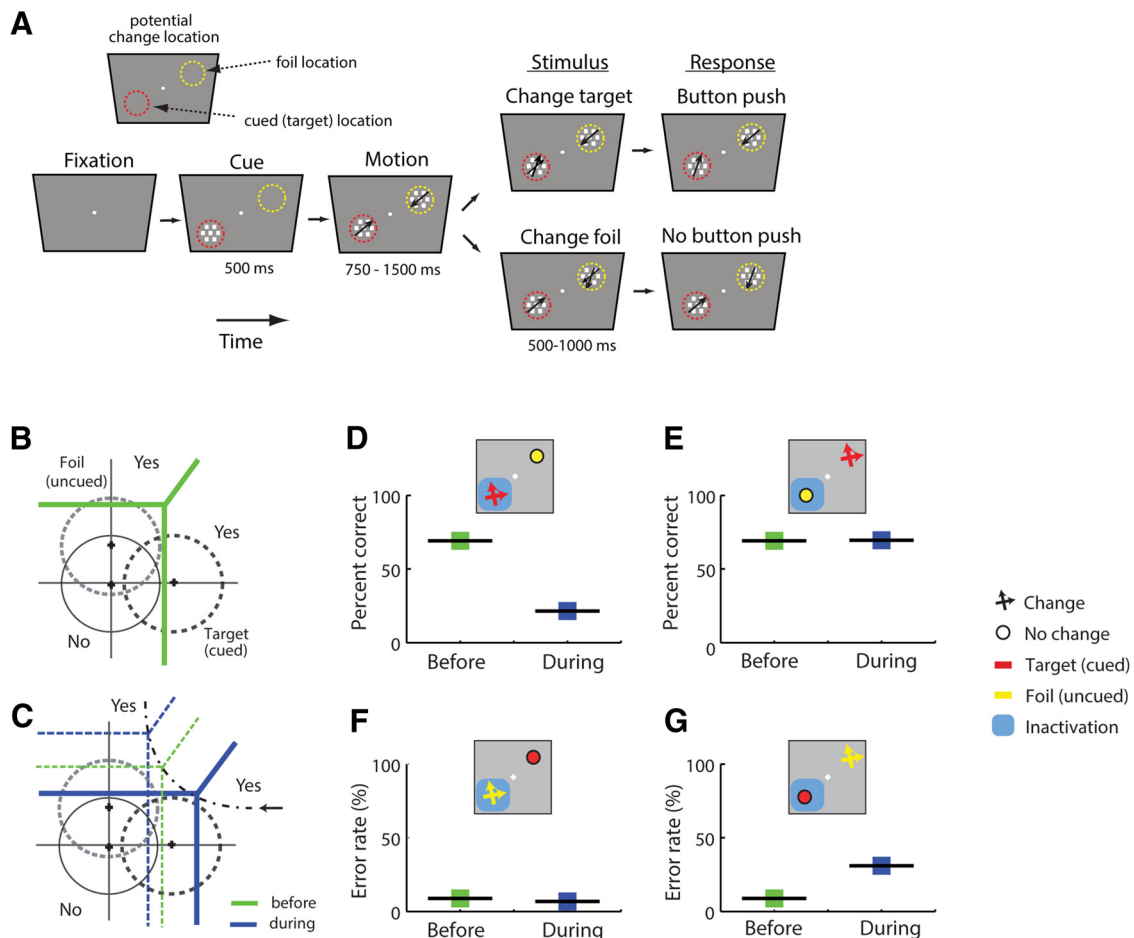
variables (total of 16 parameters), because there are 24 independent conditions in the contingency table. To make the model more parsimonious, we estimated the  $d$  and  $c$  values for the targets, but included only one parameter that modeled the ratio of the target to the foil sensitivity, and one parameter that modeled the difference between criterion values for targets and foils. In addition, because NoGo responses were not included in this task, we fixed one of the criteria to a low value ( $-3.0$ ). This is required because in a task with no NoGo responses, it is the difference of criteria, rather than their absolute values that are relevant for model fitting (for a detailed explanation, see Sridharan et al., 2014b, their section F of Supporting Information). Here, the criterion for target direction 1 was set to this value, although the same results are expected regardless of this specific choice. These assumptions reduced the overall number of parameters to nine. Despite these simplifying assumptions, the model provided good fits to data from both animals, as described in the Results.

SC inactivation produced an additional set of two  $12 \times 3$  contingency tables: one corresponding to trials in which the target was presented in the inactivated zone, and another for trials in which the foil was presented in the inactivated zone. SC inactivation could produce a response





**Figure 8.** Bias effects in a cued motion direction discrimination (filtering) task upon inactivation of the SC (Lovejoy and Krauzlis, 2010). **A**, A task paradigm involving discriminating the direction of motion of a cued target patch in the presence of multiple distractors. Following fixation, a cue (red circle) indicated the location of the upcoming target. Following this, four patches of random, incoherent dot motion appeared, one in each quadrant. Following this, the cue offset. At an unpredictable time, the target patch, and the patch in the diametrically opposite location (foil patch) moved coherently in different directions for a brief period (160 ms). The monkey was rewarded for reporting the direction of motion of the target patch with a saccade in that direction, while ignoring the direction of motion of the foil patch. **B**, 2-ADC model for perceptual discrimination in a filtering task. Dashed gray circle, Decision variable distribution for stimuli at the target and foil locations.  $x$ -axis, Sensory evidence corresponding to the target-patch motion. For this example, positive values indicate evidence in favor of motion toward the upper right, and negative values indicate evidence in favor of motion toward the lower left.  $y$ -axis, Sensory evidence corresponding to the foil-patch motion. Positive and negative values indicate evidence in favor of motion toward the lower left and upper right, respectively. Thick green line, Decision boundary at baseline (before inactivation). Central gray zone, Region of decision space corresponding to responses based on “neither” stimulus (neither target nor foil). Other conventions are the same as in Figure 7B. **C**, Same as in **B**, but following SC inactivation. Thick solid blue line, Decision boundary following inactivation of the target location. The decision boundary before inactivation is shown for reference (dashed green line). **D**, Simulated proportion of choices based on target motion direction (Correct, red), foil motion direction (Foil, yellow) or neither direction (Neither, gray) following inactivation of the cued (target) location. Dots, Proportions obtained over  $n = 30$  simulated runs by introducing limited, random variations in sensitivities and criteria (see Materials and Methods). Squares, Proportions based on mean values of sensitivities and criteria. **E**, Same as in **D**, but for inactivation of the foil location.



**Figure 9.** Bias effects in a cued motion-change detection (filtering) task upon inactivation of the SC (Zénon and Krauzlis, 2012). **A**, A two-alternative paradigm involving the detection of a change in motion direction of a cued target stimulus in the presence of a foil stimulus. Following fixation, a cue (static dot patch) indicated the location of the upcoming target. Following this, two patches of random dot motion appeared, one at the cued location (Target, red circle) and the other at the opposite location (Foil, yellow circle). At an unpredictable time, the motion direction of one of the patches changed. The monkey was rewarded for reporting a motion direction change in the target stimulus with a button push, and for ignoring changes that occurred in the foil stimulus. **B**, 2-ADC model for change detection in a filtering task. Dashed black and gray circles, Decision variable distributions for changes at the target and foil locations, respectively. Thick green line, Decision boundary at baseline (before inactivation). The boundaries partition decision space into three nonoverlapping domains. In two of the domains, the animal chooses to report a change (Yes); in the third, it reports no change (No). **C**, Same as in **B**, but following SC inactivation. Thick solid blue line, Decision boundary following inactivation of the target location. Thin dashed blue line, Decision boundary following inactivation of the foil location. The decision boundary before inactivation is shown for reference (dashed thin green line). Dot-dashed black line, Contour of choice criteria corresponding to constant false-alarm rates during catch trials. **D**, Simulated percentage correct before (green) and during (blue) inactivation when the motion change occurred at the target location (red) and the target location was inactivated. **E**, Same as in **D**, but for inactivation of the foil location (yellow). **F**, Simulated error rates before and during simulated inactivation when the motion change occurred at the foil location (yellow) and the foil location was inactivated. **G**, Same as in **F**, but for inactivation of the target location (red). **D–G**, Crossed arrows, Location of change. Circle, Location of no change.

bias against motion directions that require saccades into the inactivated location (affected quadrant). To avoid the confounding effects of this response bias in our analyses, we included trials in which neither the cued signal nor the foil motion patches pointed into the affected quadrant, as in the original study (Lovejoy and Krauzlis, 2010). To model the effects of SC inactivation, we extended the 8-ADC model to include three other parameters: one that described the ratio of sensitivity at the inactivated location to the preinactivation baseline ( $d_{IN}/d_{PRE}$ ), one that described the ratio of sensitivity at the non-inactivated location to the preinactivation baseline ( $d_{OUT}/d_{PRE}$ ), and one that described the difference in criteria between the inactivated location and the noninactivated location ( $c_{IN} - c_{OUT}$ ); as discussed previously, for a task with no NoGo responses, the relative values, rather than the absolute values of the criteria, are relevant for model fitting (Sridharan et al., 2014b). We term this extended model the “8-ADC-inact” model. With the original nine parameters of the 8-ADC model, and the three additional parameters for modeling the effects of inactivation, the 8-ADC-inact model incorporated 12 parameters.

We fit the entirety of the three  $12 \times 3$  contingency tables with the 8-ADC-inact model. To limit the domain of parameter search, we im-

posed the constraints that sensitivity at the foil location could be no higher than that at the target (cued) location and that the sensitivity at the inactivated location could not be higher than sensitivity at baseline ( $d_{IN}/d_{PRE} \leq 1$ ). MLE was used for parameter estimation, as described in the Materials and Methods, Model fitting, with the difference that constrained optimization was employed here. Goodness-of-fit testing was performed by pooling the contingency table across target and foil motion directions separately for each condition (preinactivation baseline, target in the inactivation zone, and foil in the inactivation zone). Error bars for the parameters (Fig. 10) were estimated based on a jackknife procedure by leaving out one session at a time in the estimation.

To assess whether the primary outcome of SC inactivation was a change in sensitivity versus a change in bias, we developed two other models by modifying the 8-ADC-inact model. In the first model, 8-ADC-inact<sub>sens-change</sub>, only sensitivities were allowed to change with inactivation, and criterion changes were not permitted ( $c_{IN} = c_{OUT} = c_{PRE}$ ). This resulted in a 11-parameter model. On the other hand, in the second model, 8-ADC-inact<sub>crit-change</sub>, only criteria were allowed to change with inactivation, and sensitivity changes were not permitted ( $d_{IN} = d_{OUT} =$

$d_{PRE}$ ). This resulted in a 10-parameter model. Both models were fit with the same MLE procedure described above. Comparison of these simpler models with the 8-ADC-inact model was performed based on the AIC and BIC, as described in Materials and Methods, Model comparison.

*Two-ADC model fitting of the cued motion direction-change detection (filtering) task with SC inactivation.* The task (Zénon and Krauzlis, 2012) involved detecting a change in either the target or foil motion patches or indicating that neither location changed, and the animals indicated having detected a change with a button press (Yes/No response). This produced a  $3 \times 2$  stimulus–response contingency table. A full  $3 \times 3$  contingency table was not obtained: the animal’s report did not distinguish between changes perceived at the target location and changes perceived at the foil location. Hence, behavior in this task could not be readily fit with a 2-ADC model: whereas there are only three independent observations in a  $3 \times 2$  contingency table, the 2-ADC model requires four parameters—two sensitivities and two criteria, one at each location—to be estimated.

