Sensory and decisional components of endogenous attention are dissociable.

Sanjna Banerjee*, Shrey Grover*, Suhas Ganesh, Devarajan Sridharan
**these authors contributed equally*

***Affiliation*:**
[1]Centre for Neuroscience,
Indian Institute of Science,
C. V. Raman Avenue,
Bangalore 560012, India

***Corresponding author*:**
Devarajan Sridharan
Assistant Professor and
Wellcome-Trust DBT India Alliance Fellow,
Centre for Neuroscience, Rm F104
Indian Institute of Science,
C. V. Raman Avenue,
Bangalore 560012, India
Phone: +91 802 293 3434
E-mail: sridhar@iisc.ac.in

Abstract

Endogenously cueing a stimulus for attention enhances sensory processing of the attended stimulus, and efficiently selects information from the attended stimulus for guiding behavioral decisions. Yet, it is unknown if these sensory and decisional components of endogenous attention are under the control of common, overlapping or distinct mechanisms. Here, we tested human observers in a multi-alternative visuospatial attention task with probabilistic spatial cues, whose predictive validity varied across spatial locations. Observers' behavioral choices were analyzed with a multidimensional signal detection model, derived from Bayesian decision theory. The model effectively decoupled attention's effects on perceptual sensitivity from those on decisional thresholds (bias), and revealed striking dissociations between them. Sensitivity was highest at the cued location, and not significantly different among uncued locations, suggesting a spotlight-like allocation of sensory resources at the cued location. On the other hand, bias varied systematically with endogenous cue validity, suggesting a graded distribution of decision thresholds across all locations. Cueing-induced modulations of sensitivity and bias were uncorrelated within and across subjects. Finally, bias selectively correlated with key measures of decision-making: reaction times were strongly correlated with bias, cueing-induced benefits in reaction times were correlated with bias changes independently of sensitivity changes, and subjects who exhibited higher bias, rather than higher sensitivity, toward the cued location produced more optimal behavioral decisions. Our model and findings demonstrate that endogenous cueing of attention does not engage a unitary selection mechanism, but rather involves the operation of dissociable sensory enhancement and decisional selection processes.

Keywords: Top-down attention, perceptual sensitivity, decisional bias, efficient selection, signal detection theory, Bayesian decision theory, psychophysical models.

Significance statement

The capacity for selective attention allows us to engage efficiently with the most important objects in the world around us. Attention affects the way we perceive important stimuli and also affects how we make behavioral decisions about these stimuli. In this study, we test human observers on a multiple alternative attention task and show that attention's effects on perceptual processing and decision-making can be dissociated. Specifically, attention's sensory effects alone are in line with the conventional notion of an attentional "spotlight". These insights are obtained with a novel Bayesian decision model, which we propose as a powerful tool for teasing apart component processes of cognitive phenomena, and could aid the search for the neural correlates of these phenomena.

Introduction

Attention is the remarkable cognitive capacity that enables us to select and process only the most important information in our sensory environment. In laboratory tasks, endogenous attention is engaged by cues that are predictive of upcoming stimuli or events of interest. Numerous studies have explored the behavioral consequences and neural correlates of endogenous cueing of attention, and have reported systematic effects of endogenous attention both on sensory processing (1–5) and decision-making (6–9). Yet, it remains actively debated whether sensory and decisional effects of attention are under the control of a common, unitary attention mechanism or distinct, independent mechanisms (2, 10–13).

The sensory and decisional effects of attention may be quantified, respectively, with changes in signal-to-noise ratio (SNR) associated with sensory processing, and changes in decision thresholds (criteria) associated with downstream decision processes (5, 6, 14–16). Signal detection theory (SDT), a highly successful Bayesian framework for the analysis of behavior (17–19), specifies two key psychophysical metrics to enable this quantification: i) sensitivity (d'), which measures the improvement in the quality (SNR) of sensory information processing of the attended stimulus and ii) bias (b), which measures the relative weighting of information from the attended stimulus for guiding behavioral decisions (14, 16, 19). A rich literature has reported diverse, and often contradictory, effects of endogenous cueing on sensitivity and bias. For instance, previous studies have reported a benefit (1) or no benefit (6, 8) for sensitivity at the cued location, a cost (2, 14, 20) or no cost (1) for sensitivity at uncued locations, and a graded modulation of bias (13) or no modulation of bias (1) at cued and uncued locations. Studies investigating the neural basis of sensitivity and bias have only added to these controversies. For instance neural responses in area V4 have been reported to modulate only with sensitivity, but not with bias (12), or with bias, but not with sensitivity (9), or with both (21) (see Discussion for further details).