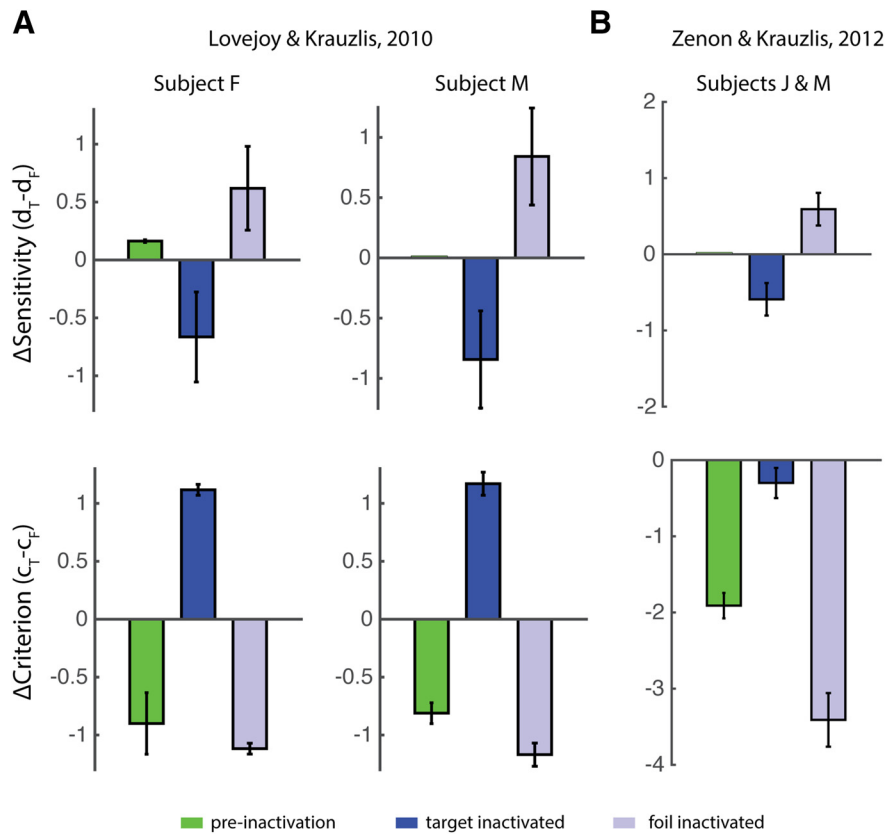
Nevertheless, as with the previous task, SC inactivation produced an additional set of two  $3 \times 2$  contingency tables, one each for when the target or foil was presented in the inactivated zone, respectively. Thus, there were now nine independent observations. We extended the 2-ADC model by incorporating the following additional four parameters: two parameters for modeling the change in sensitivity at the inactivated and noninactivated locations relative to baseline ( $d_{IN}/d_{PRE}$ ,  $d_{OUT}/d_{PRE}$ ) and two more parameters for modeling the change in criteria following inactivation ( $c_{IN}/c_{PRE}$ ,  $c_{OUT}/c_{PRE}$ ). We term this extended model the “2-ADC-inact” model. This eight-parameter model (four baseline parameters and four inactivation parameters) could now be fit to the data from the three  $3 \times 2$  contingency tables, taken together.

As with the previous analysis, we fit two reduced models: a 2-ADC-inact<sub>sens-change</sub> model, in which only sensitivities were allowed to change following inactivation ( $c_{IN} = c_{OUT} = c_{PRE}$ ), and a 2-ADC-inact<sub>crit-change</sub> model, in which only criteria were allowed to change following inactivation ( $d_{IN} = d_{OUT} = d_{PRE}$ ). Both of these models had two fewer parameters than the original 2-ADC-inact model.

Procedures for model fitting (ML parameter estimation), goodness-of-fit testing, parameter error bars, and model comparisons were the same as those for the 8-ADC-inact model, described above.

## Results

Does the SC contribute to attention by altering perceptual sensitivity, choice bias, or both? We develop the answer in three stages. In the first section, we introduce a multidimensional signal-detection model and explain how, in the model’s decision space, performance changes indicative of changes in sensitivity can, instead, be produced by changes in bias alone. In the second section, we validate the model by showing that the model accurately describes the behavior of animals engaged in different kinds of visuospatial multialternative tasks. In addition, we show that the model outperforms other candidate models in explaining behavioral results in these tasks. In the third section, we re-examine behavioral data from four seminal studies in the context of our



**Figure 10.** mADC model parameters fit to SC inactivation data (Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012). **A**, Top, Differences in sensitivity between the target and foil locations before SC inactivation (green) and during inactivation of the target (dark blue) or foil (light blue) spatial representations. Bottom, Differences in criteria between the target and foil locations. Other conventions are the same as in the top panel. Left, Data for subject F. Right, Data for subject M. Parameter estimates are based on fitting an 8-ADC-inact model to SC inactivation data from Lovejoy and Krauzlis (2010). Error bars denote SEs (jackknife) across experimental sessions ( $n = 7$  for subject F;  $n = 4$  for subject M). **B**, Same as in **A**, but based on fitting a 2-ADC-inact model to SC inactivation data from Zénon and Krauzlis (2012). Data were pooled across subjects J and M ( $n = 15$  experimental sessions).

model to identify the precise nature of the effects of SC manipulation (microstimulation or inactivation) on attention.

## A multidimensional model for decoupling perceptual sensitivity from choice bias in multialternative tasks

### Model description

We present a normative, signal-detection model that enables quantifying sensitivity and bias from behavioral responses in multialternative detection and change-detection (m-ADC) tasks. In our m-ADC model, biases represent a preferential gating of sensory evidence in the process of perceptual decision making. These biases manifest in behavior as tendencies to favor certain choices over others (Gold and Ding, 2013), and we term these “choice biases.” These could arise from perceptual or decisional processes (Mulder et al., 2012; Jogan and Stocker, 2014; Vintch and Gardner, 2014; White and Poldrack, 2014). We distinguish these from an alternative category of bias that can arise from motoric processes associated with producing an overt response: we term these “motoric response biases” (García-Pérez and Alcalá-Quintana, 2013). While our model makes it possible to distinguish the contributions of choice bias from those of perceptual sensitivity, the contribution of motoric response biases can be accounted for with an appropriate task design (Zénon and Krauzlis, 2012; García-Pérez and Alcalá-Quintana, 2013).

The m-ADC model applies to tasks that require a subject to detect a target stimulus or a change in a target stimulus at one

among multiple possible locations (Fig. 1A,E) or to detect one among multiple possible values of a stimulus feature (e.g., one among several shapes or colors) at a given location. Consider, for example, a two-alternative detection (2-ADC) task (Fig. 1A). The 2-ADC task involves reporting the appearance of a briefly presented target stimulus (positive contrast dot) that can appear at one of two possible locations or not at all. The subject reports detection of the target by making a saccadic eye movement to the response box at the corresponding location. In each trial, one of three stimulus events can occur (Fig. 1A, middle column): a target stimulus can be presented at location 1, at location 2, or not at all (“catch” trial). For each stimulus event, there are three potential response types (Fig. 1A, right column): a response to location 1, to location 2, or maintain fixation. The first two responses are termed “Go” responses, and the last is termed a “NoGo” response. Conventionally, subjects are rewarded for making accurate Go responses on stimulus trials, and for making NoGo responses on catch trials (Cavanaugh and Wurtz, 2004; Cohen and Maunsell, 2009; Zénon and Krauzlis, 2012; Steinmetz and Moore, 2014). A task that incorporates a NoGo response is also referred to as an unforced-choice task, or a nonforced-choice task (García-Pérez and Alcalá-Quintana, 2011).

A previous m-ADC model was developed to analyze behavior in such multialternative tasks that include catch trials and NoGo responses (Sridharan et al., 2014b). However, the optimal decision strategy (decision rule) and definition of choice bias specified in the previous model were limited to scenarios in which perceptual sensitivities to the various stimulus events (Fig. 1A, targets at location 1 vs location 2) were equal. This condition is frequently not satisfied, particularly in tasks that manipulate attention, as for example with spatial cueing (Herrmann et al., 2010; Carrasco, 2011; Sridharan et al., 2014a; Luo and Maunsell, 2015). Here, we extend the m-ADC model to incorporate an optimal decision rule for this more general scenario in which perceptual sensitivity to stimuli (or stimulus features) at different locations can be significantly different, and, correspondingly, generalize the definition of choice bias for such scenarios. Below, we provide a verbal description and geometric intuition for the model with a 2-ADC task. The mathematical formulation for the general (multialternative) case is presented in Materials and Methods.

SDT models perceptual decision making as an inherently noisy process. For a binary choice task, such as a 2-AFC task, decisions are based on the value of a univariate Gaussian random variable, the “decision variable,” the mean of which depends on the strength of the sensory evidence and the variance of which arises from the noisiness of the sensory evidence across trials (Green and Swets, 1988). The subject employs a cutoff (“criterion”) value of the decision variable to decide between the two competing hypotheses (e.g., stimulus at location 1 vs location 2).

For the 2-ADC task (Fig. 1A), on the other hand, we model the decision variable as a bivariate Gaussian random variable,  $\Psi$ , in a two-dimensional decision space (Fig. 1B). Each component of the decision variable ( $\Psi_1, \Psi_2$ ) encodes the strength of the sensory evidence at each potential target location. We further assume that the covariance matrix of the decision variable is an identity matrix. Hence, the variance of  $\Psi$  along each dimension is unity, and the two-decision variable components are uncorrelated and independent (see Materials and Methods, Eq. 1).

When no target stimulus is presented (as in catch trials), the  $\Psi$  distribution is a unit variance Gaussian centered at the origin; we denote this distribution as  $n$  (Fig. 1B, left). When a target stimulus is presented at a location, the mean of the  $\Psi$  distribution

increases for the decision variable component corresponding to that location (Fig. 1B, center and right). We term the  $\Psi$  distribution for a stimulus at location 1 or 2 as  $s_1$  or  $s_2$ , respectively. The change in the mean value of  $\Psi$ , measured in units of the standard deviation of  $n$ , quantifies the perceptual sensitivity,  $d_i$  ( $d_1$  or  $d_2$ ), for discriminating signal from noise (detecting a stimulus) at that location. Neurally, enhanced sensitivity may be achieved by mechanisms that enhance the gain of neural responses (Reynolds et al., 2000; Williford and Maunsell, 2006; Sridharan and Knudsen, 2015) concomitantly with those that reduce neural noise and variability (Mitchell et al., 2007; Churchland et al., 2010). Behaviorally, enhanced sensitivity at a location manifests as an enhanced ability to accurately detect the target stimulus at that location.

In our model, linear (or, in general, planar) decision surfaces partition the two-dimensional (multidimensional) decision space into nonoverlapping domains (Fig. 1D, shaded; see Materials and Methods, Eq. 19). On each trial, a response is made to the decision domain corresponding to the value of  $\Psi$  on that trial. In this 2-ADC model, the family of optimal decision surfaces in two-dimensional decision space represents the surfaces of constant posterior odds ratio for each pair of alternatives ( $s_1$  vs  $n$ ,  $s_2$  vs  $n$ , and  $s_1$  vs  $s_2$ ; see Materials and Methods, Eqs. 11 and 16). Because the joint probability distributions are Gaussians with equal variance (circularly symmetric), the decision surfaces, representing constant posterior odds, are straight lines. A family of such optimal decision surfaces, corresponding to the  $\Psi$  distributions in Figure 1B, is shown in Figure 1C. In this formulation, detection sensitivities need not be equal across locations ( $d_1 \neq d_2$ ; Fig. 1B, middle vs right). These decision domains represent zones in which the Bayesian posterior probability of each stimulus event, given the sensory evidence, is greatest (see Materials and Methods).

The family of optimal decision surfaces in the 2-ADC model is parameterized by two choice criteria,  $c_1$  ( $s_1$  vs  $n$  family; Fig. 1C, left) and  $c_2$  ( $s_2$  vs  $n$  family; Fig. 1C, middle). The third decision surface is automatically determined by the intersection of the other two surfaces and the relative values of  $d_1$  and  $d_2$  ( $s_1$  vs  $s_2$  family; Fig. 1C, right).

The model quantifies bias based on the relative values of these choice criteria. Intuitively, the value of the criterion at a location (relative to the other locations) is inversely related to the magnitude of the bias toward selecting that location. For example, decreasing the value of  $c_1$  results in a greater chance that location 1 is chosen for response, even on trials in which the strength of sensory evidence (magnitude of  $\Psi_1$  and  $\Psi_2$ ) is identical across the two locations (Fig. 1D, point marked X on diagonal line of  $\Psi_1 = \Psi_2$ ). This is equivalent to affording greater weight to sensory evidence from location 1 in the final decision. Neurally, choice bias toward a location may be achieved by increasing the gain of synaptic input to the downstream neurons that read-out activity at that location, although identifying neural mechanisms that specifically alter choice bias remains an open question (Eckstein et al., 2009). Behaviorally, choice bias for a location manifests as an increased probability for selecting that location for decision making and response.

The intuitive definition of choice bias, as being inversely related to choice criterion, is useful only when the sensitivities ( $d$  values) at the two locations are equal. This is because the value of the optimal choice criterion at a location depends on the sensitivity to a stimulus at that location: in the absence of other influences, the optimal criterion would lie at the midpoint of the signal and noise distributions ( $c_i = d_i/2$ ). In Materials and Methods,



section titled “Relationship between choice criterion, choice bias, priors, and payoffs”, we provide a definition of choice bias, based on a likelihood ratio metric that discounts the contribution of differential sensitivity. The resulting metric of choice bias remains monotonically (inversely) related to choice criterion and its numerical value depends on the difference between the actual criterion and the optimal criterion estimated in the absence of differences in priors or payoffs (see Materials and Methods, Eq. 28). We analytically demonstrate that, for optimal decisions (those that seek to maximize success), this definition of choice bias is directly related to the relative magnitudes of the priors (the prior probability of stimulus events) and the payoffs (the relative costs/benefits for each stimulus–response contingency; see Materials and Methods, Eq. 28).

Figure 1D summarizes decision variable distributions and decision surfaces for all 2-ADC stimulus and response contingencies. For each stimulus event, the integral of  $\Psi$  in each decision region represents the probability of choosing the response associated with that region. In the Materials and Methods (Eqs. 23 and 24), we derive this analytical relationship, linking the response probabilities for each stimulus event to the perceptual sensitivity and choice criterion for the extended m-ADC model.

#### *The effects of spatial cueing in a multialternative task*

Studies seeking to understand the neural bases of spatial attention have increasingly used multialternative tasks (Cavanaugh and Wurtz, 2004; Cohen and Maunsell, 2009; Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012; Steinmetz and Moore, 2014; Luo and Maunsell, 2015). In these studies, the effects of attention or of neural manipulations (microstimulation or inactivation) are typically quantified as changes in percentage correct responses (hits), changes in the percentage of incorrect responses (false alarms), and/or changes in the slope of the psychometric function. An increase in percentage correct, decrease in false alarms, and/or increase in the slope of the psychometric function induced by cueing or microstimulation are commonly interpreted as representing an improvement in the animal’s ability to detect target stimuli (an improvement in perceptual sensitivity) at the cued or microstimulated location. The converse effects (decrease in percentage correct, increase in false alarms, or decrease in the slope of the psychometric function) that result from inactivation are assumed to represent an impairment in the animal’s ability to detect targets (reduction of perceptual sensitivity) at the inactivated location. Here we test the validity of these interpretations, based on m-ADC model simulations and ideal observer analysis.

We consider a general class of task designs in which the animal is cued to attend to one of multiple ( $\geq 2$ ) potential target locations on every trial either explicitly, with a spatial cue, or implicitly, by increasing the probability of occurrence of the target stimulus across a block of trials at one of the potential locations. In both cases, the likely location of the target stimulus (or change) is determined well before the target stimulus (or change) occurs. A specific example of this class of tasks, an orientation-change detection task, is shown in Figure 1E. In this task, the animal is rewarded for correctly reporting a change in orientation of a stimulus at any location (cued or uncued), and for not responding when no change occurs anywhere (catch). Similar task designs have been widely used to measure the effect of attention on neural information processing (Williford and Maunsell, 2006; Cohen and Maunsell, 2009; Steinmetz and Moore, 2014; Luo and Maunsell, 2015).

We simulated the animal’s behavior in this task with a 2-ADC model. In our simulations, 90% of trials were change trials and

10% were catch trials, and changes were four times more likely to occur at the cued location, mimicking the proportions employed in the experimental studies (Williford and Maunsell, 2006; Cohen and Maunsell, 2009; Steinmetz and Moore, 2014). To test the hypothesis that changes in sensitivity are not required to account for changes in percentage correct performance, the modeled  $d$  values for the two locations (cued, uncued) were held equal (see Materials and Methods). In this way, changes in choice bias can be directly estimated by comparing the values of the choice criteria.

An ideal observer analysis maximized percentage correct globally across all trial types (change and catch trials) with the criteria ( $c_1$  and  $c_2$ ) as free parameters. To limit the number of free parameters, we assumed that the animal maintained a constant, low proportion of false alarms (0.1) on catch trials, a plausible assumption for such tasks (see Materials and Methods).

The simulation results are shown in Figure 1F,G. To maximize percentage correct, the ideal observer analysis automatically developed a bias toward the cued (high probability) location: the criterion at the cued location ( $c_{\text{cued}} = 1.37$ ) was about a third lower than the criterion at the uncued location ( $c_{\text{uncued}} = 2.14$ ). We plotted percentage correct as function of the various simulated  $d$  values at the cued and uncued locations (Fig. 1F). This simulation mimics the experimental scenario in which target stimuli of equal perceptual strength occur at the two locations. The consequence of maximizing overall success was a consistently higher percentage correct performance (by  $\leq 30\%$ ) at the cued, relative to the uncued location for identical values of  $d$ . In addition, we plotted percentage correct as a function of orientation-change magnitude assuming a hyperbolic ratio (Naka–Rushton) relationship between perceptual sensitivity and stimulus physical strength (here, change magnitude; see Materials and Methods; Herrmann et al., 2010; Sridharan et al., 2014a). The resulting psychometric performance curves showed a steeper slope (for orientation changes from 1 to 10°) at the cued, relative to the uncued locations (Fig. 1G).

These results demonstrate that, when data are gathered with this kind of task design, a significant improvement in percentage correct performance at the cued relative to the uncued location does not necessarily indicate an increase in perceptual sensitivity, and that changes in the slope of the psychometric performance curve is an unreliable indicator for changes of perceptual sensitivity. Rather, a change in choice bias alone is a robust strategy for increasing percentage correct at the cued location. This example highlights the importance of distinguishing bias from sensitivity effects with normative models, such as the m-ADC model, in the search for mechanisms by which attention influences neural information processing.

#### **Describing and predicting behavioral results with the m-ADC model**

The m-ADC model, for multialternative tasks, is necessarily more complex than one-dimensional binary choice signal-detection models. We evaluated the ability of the model to describe and predict behavioral responses measured in various multialternative tasks: two detection tasks and an attention task. These measurements were made in two classes of animals: birds (chickens, *Gallus domesticus*) and primates (macaques, *Macaca mulatta*).

#### *Four-alternative spatial target detection task*

Chickens performed a four-alternative target detection task, in which they had to detect and localize a brief target stimulus (pos-

**Table 2. Model fits and parameters for 4-ADC target detection task: MLEs of bias (pooled data)**

MLE (mean $\pm$ SE)
$b_1 = 0.020 \pm 0.009$
$b_2 = 0.024 \pm 0.011$
$b_3 = 0.289 \pm 0.068$
$b_4 = 0.167 \pm 0.039$

Location 1: upper right; 2: upper left; 3: lower left; 4: lower right.

itive contrast dot) by pecking at its location in one of the four quadrants of the visual field (Fig. 2A). In addition, if no target was detected, the birds could give a NoGo response by pecking once again on the zeroing cross (Sridharan et al., 2013).

In this task, the target was presented in 50% of the trials, with equal probability of appearance at each location ( $\sim 12.5\%$ ), and in the remaining 50% of the trials no target was presented (catch trials). The data shown in Figure 2B document, for each contingency (row), the number of responses made to each location (columns 1–4) as well as NoGo responses (column 5,  $\phi$ ; data pooled across  $n = 3$  birds;  $\sim 11,000$  trials in 108 experimental sessions). The diagonal entries of the contingency table represent correct responses (hits and correct rejections) for that stimulus event, whereas the off-diagonal terms represent incorrect responses (misses, false alarms, and mislocalizations; Sridharan et al., 2014b).

Examining the relative proportions of correct responses (hits) across locations revealed a substantially higher percentage correct for targets presented in the lower-left quadrant (location 3, 79% correct), relative to targets presented in the other quadrants (42, 43, or 54% correct). Do these results indicate a higher perceptual sensitivity to detect target stimuli in the lower-left quadrant?

To answer this question, we fit the data from this task with the m-ADC model. The data were well fit by the model ( $\chi^2_{(12,10944)} = 7.11$ ,  $p_r > 0.99$ , randomization test; MLE and Bayesian estimation; see Materials and Methods). Furthermore, the estimation procedure revealed that choice criteria were lowest for targets in the lower hemifield ( $c_3: 1.77 \pm 0.025$ ;  $c_4: 1.90 \pm 0.029$ ) compared with those in the upper hemifield ( $c_1: 2.80 \pm 0.074$ ;  $c_2: 2.73 \pm 0.068$ ; Fig. 2D, top; Table 1). These differences in choice criteria manifested in behavior as a high proportion of incorrect (mislocalization) responses as well as false alarms on catch trials directed to the lower hemifield locations (Fig. 2B, third and fourth columns of the contingency table), relative to the upper hemifield locations. Quantification of choice bias indicated a substantially higher bias toward lower hemifield locations compared with the upper hemifield locations (Table 2).

Having accounted for the choice biases, model estimates revealed that sensitivity to target stimuli in the lower-left quadrant (location 3:  $d_3 = 2.58 \pm 0.05$ ) was, in fact, no different from the sensitivities to stimuli in the upper hemifield quadrants ( $d_1 = 2.62 \pm 0.08$ ,  $d_2 = 2.58 \pm 0.08$ ; Fig. 2D, bottom; Table 1).

Next, we tested the ability of the model to predict the animals' behavior. We predicted response probabilities and counts for 17 contingencies (Fig. 2E, cells within highlighted square plus the bottom right cell) based on model parameters estimated using data from eight contingencies alone (misses and false alarms; Fig. 2B, last row and column of table). Predicted response probabilities (Fig. 2E, cell shading) and counts (numbers within each cell) closely matched the observed response probabilities and counts (Fig. 2B) with a  $\chi^2$  goodness of fit of  $\chi^2 = 6.61$  ( $p_r > 0.99$ ). In

**Table 3. Model fits (goodness-of-fit statistics) for the target detection task, for individual birds**

Bird identity	All contingencies		8 of 25 contingencies	
	$\chi^2$	$p_r(\chi^2)$	$\chi^2$	$p_r(\chi^2)$
Bird 1	14.90	0.717	19.18	0.467
Bird 2	24.14	0.234	98.10 (11.52)	0.002* (0.99)
Bird 3	11.01	0.945	14.68	0.786

\* $\chi^2$  Values for model fits are inflated when several cells in the contingency table contain low ( $< 5$ ) response counts, resulting in apparently poor fits (small  $p$ ).  $\chi^2$  Statistics and  $p$  values computed with aggregated counts across non-target locations for each stimulus contingency (see Materials and Methods) produced acceptable fits (values in parentheses).

addition, parameters estimated with this subset of data closely matched the estimates obtained by fitting the full contingency table.

Data from each bird analyzed separately were also fit well by the model (Table 3), and parameter estimates indicated that choice criteria were generally lower for the bottom hemifield quadrants (Fig. 2F). Finally, predicted response probabilities closely matched observed response probabilities when data were analyzed separately for each bird (Fig. 2G; Table 3).

The results indicate that, despite its simplifying assumptions, the m-ADC model accurately accounted for the observed behavioral data in this four-alternative target detection task.

In this task, the animals reported the occurrence of a target stimulus at a location with an overt movement (peck) toward the same (respective) location. Hence, it is possible that a motoric response bias toward the bottom hemifield targets contributed to these asymmetries, although the provision for a NoGo (undecided) response substantially reduces the chances of conflating response bias with choice bias, as has been demonstrated previously (García-Pérez and Alcalá-Quintana, 2011, 2013; Sridharan et al., 2014b).

#### Four-alternative, orientation-change detection task

A monkey performed a four-alternative, orientation-change detection task that involved detecting and localizing a change in orientation of one of four grating stimuli presented in each of the four visual quadrants (Fig. 3A); the task was made more demanding by the introduction of a blank period before change events (change blindness). A key difference from the previous task was in the response protocol: the animal reported the location of change with an antisaccade, a saccadic eye movement to the stimulus diagonally opposite to the location of change. This task design makes it possible to distinguish the effects of spatial attention at a location from those related to saccade preparation (Steinmetz and Moore, 2014).

The considerable difficulty of the task (change blindness) and the complex response rule (antisaccade) resulted in a high proportion of incorrect responses (Table 4), even for large orientation changes (Fig. 3G; change magnitudes,  $\sim 90^\circ$ ). With these "lapse rates" incorporated (see Materials and Methods), the m-ADC model fit and predicted the monkey's behavior accurately.

Parameters estimated by pooling trials across all orientation change magnitudes ( $10\text{--}90^\circ$ ; Fig. 3B;  $n = 22$  experimental sessions) revealed that choice criteria (Fig. 3C, top) and choice biases (Table 6) were not different across locations. With sensitivity and criterion parameters estimated using data from only 8 of 25 contingencies (misses and false alarms; Fig. 3B, last row and column of table), the model was able to predict responses (counts and probabilities) for the remaining 17 contingencies accurately

**Table 4. AP matrix for uncued orientation change detection task**

Change (C)	Response (R)			
	1	2	3	4
1	0.000 <sup>b</sup>	0.143	0.714 <sup>a</sup>	0.143
2	0.038	0.019 <sup>b</sup>	0.057	0.887 <sup>a</sup>
3	0.931 <sup>a</sup>	0.035	0.000 <sup>b</sup>	0.035
4	0.050	0.775 <sup>a</sup>	0.175	0.000 <sup>b</sup>

Each entry corresponds to the proportion of Go responses at a location R (columns) when a salient change (largest rotation magnitudes, 70–90°, and the shortest flash durations, <200 ms) were presented at location C.