A key reason for these contradictions is the lack of an appropriate psychophysical (SDT) model for analyzing behavioral responses in spatial attention tasks. In the spatial attention task design, employed by the vast majority of studies in the attention literature (Fig. 1A) (22), the observer is cued to attend to one of two (or multiple) stimulus locations. At an unpredictable time following cue onset an event of interest, for example a change in grating orientation, occurs at one of the locations ("change" trials). On some trials, no events (changes) occur at any location ("catch" trials). On change trials, the observer must detect and report the location of the change or, on catch trials, must indicate that no change occurred. Such attention tasks have been commonly analyzed with a combination of multiple

binary choice (Yes/No) one-dimensional SDT models, with independent decision variables at each location (Fig. 1B) (1, 8, 12, 23). However, such a formulation is inadequate, at best, and incorrect, at worst, for modeling behavior in such tasks. Because the observer must report a single location of change, modeling behavior with multiple independent Yes/No decisions models produces a response conflict if a “Yes” decision occurs at more than one location on the same trial (Fig. 1B, see also Discussion).

Here, we analyze human attention behavior by extending a multidimensional signal detection model, the m-ADC model, which we recently developed to overcome these pitfalls (24, 25) (Fig. 1C). The model extends conventional signal detection theory and derives an optimal decision rule from Bayesian decision theory to effectively decouple sensitivity from bias in attention tasks. We show that the model successfully fits human behavioral choices in a multi-alternative (five-alternative) attention task employing probabilistic, endogenous cueing, and outperforms alternative models in model selection analysis. We apply the model to test if endogenous cueing engages common or distinct mechanisms to modulate sensitivity and bias. The results reveal essential dissociations among these fundamental components of attention.

Results

A model for predicting and fitting behavior in multi-alternative attention tasks

We measured the effect of endogenous cueing of attention on sensitivity and bias by extending the recently developed m-ADC model framework (24). We developed the model, from first principles, for multi-alternative attention tasks that employ the method of constant stimuli, i.e. in which stimuli can occur at different, unpredictable strengths at each location (SI Methods; Fig. 1C, SI Fig. S1A). This new model is essential, for example, for studies that seek to measure the effect of attention on the psychometric function at cued and uncued locations (25). The decision manifold in this model comprises a family of intersecting hyperplanes in a multidimensional decision space (Fig. 1C, SI Methods). These hyperplanes represent the decision criteria. These criteria are optimal for distinguishing each class of signal from noise (e.g. change at a given location versus no change), and provide a close approximation to the theoretical optimum for distinguishing signals of one class from another (e.g. changes at one location from another, SI Fig. S1B; SI Methods). The model is able to quantify perceptual sensitivities and decision criteria from stimulus-response contingency tables for tasks with any number of alternatives.