<sup>a</sup>Proportion correct (hit rates), corresponding to proportion of Go responses to the location opposite to change (antisaccade task).

<sup>b</sup>Proportion of Go responses to change location. The low probabilities in the other entries indicated that the monkey had learned the antisaccade rule, but some confusion persisted when reporting the change, indicated by the responses to the locations adjacent to the response location (italicized cells).

**Table 5. AP matrix for cued orientation change detection task (100% valid cues)**

Change (C)	Response (R)			
	1	2	3	4
1	0.005	0.065	0.859 <sup>a</sup>	0.071
2	0.045	0.000	0.045	0.910 <sup>a</sup>
3	0.921 <sup>a</sup>	0.052	0.000	0.026
4	0.039	0.945 <sup>a</sup>	0.006	0.011

<sup>a</sup>Upon cueing (100% valid), asymptotic performance (for the most salient changes, as above) improved significantly (>90% correct, on average). Other conventions are the same as in Table 4.

(Fig. 3D,E, highlighted square;  $\chi^2_{(12,5416)} = 25.97$ ,  $p_r = 0.17$ , randomization test).

Next, we estimated psychophysical functions by estimating sensitivities for nonoverlapping bins of orientation change magnitudes, based on the miss rates for each orientation change bin and the overall false-alarm rates (Fig. 3F; see Materials and Methods). Psychometric functions of percentage correct (hits) and incorrect responses (misidentifications) predicted by the m-ADC model (Fig. 3G, inverted triangles) closely matched experimentally observed psychometric functions across all orientation change values (Fig. 3G, filled circles).

Thus, as before, the m-ADC model accurately accounted for the observed behavioral data in this four-alternative change detection task. The results indicated that, in this task, the monkey did not exhibit a choice bias toward any specific location.

The multialternative tasks presented so far provided the subjects with no prior information regarding the location of the impending target stimulus. In visuospatial attention tasks, on the other hand, the likely location of the target is indicated beforehand with an informative spatial cue, which improves performance by inducing a shift of attention to the cued location (Carrasco, 2011). The relative contributions of sensitivity and bias mechanisms to the effects of spatial cueing remain unknown. We sought to answer this question with our m-ADC model.

#### Four-alternative, cued orientation-change detection task

The monkey performed the same orientation-change detection task described in the previous section, except that on each trial a cue (Fig. 4A, directed line segment close to fixation spot), indicating the location of the upcoming change with ~93% validity, appeared shortly after stimulus onset (Steinmetz and Moore, 2014). The superior task performance enabled by the cue obviated the need for accounting for lapse rates in this task (Table 5; see Materials and Methods).

Estimates of sensitivities and criteria, based on behavioral responses pooled across cue locations (Fig. 4B; see Materials and Methods) are shown in Table 7 and Figure 4C. The choice criterion was lowest ( $c_C$ ,  $0.13 \pm 0.020$ ), and choice bias highest ( $b_C$ ,  $1.477 \pm 0.293$ ), for detection at the cued location. On the other

**Table 6. Model fits and parameter estimates for uncued four-alternative orientation-change detection task**

Parameter	MLE (mean $\pm$ SE)
Sensitivity	$d_1 = 1.60 \pm 0.19$
	$d_2 = 0.87 \pm 0.18$
	$d_3 = 1.02 \pm 0.05$
	$d_4 = 0.82 \pm 0.20$
Criterion	$c_1 = 2.17 \pm 0.06$
	$c_2 = 1.71 \pm 0.04$
	$c_3 = 0.13 \pm 0.02$
	$c_4 = 1.44 \pm 0.03$
Bias	$b_1 = 0.310 \pm 0.072$
	$b_2 = 0.357 \pm 0.080$
	$b_3 = 0.343 \pm 0.077$
	$b_4 = 0.278 \pm 0.066$

Location 1, Upper right; location 2, upper left; location 3, lower left; location 4, lower right.

**Table 7. Model fits and parameter estimates for cued four-alternative orientation-change detection task**

Parameter	MLE (mean $\pm$ SE)
Sensitivity	$d_R = 1.60 \pm 0.19$
	$d_O = 0.87 \pm 0.18$
	$d_C = 1.02 \pm 0.05$
	$d_S = 0.82 \pm 0.20$
	$d_{\text{diag}} = 0.82 \pm 0.20$
Criterion	$c_R = 2.17 \pm 0.06$
	$c_O = 1.71 \pm 0.04$
	$c_C = 0.13 \pm 0.02$
	$c_S = 1.44 \pm 0.03$
	$c_{\text{diag}} = 1.44 \pm 0.03$
Bias	$b_R = 0.112 \pm 0.047$
	$b_O = 0.330 \pm 0.138$
	$b_C = 1.477 \pm 0.293$
	$b_S = 0.427 \pm 0.157$
	$b_{\text{diag}} = 0.427 \pm 0.157$

C, Cued location; R, location opposite cued location; S, same hemifield as cued location; O, diagonally opposite cued location.

**Table 8. Model fits (goodness-of-fit statistic) for individual cued positions**

Cue location	All contingencies		8 of 25 contingencies	
	$\chi^2$	$p_r(\chi^2)$	$\chi^2$	$p_r(\chi^2)$
Combined	23.31	0.28	29.30	0.093
Cue: location 1	29.08 (4.96)	0.127 (0.99)	79.96 (31.51)	0.003* (0.99)
Cue: location 2	17.45 (0.29)	0.618 (0.99)	21.11 (2.15)	0.395 (0.99)
Cue: location 3	56.29 (7.51)	0.007* (0.99)	83.32 (22.74)	0.000* (0.99)
Cue: location 4	18.39 (0.95)	0.555 (0.99)	25.99 (6.44)	0.208 (0.99)

\*Poor fits due to inflated  $\chi^2$  values. Values in parentheses,  $\chi^2$  statistic and p values computed with counts aggregated across no-change response locations for each stimulus contingency (see Table 3, Materials and Methods).

hand, the choice criterion was the highest ( $c_R$ ,  $2.17 \pm 0.060$ ) and choice bias lowest ( $b_R$ ,  $0.112 \pm 0.047$ ) for detection at the diagonally opposite location. Criteria and biases were of intermediate values at the other locations ( $b_S$ ,  $0.427 \pm 0.157$ ;  $b_O$ ,  $0.330 \pm 0.138$ ;  $c_S$ ,  $1.71 \pm 0.040$ ;  $c_O$ ,  $1.44 \pm 0.033$ ). The results indicate a strong cue-induced increase in choice bias relative to the uncued task (compare Tables 6, 7; see Discussion). In addition, sensitivity estimates (Fig. 4C, bottom) were nominally the highest at the cued location and response location ( $d_C$ ,  $1.02 \pm 0.05$ ;  $d_R$ ,  $1.60 \pm 0.19$ ), relative to the other locations ( $d_O$ ,  $0.87 \pm 0.18$ ;  $d_S$ ,  $0.82 \pm 0.20$ ), although these differences were not significant.

As before, the model fit the monkey's behavioral responses accurately in this task ( $\chi^2_{(12,4986)} = 23.31$ ,  $p_r = 0.28$ , randomization test,  $n = 12$  experimental sessions). In addition, the model



was able to accurately predict the monkey's responses to both validly and invalidly cued locations with parameters derived from only the misses and false alarms (Fig. 4D;  $\chi^2 = 29.30$ ,  $p_r = 0.093$ , randomization test). The model continued to both fit (Table 8) and predict (Fig. 4E) responses robustly when data for each cue location were considered separately.

We asked whether, in this task, the monkey adjusted its choice criterion at the cued location (Fig. 4C) following a rational decision policy. As shown in Materials and Methods (Eq. 28), the optimal value of choice bias varies with the prior odds ratio of a change at a location versus no change ( $p_X/p_0$ ). In both the uncued (Fig. 3A) and the cued (Fig. 4A) tasks, the probabilities of no change on any trial ( $p_0$ ) were identical (50%). Thus, any difference in decisional policy ( $p_X$ ) between the cued and uncued tasks would reflect a difference in prior probabilities of change between these tasks. In the uncued change detection task, changes were equally likely (25%) at all locations. Thus, the conditional probability of change at any location  $p_{X-UC}$  was 0.25. The monkey did not show a bias toward any particular location, and choice criteria (Fig. 3C) and biases were similar across locations (Table 6; median  $b_{UC} = 0.327$ ). On the other hand, in the cued change detection task (Fig. 4A), 93% of changes occurred at the cued location ( $p_{X-C} = 0.93$ ). For optimal behavior, the value of the choice bias at the cued location (Table 7) should reflect the change in the prior odds ratio across the two tasks. Indeed, the choice bias at the cued location estimated from the animal's behavior with the m-ADC model ( $b_C = 1.48$ ) was numerically similar to that estimated from optimal decision analysis ( $b_{UC} \times (p_{X-C}/p_{X-UC}) = 1.22$ ). Thus, the m-ADC model analysis indicated a rational, albeit slightly suboptimal, shift in the animal's choice bias favoring the cued location.

The results demonstrate the ability of the model to explain and predict behavior in this spatial cueing task. The model revealed that, with spatial cueing, the monkey developed a strong, near-optimal bias to report changes at the cued location. Having controlled for these bias effects, the model provided an interesting insight: the animal's sensitivity (Fig. 4C, bottom) was numerically highest at the location of the impending saccade, indicating that although the animal was cued to attend to one location, the planning of a motor response to the opposite location may have caused some shift of attention to that opposite location.

#### Comparison of the m-ADC model with other candidate models

With eight parameters (one sensitivity and one criterion for each of the four locations) the 4-ADC model was able to accurately describe behavior in each of these tasks. Can a simpler model, with fewer parameters, explain the data equally well? Alternatively, do more complex models, with additional parameters, explain the data better?

We compared 4-ADC model fits against those of three other candidate models (Fig. 5): (1) a five-parameter 4-ADC eq crit model, a signal-detection model that assumes equal criterion values at all locations (Fig. 5C,D; Macmillan and Creelman, 2005); (2) a five-parameter 4-ADC eq sens model, a signal-detection model that incorporates unequal criteria, but assumes equal sensitivity values at all locations (Fig. 5E,F); and (3) a nine-parameter 5-AFC model that treats NoGo responses as a distinct category of Go responses (Fig. 5G,H; Macmillan and Creelman, 2005). Each of these models was tested for its ability to fit responses from the four-alternative target detection task (chickens; Fig. 2A) and the four-alternative, cued orientation-change detection task (monkey; Fig. 4A). We did not attempt to fit the data from the four-alternative orientation-change detection task

**Table 9. Model comparison based on AIC and BIC: four-alternative target detection task (chickens)**

Model	$n_p$	df	$\chi^2$	$p'(p_{\chi^2})$	AIC	BIC	$\Delta$ AIC	$\Delta$ BIC
4ADC	8	12	7.11	0.99 (0.85)	12222	12280	0	0
4ADC eq crit	5	15	415.7	0.00 (0.00)	12693	12730	471	450
4ADC eq sens	5	15	111.04	0.00 (0.00)	12318	12354	96	74
5AFC	9	11	16.22	0.69 (0.13)	12229	12295	7	15

$n_p$ , Number of parameters; df, degrees of freedom;  $p_r$ ,  $p$  value (randomization test);  $p_{\chi^2}$ ,  $p$  value  $\chi^2$  test;  $\Delta$ AIC, AIC – AIC<sub>min</sub>;  $\Delta$ BIC, BIC – BIC<sub>min</sub>.

**Table 10. Model comparison based on AIC and BIC: four-alternative cued orientation-change detection task (monkey)**

Model	$n_p$	df	$\chi^2$	$p'(p_{\chi^2})$	AIC	BIC	$\Delta$ AIC	$\Delta$ BIC
4ADC	8	12	23.31	0.28 (0.03)	9422	9474	0	0
4-ADC eq crit	5	15	3864.6	0.00 (0.00)	12625	12658	3203	3184
4-ADC eq sens	5	15	33.73	0.04 (0.00)	9425	9458	3	–16
5AFC	9	11	21.09	0.40 (0.03)	9425	9484	3	10

Conventions are as in Table 9.

(monkey; Fig. 3A) because of the additional complexity involved with modeling lapse rates for this task.

The models differed widely in their ability to fit the behavioral data (Tables 9, 10). The 4-ADC and 5-AFC models could successfully account for responses in both tasks ( $\chi^2$  goodness-of-fit statistic, randomization test; Tables 9, 10). On the other hand, the 4-ADC eq crit and 4-ADC eq sens models could not account for responses in either task (randomization test). However, for both tasks, the 4-ADC eq sens model accounted for the observations better than the 4-ADC eq crit model; because these two models have the same number of parameters, the model with a lower AIC value indicates a better fit.

Could these differences in ability to fit the behavioral data arise from overfitting by the more complex models due to the additional free parameters? To address this question, we compared the relative performances of the models with two information metrics that penalize more complex models: the BIC and the AIC; the BIC penalizes more severely for the number of parameters (Materials and Methods).

For the target detection task, the 4-ADC model yielded the lowest AIC and BIC values (Table 9). Thus, for this task the 4-ADC model provided the best balance of complexity and goodness of fit. Again, for the cued orientation-change detection task, the 4-ADC model yielded the lowest AIC and BIC scores (Table 10) among the models that fit the data; although the 4-ADC eq sens model had the lowest AIC value, it did not fit these data per the randomization test, as mentioned above. Absolute and relative values of sensitivity and criteria were similar among the 4-ADC and 5-AFC models for both tasks (Fig. 5A,B,E–H), demonstrating that the more parsimonious 4-ADC model was sufficient to provide reliable estimates of these parameters.

In summary, the results indicate that the 4-ADC model provided the best trade-off between complexity and goodness of fit when compared with simpler (4-ADC eq crit, 4-ADC eq sens) or more complex (5-AFC) models. Incorporating differences in criteria into the model was essential to fit, and to correctly interpret, responses in these multialternative tasks: a model that incorporated differences in criteria but with equal sensitivities accounted for the observations better than a model with equal criteria but different sensitivities, indicating that differences in criteria were important for explaining the observed differences in performance across locations.



### The mechanistic role of the SC in spatial attention as revealed by the m-ADC model

Having established the parsimony and validity of the m-ADC model, we applied the model to published behavioral data from four studies that tested the effects of either microstimulation or inactivation of the SC on attention-demanding tasks (Cavanaugh and Wurtz, 2004; Müller et al., 2005; Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012). We sought to determine whether the reported effects (namely, changes in percentage correct and/or false-alarm rates) could have been produced by changes in choice bias (decision strategies) alone without changes in the animal's perceptual sensitivity ( $d$ ).

#### *The effects of SC microstimulation in a spatial, change detection task*

The first of the four studies that we considered employed a "change blindness" motion direction-change detection task (Fig. 6; Cavanaugh and Wurtz, 2004). In this task, either a spatial cue was presented or focal microstimulation was delivered on some trials, but not on others. The cue was presented well before ( $\sim 1$ – $2$  s) the motion direction change occurred, and was fully informative (100% valid) about the location of the upcoming change. Similarly, focal microstimulation of the SC was delivered well before ( $\sim 450$  ms) the motion direction change occurred: although the microstimulation was subthreshold for evoking movement, its effects could provide the animal with advance information indicating, with 100% validity, the location of change. The blanking of the screen (Fig. 6A) served a temporal cue, thereby eliminating temporal uncertainty associated with the onset of the motion direction change (in the original study, data from trials without blanks were eliminated from the analyses).

The effect of the cue or microstimulation was evaluated based on differences in performance between cued and uncued trials, with performance quantified as percentage correct and percentage false-positive responses. The percentage of correct responses (hits) was based on the proportion of trials in which a saccade was made to the location of change as a fraction of the number of trials in which a change occurred in any one of the patches, and the percentage of false positives was based on the proportion of trials in which a saccade was made to a location at which no change had occurred, regardless of whether a change had occurred anywhere else or nowhere (catch). The percentages of false alarms during catch trials and change trials, respectively, were not reported separately. An increase in percentage correct responses on cued (relative to uncued) trials, accompanied by no change, or a reduction, in percentage false-positive responses, was taken as evidence that shifting attention to the cued location counters "change blindness" (Cavanaugh and Wurtz, 2004, their Figs. 2A, 3A). Quantitatively similar increases in percentage correct were observed when the SC was microstimulated (Cavanaugh and Wurtz, 2004, their Fig. 4B). Hence, it has been inferred that microstimulating the SC produces an effect similar to that of a cued shift of attention.

Before we present our simulation results, we note a difference between the experimental task and our simulated task. In the original study (Cavanaugh and Wurtz, 2004), the change in the direction of motion could occur at one of three patch locations, which we could have simulated with a 3-ADC model. To simplify the presentation of the results, however, we simulated a two-alternative version of the task (Fig. 6A), in which changes in motion direction could occur at one of only two patch locations, with a 2-ADC model. The

essential conclusions remain the same, regardless of the number of potential change locations (see Materials and Methods).

In the simulated task, changes occurred with equal probability (50%) at each of the two locations on uncued change trials, whereas all changes (100%) occurred at the cued/microstimulated location on cued/microstimulated change trials. The proportions of trials were also the same as in the original study (65% change and 35% catch trials; Cavanaugh and Wurtz, 2004). To reduce the number of free parameters in the model, we assumed a constant proportion of false alarms on catch trials (0.2). As in the previous example,  $d$  values for target stimuli at both locations and for both cued and uncued trials were held equal.

Figure 6D–F (open symbols) shows the results of ideal observer analysis for uncued/nonmicrostimulated trials analyzed with the 2-ADC model (Fig. 6B). Because the number of change trials and the distribution of  $d$  values were identical across locations, the ideal observer analysis selected a symmetric decision rule with no differences in criteria across locations. The maximum percentage correct (with an average false-positive rate of  $\sim 11\%$ ) increased monotonically with the  $d$  value (Fig. 6D, open circles) and with the magnitude of motion direction change (Fig. 6E, open circles).

Next, we simulated the animal's behavior on the trials with cueing or microstimulation. We matched the percentage of false-positive responses across cued/microstimulated versus uncued trials (Fig. 6F; see Materials and Methods). Ideal observer analysis of cued/microstimulated trials yielded a model with a high bias toward the cued/microstimulated location and away from the uncued/other locations (Fig. 6C; see Materials and Methods, section titled "Optimal two-alternative detection model for 100% valid cues is a two-alternative forced choice (Yes-No) model"). Simulated behavioral results revealed that for identical  $d$  values and similar false-positive rates across cued/microstimulated and uncued/other locations, percentage correct performance was consistently greater on cued/microstimulated trials than on uncued/other trials (Fig. 6D, E, closed vs open circles).

Thus, the effects of spatial cueing, as well as those of SC microstimulation, on task performance could be entirely accounted for by changes in choice bias (including a potential oculomotor response bias) alone without involving changes in perceptual sensitivity.

#### *The effects of SC microstimulation in a perceptual discrimination task*

The second of the four studies employed a two-alternative perceptual discrimination task in which monkeys had to report (with a saccadic eye movement) the direction of motion (left vs right) of a patch of coherently moving dots (Müller et al., 2005). In each experimental session, the moving patch could occur at one of two potential locations symmetrically distributed about the vertical meridian (Fig. 7A). This study demonstrated that electrical microstimulation applied at the target's spatial representation in the SC improves percentage correct in this task. This improvement has been interpreted as demonstrating a causal role for the SC in covert spatial attention (Müller et al., 2005).

Unlike the previously discussed detection paradigm (Cavanaugh and Wurtz, 2004), this study employed a forced-choice discrimination paradigm. Such a paradigm provides an unbiased estimation of the effects of microstimulation on performance, because the spatial microstimulation carries with it no information regarding the correct behavioral response (leftward vs rightward motion), thereby eliminating the contribution of oculomotor response biases to the observed effects on task performance. Nevertheless, could the mi-

crostimulation have induced a spatial choice bias that resulted in the improvement in percentage correct?

To answer this question, we employed a variant of the 2-ADC model for discrimination tasks. The analytical development of the model specifically adapted for this task is presented in Materials and Methods, section titled “Extension of the model to perceptual discrimination tasks.” In this model, the decision to respond to one of the directions of motion is based on sensory evidence at both potential target locations. Hence, in contrast to the one-dimensional decision space of the conventional 2-AFC model (Macmillan and Creelman, 2005), the subject’s decision in the 2-ADC discrimination task is modeled in a two-dimensional decision space (Fig. 7B). In this model, the subject gives a Go response indicating the motion direction with the greatest strength of sensory evidence. However, if the sensory evidence for motion in either direction falls below the criteria at both locations (Fig. 7B,C, gray zone), the subject gives a NoGo response, indicating no target detected anywhere. In the latter case, if the subject were forced to choose one of the two directions (as in the task in Müller et al., 2005), we posit that the subject would “guess” at each of the two directions with equal probability (50%), with no response bias for either direction.

First, we simulated the 2-ADC model for the case without microstimulation. As before, the model produced a monotonic improvement in percentage correct with the strength of motion coherence in the target patch (proportion of coherent dots; Fig. 7D,E, gray). In these simulations, the criterion values were equal at both locations ( $c_1 = c_2 = 1.5$ ).

Next, we simulated the effect of microstimulation as a 20% decrease in the value of the criterion at the microstimulated location (Fig. 7C); an identical decrease was applied symmetrically across the axis, implying that the microstimulation did not bias any particular direction of motion. Nevertheless, this configuration resulted in a spatial choice bias favoring the microstimulated location for both directions of motion. In this case, when sensory evidence at both locations was of equal strength, but indicated conflicting directions (Fig. 7C, red circles), the monkey would be more likely (biased) to make a choice based on the evidence at the microstimulated location. The effect is a shift in bias that preferentially gates sensory evidence at the microstimulated location.

Model simulations produced results strikingly similar to the observed effects of SC microstimulation in this study (Fig. 7D, black vs gray; compare with Müller et al., 2005, their Fig. 4a): increased percentage correct and a shift in the psychometric function toward lower target strengths resulting in a lower psychophysical threshold (strength of coherent motion to achieve a performance level of 82% correct; Müller et al., 2005). On the other hand, there was little change in the psychometric function at the opposite (nonmicrostimulated) location (Fig. 7E; compare with Müller et al., 2005, their Fig. 4b), as observed in the original study. Thus, again, the experimentally observed effects of SC microstimulation could be explained entirely by changes in spatial choice bias alone.

#### *The effects of SC inactivation in a spatially cued perceptual discrimination (filtering) task*

The third study we considered employed a task in which monkeys had to detect a change in a cued stimulus in the presence of a similar foil stimulus, a task referred to as a “filtering task” (Fig. 8A). In such filtering tasks, multiple stimuli are presented on each trial, and the animal is cued beforehand to attend to the location of one of the stimuli (target). The animal is rewarded for accurately discriminating a feature (e.g., motion direction) of the cued target stimulus while ignoring the foil (distractor) stimuli.

The task in this study (Lovejoy and Krauzlis, 2010) involved the presentation of four patches of randomly moving dots, one in each quadrant of the visual field, with one of the four patches cued as the target patch (Fig. 8A). Following a brief delay, the cued (target) patch and the diametrically opposite uncued (foil) patch moved coherently for a brief period (160 ms). The animal was rewarded for making a saccade in the direction indicated by the target motion patch, while ignoring the motion of the foil patch. The effect of SC inactivation was measured in terms of changes in proportion of responses based on the target versus foil-patch motion direction.

Lovejoy and Krauzlis discovered that upon inactivating the SC representation at the cued (target) location, the animals responded more frequently based on the direction indicated by the foil stimulus ( $\sim 20\%$  before inactivation to  $\sim 70\%$  after inactivation; Lovejoy and Krauzlis, 2010, their Fig. 4a,c). However, the proportion of “neither” responses, in which the animal’s response was based on neither target nor foil motion did not change upon inactivation, indicating that the change in behavior was not simply a result of the animal guessing more frequently. In addition, inactivation of the SC representation corresponding to the foil location produced a complementary effect such that the animals based their decision on target motion even more frequently than at baseline (Lovejoy and Krauzlis, 2010, their Fig. 4b,d).

We simulated this task with a model for discrimination tasks adapted for such filtering tasks (Fig. 8B; see Materials and Methods, Extension of the model to filtering tasks involving perceptual discrimination). We posited that cueing a location for attention would produce a greater sensitivity at the cued (target) location relative to the foil location ( $d_{\text{target}} = 1.35$ ;  $d_{\text{foil}} = 1.05$ ). Sensory evidence from the target was biased in the decision by setting the choice criterion for the target stimulus ( $c_t = 0.2$ ) to be substantially ( $7\times$ ) lower than the criterion for the foil location ( $c_f = 1.4$ ; Fig. 8B). We further proposed that when sensory evidence was weak at both locations (Fig. 8B, central gray zone), the animal would indicate a location that was different from both the target and the foil locations (“neither”). Finally, we modeled the effect of SC inactivation of the cued (target) location by reversing the relative magnitudes of the criteria at the target and foil locations (Fig. 8C); sensitivity was assumed to remain unaltered.