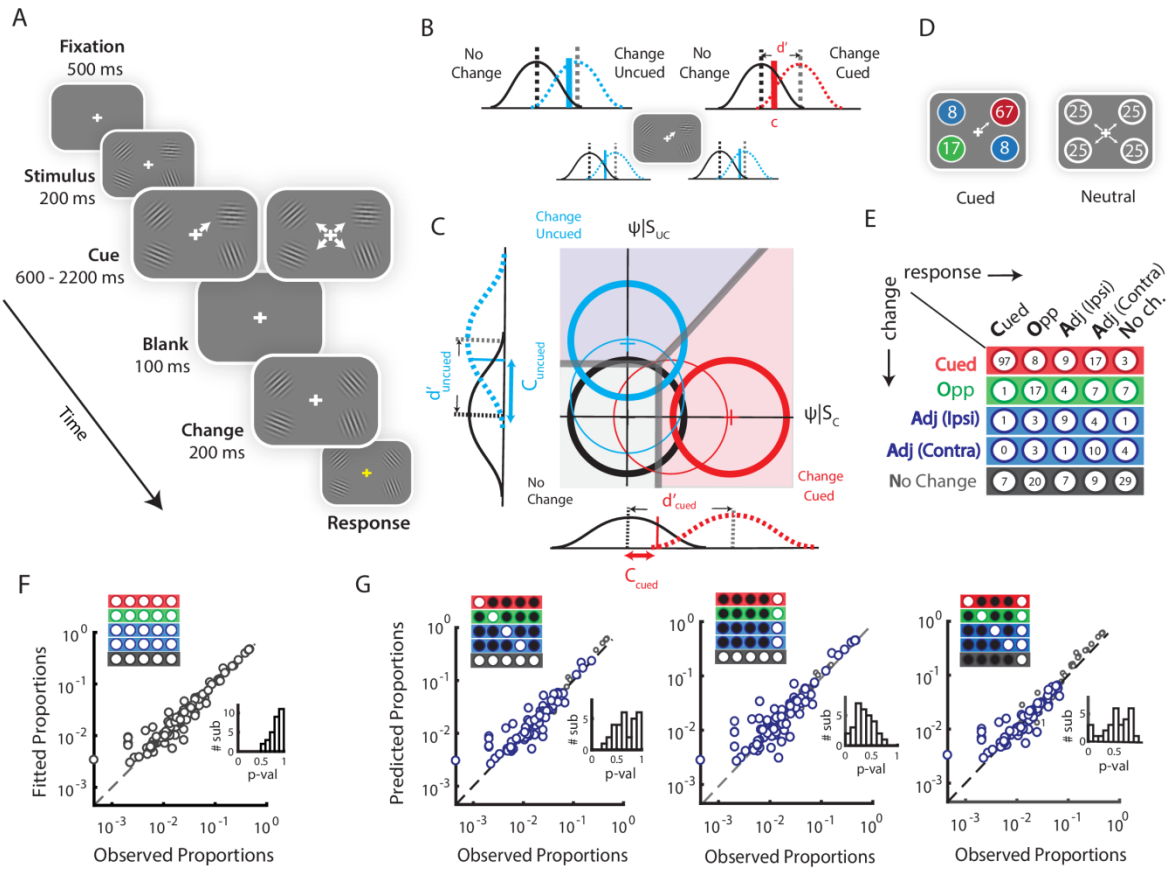


Figure 1. Evaluating a model for analyzing behavior in conventional attention tasks. A. Attention task protocol involving change detection. After fixation (500 ms), four Gabor patches appeared (200ms) followed by an attentional cue (central arrow). After a variable delay (600-2200 ms; exponentially distributed) all patches disappeared briefly (100 ms). Upon reappearance, either one of four patches changed in orientation ("change" trials) or none changed ("no change" trials). After 200 ms the fixation cross changed to yellow, instructing the subject to indicate the location of change (or no change) by pressing one of five different buttons. **B.** Multiple one-dimensional signal detection models representing independent Yes vs. No (change vs. no change) decisions at each location. In such a composite model a response conflict may arise when decision variables exceed their respective criteria simultaneously at multiple locations. Black Gaussians: noise distribution for "no change" trials; red Gaussian: signal distribution for "change" trials at the cued location; light blue Gaussians: signal distribution for "change" trials at an uncued location. Dashed vertical lines: means of signal and noise distributions, whose separation quantifies sensitivity (d') at each location. Solid vertical lines: Criterion at each location (c). **C.** A multidimensional (m-ADC) signal detection model, schematized here for changes at the cued and an uncued location. Decision variables at each location are represented along orthogonal decision axes in a multidimensional perceptual space. Circles: Contours of multivariate Gaussian decision variable distributions. Black: noise distribution; red: signal distribution for "change" trials at the cued location; light blue: signal distribution for "change" trials at an uncued location. Thin contours: lower sensitivity values; thick contours: higher sensitivity values. Thick gray lines: Decision manifold delineating domains in the perceptual space corresponding to cued, uncued and no change decisions. Gaussians at the side represent the marginals at each location**. D.** Proportion of change events at each location, relative to the cued location, in a block with high cue validity (left) and in a block with neutral cues (right). Red: Cued location; green: location opposite to the cue (Opp); blue: location adjacent to the cue, either in the visual hemifield ipsilateral to the cue (Adj-Ipsi) or in the visual hemifield contralateral to the cue (Adj-Contra).**E.** A contingency table from a representative behavioral session. The rows represent locations of change and the columns represent locations of response, both measured relative to the cue (data pooled across angles tested). Grey: "No change" trials. Color conventions are as in panel B.**F.** Fitting and predicting responses with a 4ADC model. Choice proportions fitted with the m-ADC model (ordinate) plotted against actual

choice proportions (abscissa). **G.** The three panels plot predicted proportions against actual response proportions by fitting either hits and false-alarms (case i), misses and false-alarms (case ii) or hits and misses (case iii). In each case, correct rejections were included in the fitting. Small grey circles – fitted proportions**.** Data points: Choice probabilities (joint) for each stimulus-response contingency at each angle (data pooled across subjects). Upper left insets: Subset of stimulus-response contingencies used for fitting. Lower right insets: Distribution of goodness of fit p-values across subjects.

We validated the model on a five-alternative visual change detection task (Fig. 1A). Following fixation, four oriented Gabor gratings appeared, one in each visual quadrant, and one of the four gratings was cued for attention with a central, endogenous cue (Fig. 1A; central, directed arrow). After an unpredictable delay, the screen was blanked and the four gratings reappeared. Following reappearance, either one of the four gratings had changed in orientation ("change" trials, 75%) or none had changed ("no change" trials, 25%). Participants indicated the location of change (one of four possibilities) or no change, with five distinct button press responses.