The results of simulating the model produced the same pattern of effects, both qualitative and quantitative, compared with the reported effects of SC inactivation of the cued location (Lovejoy and Krauzlis, 2010): the proportion of choices based on the foil, initially at  $\sim 23\%$  increased to  $\sim 67\%$ , whereas the proportion of choices based on the target fell to 28% from 72%, with the proportion of “neither” responses remaining unchanged at  $\sim 5\%$ . Again, simulating the effect of inactivating the foil location produced precisely the same complementary effect as observed experimentally, with a slight increase in the proportion of choices based on the target (89%), and a decrease in the proportion of choices based on the foil (6%), again with no change in the proportion of “neither” responses. Thus, the precise pattern of effects in a cued discrimination task following SC inactivation could be readily replicated in our model framework by changes of choice bias alone (Krauzlis et al., 2013).

#### *The effects of SC inactivation in a spatially cued, change detection (filtering) task*

The final study we considered employed a filtering task, similar to the one described previously, except that the animal is rewarded for detecting and reporting a change in the cued stimulus (tar-

get), and for ignoring any changes in uncued, foil stimuli (Zénon and Krauzlis, 2012).

In the task in this study (Fig. 9A), the presentation of a spatial cue was followed by two motion stimuli (moving dot patches), one at the cued location (target) and the other at the diagonally opposite location (foil). The animal was rewarded for reporting a change in the target's motion direction by pushing a button ("Yes") and for abstaining from responding ("No") when the change occurred at the foil location (Zénon and Krauzlis, 2012). Task performance following SC inactivation was measured in terms of changes in percentage correct (hit rates: percentage of correctly reported Yes responses to changes in the target stimulus) and percentage incorrect (error rates: percentage of incorrectly reported Yes responses to changes in the foil stimulus). The percentage of false alarms (Yes responses on catch trials) was not reported.

This study demonstrated that when a target stimulus was presented inside the SC inactivated zone, the monkey rarely reported direction changes in the target stimulus, resulting in a dramatic decrease in percentage correct performance (Zénon and Krauzlis, 2012, their Fig. 1c, top left). At the same time, changes in the foil stimulus were reported more frequently, corresponding to an increase in error rates (Zénon and Krauzlis, 2012, their Fig. 1c, bottom right). In contrast, when the target stimulus was presented outside the inactivated zone, performance was not different from the preinactivation baseline (Zénon and Krauzlis, 2012, their Fig. 1c, top right and bottom left). The button-press protocol in this task design excluded the contribution of any oculomotor response biases to the effects on task performance.

We simulated this filtering task with a 2-ADC model (Fig. 9B). To account for the ability of the monkey to perform this task before SC inactivation, we hypothesized that cueing increased the monkey's perceptual sensitivity for detecting motion direction changes at the cued location ( $d_{\text{target}} = 2.0$ ; see Discussion) compared with at the foil location ( $d_{\text{foil}} = 1.0$ ); the specific choice of  $d$  values did not alter the results. Next, to simulate the  $\sim 70\%$  hit rates and the  $\sim 5\%$  error rates measured before inactivation, we chose, respectively, an intermediate value of the criterion at the target location ( $c_{\text{target}} = 1.5$ ) and a high criterion at the foil location ( $c_{\text{foil}} = 3.0$ ). As before, we assumed that the animal maintained a constant, low percentage of false alarms during catch trials ( $\sim 7\%$ ); the conclusions were robust to variations in these specific parameter values (see Materials and Methods).

We employed this model to explore the effects of changes in choice bias, as might be caused by SC inactivation. Specifically, we hypothesized that SC inactivation shifts the animal's choice bias away from the inactivated location. We modeled this shift by increasing the criterion at the inactivated location by  $\Delta c_{\text{inact}} = 1.5$ , to match the observed reduction in percentage correct with inactivation. Figure 9C depicts the 2-ADC model of this inactivation-induced bias effect (blue): the criterion at the inactivated location was increased and the criterion at the noninactivated location was decreased, so that performance traveled along the contour of constant false-alarm rate (Fig. 9C, dot-dashed black curve; see Materials and Methods, Eq. 25).

Simulating this model produced results strikingly similar to the observed effects of SC inactivation on behavioral performance. All of the effects on hit rates (percentage correct) and error rates (percentage incorrect) to the target and foil stimuli (four independent effects; Fig. 9D–G; compare with Zénon and Krauzlis, 2012, their Fig. 1c) were accurately reproduced by increasing the single free parameter,

$\Delta c_{\text{inact}}$ , which simulates a change in choice bias away from the inactivated location (Fig. 9C).

The results are consistent, once again, with the hypothesis that, rather than affecting perceptual sensitivity ( $d$ ), inactivating the SC resulted in a change in choice bias away from the inactivated region.

#### *m-ADC model analysis of SC inactivation studies*

We have shown, thus far, that the effects of SC manipulation (microstimulation or inactivation) can be entirely accounted for by changes in choice bias alone, indicating that the SC acts predominantly by controlling the choice bias for attended locations. These results were based on simulating the m-ADC model with a fixed set of parameters. We sought to confirm these findings by fitting the m-ADC model to behavioral data from the last two seminal SC inactivation studies (Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012).

First we analyzed the SC inactivation study involving motion direction discrimination (Fig. 8A; Lovejoy and Krauzlis, 2010). The task in this study involved the simultaneous presentation of target and foil motion patches, each moving in one of four directions, although motion directions for the target and foil patch were never identical (Fig. 8A). Animals responded indicating the direction of motion of the target patch (with a saccade or button press). Each response could correspond to the direction of target-patch motion, foil-patch motion, or neither (three response types). The 12 stimulus event types and three response types resulted in a  $12 \times 3$  contingency table. In the simulations (Fig. 8), we had modeled only a pair of motion directions to illustrate the effect of a change in choice bias. Here we fit these data with a fully descriptive 8-ADC model. The model represented evidence for each of the four directions of motion at the target and foil locations along each axis of an eight-dimensional decision space (see Materials and Methods).

SC inactivation produced two more  $12 \times 3$  contingency tables, corresponding to trials in which targets or foils, respectively, were presented at the inactivated location. We modeled the entirety of these three  $12 \times 3$  contingency tables by extending the 8-ADC model to incorporate three additional parameters: changes in sensitivities at inactivated locations, changes in sensitivities at noninactivated locations, and changes in criterion at the noninactivated location relative to the inactivated location (see Materials and Methods). We term this model the 8-ADC-inact model. To avoid the effects of induced motoric biases, we only included trials in which neither the cued target nor the foil motion patches pointed into the inactivated quadrant.

The 8-ADC-inact model provided a good fit to the monkeys' behavior both before and after inactivation (responses pooled across motion directions,  $\chi^2_{(6,3687)} = 12.07$ ,  $p_r = 0.062$  for monkey F, and  $\chi^2_{(6,1593)} = 2.09$ ,  $p_r = 0.915$  for monkey M randomization test; see Materials and Methods). Before inactivation, sensitivity at the foil location was closely similar to that at the target location ( $d_f/d_T = 0.93$  for monkey F and 0.99 for monkey M). On the other hand, the criterion for foil-based decisions was consistently higher than for target-based decisions in both animals ( $c_F - c_T = 0.9$  for monkey F, and 0.81 for monkey M). Following SC inactivation, the model indicated that sensitivity marginally decreased at the inactivated location ( $d_{\text{IN}}/d_{\text{PRE}} = 0.65$  for monkey F and 0.75 for monkey M), and was relatively unaltered or marginally increased at the noninactivated location ( $d_{\text{OUT}}/d_{\text{PRE}} = 0.94$  for monkey F and 1.21 for monkey M). Moreover, the criterion for decisions based on motion at the inactivated location was consistently greater than the criterion for decisions based on motion at



**Table 11. Model comparison for SC inactivation data in a cued, motion direction discrimination task (Lovejoy and Krauzlis, 2010)**

Model	$n_p$	Monkey F		Monkey M	
		AIC	BIC	AIC	BIC
8-ADC-inact	12	5369.68 <sup>a</sup>	5444.14	2592.30 <sup>a</sup>	2656.59
8-ADC-inact sens-change	11	5389.73	5458.00	2595.64	2654.58
8-ADC-inact crit-change	10	5378.02	5440.09 <sup>a</sup>	2594.70	2648.29 <sup>a</sup>

$n_p$ , Number of parameters.

<sup>a</sup>Model with the lowest AIC or BIC value.

the noninactivated location for both animals ( $c_{IN} - c_{OUT} = 1.12$  for monkey F and 1.17 for monkey M). In Figure 10, we plot the effects of SC inactivation on the difference between the sensitivities, and the difference between the criteria, at the target and foil locations for both animals (monkey F and monkey M). These results indicate that SC inactivation produced consistent effects on sensitivities and criteria in both animals. Moreover, for both animals, criterion differences were dramatically reversed (reversal of bias) following inactivation of the attended target's representation in the SC (Fig. 10A, bottom, green vs dark blue).

We next asked which of these effects of SC inactivation—those on sensitivity, on criteria, or both—could best account for the observed behavioral data. To answer this question, we computed the AIC and BIC metrics for the original model (the 8-ADC-inact model) as well as two adaptations of the 8-ADC-inact model: one that modeled only changes in sensitivity with inactivation (8-ADC-inact<sub>sens-change</sub>), and one that modeled only changes in criteria with inactivation (8-ADC-inact<sub>crit-change</sub>). In each monkey, and regardless of the choice of measure (AIC or BIC), the 8-ADC-inact<sub>crit-change</sub> model outperformed the 8-ADC-inact<sub>sens-change</sub> model (Table 11;  $\Delta AIC = -11.71$ ,  $\Delta BIC = -17.91$  for monkey F;  $\Delta AIC = -0.94$ ,  $\Delta BIC = -6.19$  for monkey M). Even when we included in the comparisons the full 8-ADC-inact model that incorporated both sensitivity and criterion changes with inactivation, the 8-ADC-inact<sub>crit-change</sub> model exhibited lower BIC values consistently across the two monkeys (Table 11). These results indicate that criterion changes, as compared with sensitivity changes, better account for the behavioral effects of SC inactivation. These criterion changes represent a shift in choice bias away from the inactivated location (see Materials and Methods, Eq. 27).

Next, we analyzed the SC inactivation study involving motion change detection (Zénon and Krauzlis, 2012). This study required the animal to detect a change in the direction of motion of a target patch and ignore changes in motion in a foil patch presented concurrently. Changes could occur in the target patch, foil patch, or neither, and the animals responded with a Yes (change in target patch) or No (no change in target patch) response, resulting in a  $3 \times 2$  contingency table. A 2-ADC model does not suffice to model behavior in this task (see Materials and Methods). However, the SC inactivation produced two other  $3 \times 2$  contingency tables, one when the target was presented in the inactivated location and one when the foil was in the inactivated location. As before, we modeled the entirety of these three  $3 \times 2$  tables by extending the 2-ADC model with four parameters: (1) change in sensitivity at the inactivated location, (2) change in sensitivity at the noninactivated location, (3) change in criterion at the inactivated location, and (4) change in criterion at the noninactivated location (the explanation for four additional parameters vs three in the previous case is given in Materials and Methods). We term this model the 2-ADC-inact model. Because the task design did not distinguish correct rejections based on

**Table 12. Model comparison for SC inactivation data in a cued, motion change detection task (Zénon and Krauzlis, 2012)**

Model	$n_p$	AIC	BIC
2-ADC-inact	8	4999.82 <sup>a</sup>	5053.79
2-ADC-inact sens-change	6	5042.33	5082.82
2-ADC-inact crit-change	6	5003.01	5043.49 <sup>a</sup>

Conventions are as in Table 11.

perceived foil motion from those based on no change, we pooled these model responses in the fitting procedure. As the data from monkeys J and M were highly similar (Zénon and Krauzlis, 2012), data from these animals were combined for this analysis.

Again, the 2-ADC-inact model provided a good fit to the behavioral data, both before and after SC inactivation ( $\chi^2_{(4,6306)} = 11.65$ ,  $p_r = 0.23$ , randomization test; see Materials and Methods). Before inactivation, sensitivity at the target and foil locations were not different ( $d_F = d_T = 2.64$ ), whereas the criterion at the foil location was 1.8 times higher than at the target location ( $c_F = 2.33$ ,  $c_T = 4.24$ ). Following SC inactivation, the model indicated that sensitivity was marginally reduced at the inactivated location relative to preinactivation baseline ( $d_{IN}/d_{PRE} = 0.66$ ) and was relatively unaltered at the noninactivated location ( $d_{OUT}/d_{PRE} = 0.88$ ). In addition, the criterion for decisions based on changes at the inactivated location increased relative to preinactivation baseline ( $c_{IN}/c_{PRE} = 1.22$ ), whereas the criterion for decisions based on changes at the noninactivated location decreased relative to baseline ( $c_{OUT}/c_{PRE} = 0.74$ ). These results are fully concordant with those obtained in the previous analysis based on the Lovejoy and Krauzlis (2010) dataset (Fig. 10A, B).

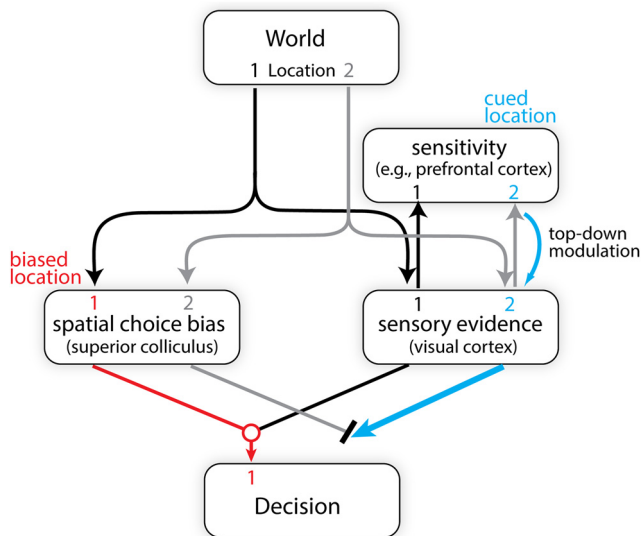
As before, we performed model comparisons by computing AIC and BIC metrics for the 2-ADC-inact model (the original model), a 2-ADC-inact<sub>sens-change</sub> model that modeled only changes in sensitivity with inactivation, and a 2-ADC-inact<sub>crit-change</sub> model that modeled only changes in criteria with inactivation. Fully in line with the results of the previous analysis, the 2-ADC-inact<sub>crit-change</sub> model outperformed the 2-ADC-inact<sub>sens-change</sub> model regardless of the choice of comparison measure (AIC or BIC; Table 12;  $\Delta AIC = -39.32$ ,  $\Delta BIC = -39.33$ ). As before, in comparison with the full 2-ADC-inact model, the 2-ADC-inact<sub>crit-change</sub> model exhibited lower BIC values (Table 12).

These convergent results across two completely different experimental datasets strongly suggest that the predominant consequence of SC inactivation is to change the choice criterion in a space-specific manner. Changes in the choice criterion, in turn, indicate that the SC biases the gating of sensory information into decision-making networks at the attended location (see Discussion; Fig. 11). In addition, the results demonstrate the utility and power of our m-ADC modeling framework for distinguishing the effects of sensitivity from those of choice bias in such multialternative experiments.

## Discussion

Multialternative task designs are increasingly used in animal studies that seek to identify the neural bases of cognitive processes, including selective attention (Cavanaugh and Wurtz, 2004; Williford and Maunsell, 2006; Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012; Sridharan et al., 2013; Steinmetz and Moore, 2014) and decision making (Churchland et al., 2008; Niwa and Ditterich, 2008; Ditterich, 2010; Churchland and Ditterich, 2012). The m-ADC model, developed and validated here with behavioral data, provides a principled framework for ana-





**Figure 11.** A schema for how sensitivity and bias mechanisms in the forebrain and midbrain interact to control visual spatial attention. Sensitivity (blue) and bias (red) mechanisms act along parallel, convergent neural pathways to enhance performance in spatial attention tasks. Cueing a location engages top-down mechanisms (e.g., from the frontoparietal network) that selectively enhance neural sensitivity to sensory evidence at the cued location (blue arrow). The enhancement is manifest in neural encoding in the visual cortex. At the same time, neural signals from the SC bias particular spatial locations by giving sensory evidence from these locations the greatest weight (red arrow). The SC spatial bias influences the animal's perceptual decisions by selectively gating (up/down weighting) the sensory evidence from different spatial locations to the decision process (evidence accumulator). Typically, the location of enhanced sensitivity (blue) and the location biased by the SC (red) are the same. The distinct contribution of each mechanism is revealed by experimental manipulations (such as SC stimulation or inactivation) that cause these mechanisms to favor different spatial locations, as shown in this schema.

lyzing and interpreting data collected in such multialternative task designs.

Applying our model to published results from multialternative tasks provided important insights into how the SC controls spatial attention. The analyses revealed that the SC contributes to attention, predominantly through mechanisms that alter spatial choice bias, which is produced by differential control of choice criteria at attended versus unattended locations.

### Fundamental components of choice bias

At least two components of task design influence choice bias: (1) the prior probability (“priors”) of each stimulus event, and (2) the magnitude of reward (“payoffs”) for a correct response for each stimulus event. Each of these has been manipulated in previous studies to induce choice bias in perceptual decision tasks (Mulder et al., 2012; Vintch and Gardner, 2014). From a Bayesian view of decision making, priors and payoffs are optimally combined to maximize success during task performance (Green and Swets, 1988; Gold and Shadlen, 2007). Our model formalizes the relationship among Bayesian priors, payoffs, and choice criteria and biases in multialternative tasks (see Materials and Methods, Eqs. 26–28).

Application of the m-ADC model to behavioral results revealed the existence of two kinds of spatial bias. Chickens exhibited an innate bias for targets (small dots) in the lower visual hemifield. This bias likely reflects the significantly higher chance of encountering rewarding stimuli (food items) in the lower versus the upper visual field when they search the ground for food (Sridharan et al., 2013). The monkey exhibited a bias, associated

with a near-optimal criterion shift that was induced by a highly informative spatial cue. This bias reflects a learned expectation that is expressed only in the context of the task. Thus, these biases may reflect Bayesian priors that arise from adaptation to the statistics of the environment over various timescales. Rather than treating these biases as a nuisance that must be trained away, our model quantifies such biases and, thereby, enables deciphering the complex relationship between an animal's perception and its behavior (Carandini and Churchland, 2013).

### Assumptions and extensions of the m-ADC framework

The m-ADC framework incorporates simplifying assumptions, including uncorrelated noise distributions and stationarity of the noise variance (across trials). In addition, it does not distinguish between external and internal noise (Lu and Doshier, 1998) and does not explicitly model response bias. Notice that these assumptions impose certain constraints on the kinds of responses that can be modeled with the mADC framework. For example, the assumption of uncorrelated noise implies that the probability of a false alarm for a particular response type (indicating response  $r_1$  when no stimulus was presented), can never be less than the probability of misidentification (indicating response  $r_1$  when stimulus type  $s_2$  was presented). Despite these assumptions, our parsimonious model provided an excellent fit to behavioral data from different kinds of animals across a wide range of experiments, and it consistently outperformed other candidate models.

The model may be readily extended to incorporate additional structure both in the decision variables and decision rule. For instance, correlations between the noise distributions can be modeled by altering the covariance matrix  $C$  of decision variables (see Materials and Methods, Eq. 1), so that  $C$  is no longer a diagonal matrix, but also incorporates off-diagonal terms. In this case, the optimal decision rule would have to be updated based on statistical decision theory (Eckstein et al., 2009), and is likely to involve quadratic decision surfaces (unlike the simpler, linear decision surfaces of the m-ADC model). Similarly, nonstationarity of the variance of the noise distribution (across trials) could be modeled by considering the distribution of the noise variance as an additional parameter of the model. Were signal and noise variance to scale unequally (for example, if the internal noise variance covaried with signal strength), the decision variable axis would no longer be monotonically related to the log-likelihood ratio of the hypothesis, and the optimal decision strategy would involve a much more complex decision rule involving more than one cutoff criterion per decision axis (Sridharan et al., 2014b; Cabrera et al., 2015). In addition, the model can be extended to incorporate motoric response bias by including, for each response type, an additional multiplicative parameter whose value scales with the magnitude of the bias for that response. Generally, such a model must be accompanied by an appropriate task design for distinguishing response bias from choice bias (Garcia-Perez and Alcalá-Quintana, 2013; Sridharan et al., 2014b). For example, for the tasks shown in Figure 4 or 8, such an extended model has the potential to decouple choice bias from response bias, as the choice is made based on sensory evidence from a location that is distinct from the location of the response.

### Neural mechanisms of attention: sensitivity or bias?

Our results are highly relevant to resolving a fundamental question regarding the neurobiology of attention (Carrasco, 2011): do improvements in behavioral performance at the attended location reflect the action of neural mechanisms that enhance the quality of sensory information at the attended location (percep-

tual sensitivity), those that selectively gate sensory evidence at the attended location (choice bias), or both?

Analysis of behavior in the spatial cueing task (Fig. 4) with the m-ADC model yielded preliminary insights into this question. The monkey's perceptual sensitivity was marginally but reliably higher at two locations—the spatially cued location (C) and the location of the impending saccade (R)—indicating that the animal was dividing its attention between the cue and response locations. On the other hand, choice criteria were lowest for changes at the cued location and highest for changes at the response location. These results indicate that spatial attention can induce changes in bias, sensitivity, or both, and that mechanisms that alter choice criterion and bias can be dissociated from those that alter perceptual sensitivity. These observations resonate with a recent proposal: performance improvements in attention tasks can arise from changes in the animal's decision policies (bias) regardless of changes of sensitivity (Krauzlis et al., 2014).

Moreover, when we simulated the effects of attentional cueing with the m-ADC model (Figs. 1E–G, 6B–F), we found that changes in choice bias (decision rule) could result in exactly the same pattern of changes in hit, error, and false-alarm rates, or changes in the slope of the psychometric function that are routinely assumed to represent changes in perceptual sensitivity when analyzed with conventional one-dimensional models. The results urge caution when inferring changes in sensitivity based on these conventional behavioral metrics alone (Eckstein et al., 2013). Although in some cases the location of greatest bias may be directly inferred from the pattern of false-alarm and miss rates (Figs. 2B, 4B), the m-ADC model permits quantifying the magnitude of this bias for multialternative behavioral data. In addition, the m-ADC model fit behavioral data in multialternative tasks better and more parsimoniously (Tables 9, 10) than alternative candidate models. By quantifying the distinct contributions of bias and sensitivity to behavioral data, the parsimonious framework of the m-ADC model provides a valuable tool for identifying underlying neural mechanisms of selective attention.

### Role of the SC in visuospatial attention

Simulations with the m-ADC model provided important insights into the contribution of the SC to spatial attention. The behavioral effects of microstimulation or inactivation of the SC, from four important studies, were quantitatively reproduced by the model with changes in spatial choice bias alone, and fitting the model to behavioral data independently confirmed these findings. In addition to our results, converging evidence from the original studies indicate that, in each case, the behavioral effects of SC manipulation were not due to changes in perceptual sensitivity.

In the first study (Cavanaugh and Wurtz, 2004), additional behavioral evidence confirmed the spatial choice bias effects of SC microstimulation. Had the microstimulation improved perceptual sensitivity without altering choice criteria, false-positive rates should have been unchanged at the microstimulated location. A geometric intuition for this result is obtained by translating the decision variable distribution for a stimulus at one location (the microstimulated location) further from the origin, without concurrently moving the decision surfaces in Figure 6B. On the other hand, the data from this study overwhelmingly indicated that in experiments in which SC microstimulation increased percentage correct (hits) significantly, false-positive rates were also increased (in five of six experiments; Cavanaugh and Wurtz, 2004, their Fig. 5A,B). These results are consistent with a

selective effect of SC microstimulation on increasing spatial choice bias toward the microstimulated location.

The second study (Müller et al., 2005) used an elegant, forced-choice design involving a motion-direction discrimination task. In this task, the location of SC microstimulation could not have provided the animal with information regarding the correct answer (motion direction). Nevertheless, when analyzed with a 2-ADC model (a two-dimensional model), the effects of SC microstimulation were accurately accounted for by an increase in spatial choice bias toward the microstimulated location. On the other hand, a conventional (one-dimensional) 2-AFC model would predict that the shift in the psychometric function represented an improvement in perceptual sensitivity. Which interpretation is correct?

Attentional enhancement of perceptual sensitivity is thought to operate by differential allocation of limited perceptual resources (Eckstein et al., 2009; Carrasco, 2011). In the study of Müller et al. (2005), while percentage correct performance increased at the SC microstimulated location, no concomitant decrease in percentage correct occurred at the opposite (nonmicrostimulated) location. Had SC microstimulation increased performance by enhancing perceptual sensitivity at a location, the limited resources theory predicts that performance should have deteriorated at the other location. The lack of such an effect (Müller et al., 2005, their Fig. 4b) is consistent with our proposal that SC microstimulation induced a choice bias toward the microstimulated location (Fig. 7C). Further experiments incorporating an unforced-choice design (with a NoGo response) could distinguish between these interpretations.

The third and fourth studies (Lovejoy and Krauzlis, 2010; Zénon and Krauzlis, 2012) demonstrated that SC inactivation of the target's sensory representation resulted in a detriment to behavioral performance because the animal based its decisions on sensory information from the foil.

Analysis of both of these studies with our m-ADC model framework indicated a highly convergent pattern of results (Fig. 10). The main difference was that  $c_T - c_F$  became positive following inactivation in one case (Fig. 10A), but not in the other. This was perhaps due to its much larger negative value ( $c_T \ll c_F$ , high choice bias toward the target) even during the preinactivation baseline in the second study (Fig. 10B), although the extent of criterion modulation remained about the same across the studies.

Our model comparison analysis (Tables 11, 12) demonstrated that, in both studies, the effects of SC inactivation could be best accounted for by a model that incorporated only changes in criteria, rather than only changes in sensitivity. Furthermore, the model that incorporated criterion changes alone also outperformed the full model that incorporated both changes in criterion and sensitivity. These results indicate that the predominant effect of SC inactivation is a change in choice bias. This conclusion is consistent with observations by Lovejoy and Krauzlis (2010): they reported that when no foil was presented, and the target was presented by itself, the animal did not show appreciable performance deficits, suggesting that sensitivity at the inactivated location was relatively unimpaired.

The fourth study (Zénon and Krauzlis, 2012) provided neurophysiological evidence that is also consistent with our findings. This study measured the effects of spatial cueing on behavioral detection of motion change and, simultaneously, on neural activity in a visual cortical area (MT). The results demonstrated that, with SC inactivation, behavioral performance plummeted at the cued location while the neural signatures associated with improved sensory encoding for that location were unaffected.

These apparently paradoxical findings are readily explained within the parsimonious framework of our m-ADC model (Fig. 9B,C). In this framework (Fig. 11), the SC contributes to improved behavioral performance, specifically, by enhancing spatial choice bias. Neural signals from the SC improve performance at a particular spatial location by giving sensory evidence from that location greater weight (Fig. 11, red). Such SC-biased sensory evidence dominates the animal's perceptual decisions by selectively gating (up/down weighting) sensory evidence, and acts downstream of forebrain sensitivity enhancement mechanisms (Fig. 11, blue). Consequently, the effects of SC inactivation are explained by a choice bias (criterion shift) away from the inactivated location (Fig. 9C) without affecting sensitivity to sensory information in the forebrain.

Our model provides a framework for gaining a mechanistic understanding of the contribution of the midbrain network to selective spatial attention from the analysis of behavioral data (Fig. 11). This framework also explains other previous reports of attention deficits following SC lesions. For example, viewed from the perspective of this framework, reported effects of spatial hemineglect that appear following SC lesions (Dean and Redgrave, 1984a,b; Overton et al., 1985; Midgley et al., 1988) represent the effects of severe spatial choice bias that accompanies SC lesions. Our framework also offers a parsimonious explanation for the so-called "Sprague effect," the observation that the hemianopsia that results from a unilateral lesion of the visual cortex is greatly diminished by the removal of the contralateral SC. According to our schema (Fig. 11), removal of the contralateral SC causes a powerful spatial choice bias that blocks sensory information from the normal hemifield and enhances information from the previously anoptic hemifield (Sprague, 1966).

Our findings demonstrate the need for careful experimental designs for distinguishing perceptual sensitivity from choice bias in attention tasks. Our model and framework will find important application in teasing apart the contribution of various brain regions and circuits to these fundamental components of attention.

## Notes

Supplemental material for this article is available at <http://cns.iisc.ac.in/sridhar/publications.html>. This material has not been peer reviewed.

## References

- Bisley JW, Goldberg ME (2010) Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci* 33:1–21. [CrossRef Medline](#)
- Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M (2008) Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J Neurosci* 28:10056–10061. [CrossRef Medline](#)
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference. New York: Springer.
- Cabrera CA, Lu ZL, Doshier BA (2015) Separating decision and encoding noise in signal detection tasks. *Psychol Rev* 122:429–460. [CrossRef Medline](#)
- Carandini M, Churchland AK (2013) Probing perceptual decisions in rodents. *Nat Neurosci* 16:824–831. [CrossRef Medline](#)
- Carrasco M (2011) Visual attention: the past 25 years. *Vision Res* 51:1484–1525. [CrossRef Medline](#)
- Cavanaugh J, Wurtz RH (2004) Subcortical modulation of attention counters change blindness. *J Neurosci* 24:11236–11243. [CrossRef Medline](#)
- Churchland AK, Ditterich J (2012) New advances in understanding decisions among multiple alternatives. *Curr Opin Neurobiol* 22:920–926. [CrossRef Medline](#)
- Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. *Nat Neurosci* 11:693–702. [CrossRef Medline](#)
- Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB, Bradley DC, Smith MA, Kohn A, Movshon JA, Armstrong KM, Moore T, Chang SW, Snyder LH, Lisberger SG, Priebe NJ et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci* 13:369–378. [CrossRef Medline](#)
- Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12:1594–1600. [CrossRef Medline](#)
- Corbetta M, Shulman GL (2011) Spatial neglect and attention networks. *Annu Rev Neurosci* 34:569–599. [CrossRef Medline](#)
- Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58:306–324. [CrossRef Medline](#)
- Dean P, Redgrave P (1984a) The superior colliculus and visual neglect in rat and hamster. I. Behavioural evidence. *Brain Res* 320:129–141. [Medline](#)
- Dean P, Redgrave P (1984b) The superior colliculus and visual neglect in rat and hamster. II. Possible mechanisms. *Brain Res* 320:143–153. [Medline](#)
- DeCarlo LT (2012) On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *J Math Psychol* 56:196–207. [CrossRef](#)
- Ditterich J (2010) A comparison between mechanisms of multi-alternative perceptual decision making: ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory. *Front Neurosci* 4:184. [CrossRef Medline](#)
- Eckstein MP, Peterson MF, Pham BT, Droll JA (2009) Statistical decision theory to relate neurons to behavior in the study of covert visual attention. *Vision Res* 49:1097–1128. [CrossRef Medline](#)
- Eckstein MP, Mack SC, Liston DB, Bogush L, Menzel R, Krauzlis RJ (2013) Rethinking human visual attention: spatial cueing effects and optimality of decisions by honeybees, monkeys and humans. *Vision Res* 85:5–19. [CrossRef Medline](#)
- García-Pérez MA, Alcalá-Quintana R (2011) Interval bias in 2AFC detection tasks: sorting out the artifacts. *Atten Percept Psychophys* 73:2332–2352. [CrossRef Medline](#)
- García-Pérez MA, Alcalá-Quintana R (2013) Shifts of the psychometric function: distinguishing bias from perceptual effects. *Q J Exp Psychol (Hove)* 66:319–337. [CrossRef Medline](#)
- Gold JI, Ding L (2013) How mechanisms of perceptual decision-making affect the psychometric function. *Prog Neurobiol* 103:98–114. [CrossRef Medline](#)
- Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574. [CrossRef Medline](#)
- Green DM, Swets JA (1988) Signal detection theory and psychophysics. Los Altos, CA: Peninsula.
- Herrmann K, Montaser-Kouhsari L, Carrasco M, Heeger DJ (2010) When size matters: attention affects performance by contrast or response gain. *Nat Neurosci* 13:1554–1559. [CrossRef Medline](#)
- Jogan M, Stocker AA (2014) A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *J Vis* 14(3):20. [CrossRef Medline](#)
- Knudsen EI (2011) Control from below: the role of a midbrain network in spatial attention. *Eur J Neurosci* 33:1961–1972. [CrossRef Medline](#)
- Krauzlis RJ, Lovejoy LP, Zénon A (2013) Superior colliculus and visual spatial attention. *Annu Rev Neurosci* 36:165–182. [CrossRef Medline](#)
- Krauzlis RJ, Bollimunta A, Arcizet F, Wang L (2014) Attention as an effect not a cause. *Trends Cogn Sci* 18:457–464. [CrossRef Medline](#)
- Lovejoy LP, Krauzlis RJ (2010) Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nat Neurosci* 13:261–266. [CrossRef Medline](#)
- Lu ZL, Doshier BA (1998) External noise distinguishes attention mechanisms. *Vision Res* 38:1183–1198. [CrossRef Medline](#)
- Luo TZ, Maunsell JH (2015) Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron* 86:1182–1188. [CrossRef Medline](#)
- Macmillan NA, Creelman DC (2005) Detection theory: a user's guide. Mahwah, NJ: Lawrence Erlbaum.
- Midgley GC, Wilkie DM, Tees RC (1988) Effects of superior colliculus lesions on rats' orienting and detection of neglected visual cues. *Behav Neurosci* 102:93–100. [CrossRef Medline](#)
- Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55:131–141. [CrossRef Medline](#)

- Mulder MJ, Wagenmakers EJ, Ratcliff R, Boekel W, Forstmann BU (2012) Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J Neurosci* 32:2335–2343. [CrossRef Medline](#)
- Müller JR, Philiastides MG, Newsome WT (2005) Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proc Natl Acad Sci U S A* 102:524–529. [CrossRef Medline](#)
- Niwa M, Ditterich J (2008) Perceptual decisions between multiple directions of visual motion. *J Neurosci* 28:4435–4445. [CrossRef Medline](#)
- Noudoost B, Moore T (2011) Control of visual cortical signals by prefrontal dopamine. *Nature* 474:372–375. [CrossRef Medline](#)
- Overton P, Dean P, Redgrave P (1985) Detection of visual stimuli in far periphery by rats: possible role of superior colliculus. *Exp Brain Res* 59:559–569. [Medline](#)
- Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26:703–714. [CrossRef Medline](#)
- Shaw ML (1980) Identifying attentional and decision-making components in information processing. In: *Attention and performance, VIII* (Nickerson RS, ed), pp 277–296. Hillsdale, NJ: Lawrence Erlbaum.
- Soltani A, Noudoost B, Moore T (2013) Dissociable dopaminergic control of saccadic target selection and its implications for reward modulation. *Proc Natl Acad Sci U S A* 110:3579–3584. [CrossRef Medline](#)
- Sperling G, Doshier BA (1986) Strategy and optimization in human information processing. In: *Handbook of perception and human performance* (Boff KR, Kaufman L, Thomas JP, eds), pp 1–65. New York: Wiley.
- Sprague JM (1966) Interaction of cortex and superior colliculus in mediation of visually guided behavior in the cat. *Science* 153:1544–1547. [CrossRef Medline](#)
- Squire RF, Noudoost B, Schafer RJ, Moore T (2013) Prefrontal contributions to visual selective attention. *Annu Rev Neurosci* 36:451–466. [CrossRef Medline](#)
- Sridharan D, Knudsen EI (2015) Selective disinhibition: a unified neural mechanism for predictive and post hoc attentional selection. *Vision Res* 116:194–209. [CrossRef Medline](#)
- Sridharan D, Ramamurthy DL, Knudsen EI (2013) Spatial probability dynamically modulates visual target detection in chickens. *PLoS One* 8:e64136. [CrossRef Medline](#)
- Sridharan D, Ramamurthy DL, Schwarz JS, Knudsen EI (2014a) Visuospatial selective attention in chickens. *Proc Natl Acad Sci U S A* 111:E2056–E2065. [CrossRef Medline](#)
- Sridharan D, Steinmetz NA, Moore T, Knudsen EI (2014b) Distinguishing bias from sensitivity effects in multialternative detection tasks. *J Vis* 14(9):pii:16. [CrossRef Medline](#)
- Steinmetz NA, Moore T (2014) Eye movement preparation modulates neuronal responses in area V4 when dissociated from attentional demands. *Neuron* 83:496–506. [CrossRef Medline](#)
- Vintch B, Gardner JL (2014) Cortical correlates of human motion perception biases. *J Neurosci* 34:2592–2604. [CrossRef Medline](#)
- White CN, Poldrack RA (2014) Decomposing bias in different types of simple decisions. *J Exp Psychol Learn Mem Cogn* 40:385–398. [CrossRef Medline](#)
- Williford T, Maunsell JH (2006) Effects of spatial attention on contrast response functions in macaque area V4. *J Neurophysiol* 96:40–54. [CrossRef Medline](#)
- Zénon A, Krauzlis RJ (2012) Attention deficits without cortical neuronal deficits. *Nature* 489:434–437. [CrossRef Medline](#)