We tested two versions of the task on two different groups of subjects: a version that included only validly cued and invalidly cued trials and a version that also included neutrally cued trials), presented in pseudorandomly interleaved blocks (Fig. 1A, right). Valid cues indicated the location of the upcoming change with 66.7% validity and invalid cues with 8.3%-16.7% validity (Fig. 1D, left). Neutral cues were not predictive of the upcoming change (25% validity; Fig. 1D, right). In the following analyses, results from the validly and invalidly cued trials were combined across the two groups of subjects (n=30 subjects), except when reporting neutral cueing effects (n=10 subjects).

We examined how well the modified m-ADC model described human observers' behavior in this multialternative attention task.

First, we fitted a 4-ADC model for each subject's 5x5 stimulus-response contingency table obtained from the five-alternative task (exemplar table in Fig. 1E). Because changes could occur at one of 6 orientation change values, the contingency table for each subject contained 63 independent observations (details in Methods). For these analyses, responses across two of the uncued locations (ipsilateral and contralateral) were averaged into a single contingency (adjacent) because responses were not significantly different across these locations ($p>0.2$ for hits, misses and false alarms at these locations, signrank test). This simplification significantly reduced the number of model parameters to be estimated (Methods). Goodness-of-fit p-values obtained from a randomization test (based on the chi-squared statistic) were generally greater than 0.7 (median: 0.84; range: 0.57- 0.98; Fig. 1F,

lower inset), indicating that the model successfully fit observers' responses in this multialternative attention task.

Next, we tested the model's ability to predict individual subjects' responses: for this we fit the model using only a subset of the observers' behavioral choices (33%-58%) and tested its ability to predict their remaining (42%-67%) choices (Fig. 1G). Three different subsets of contingencies were selected for fitting: i) hits and false alarms (Fig. 1G, left), ii) false-alarms and misses (Fig. 1G, middle) or iii) hits and misses (Fig. 1G, right). Correct rejection responses were included either implicitly (cases i and ii) or explicitly (case iii; see SI Methods for details). The model was able to predict all of the remaining observations in the contingency table with high accuracy (Fig. 1G); p-values of randomization fit for predictions (median [95% CI]) case i: 0.69 [0.58-0.87]; case ii: 0.33 [0.25-0.48]; case iii: 0.60 [0.50-0.75]). Parameter estimates from some of these cases (e.g. cases i and iii) were closely similar to the standard model, but differed systematically in the remaining case (case ii; SI Fig. S3B-E; see also Discussion).

Taken together, these results demonstrate that our model accurately fit observers' behavioral responses in this five-alternative attention task. To our knowledge, this is the first demonstration of a model capable of predicting human choices in a multialternative attention task.

Endogenous cueing effects on sensitivity, criterion and bias

Does endogenous cueing of attention induce changes of sensitivity, changes of bias or both? Changes were twice as likely at the cued location (66.7%), as all of the other locations combined (opposite: 16.7%, adjacent ipsi/contra: 8.3%, Fig. 1D, left). We tested the effect of these different cue validities on modulations of sensitivity and bias, as estimated by the m-ADC model. Bias was quantified based both on the constant criterion (b_{CC}) and on the likelihood ratio (b_{LR}) measures (Methods). Cue validity was not different between cue-ipsilateral and cue-contralateral locations, and parameter values were strongly correlated and not significantly different between these locations (SI Fig. S4); these locations were treated as a single cue-"adjacent" location for these analyses (Methods).

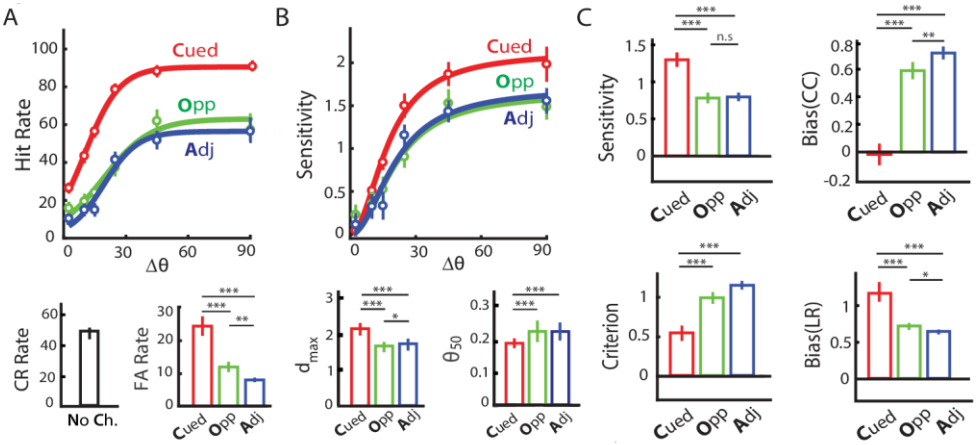


Figure 2. Sensitivity and bias changes induced by endogenous cueing of attention. A. (Top) Average psychometric function (n=30 subjects) showing hit rates across different change-angles (Δθ). Curves: Sigmoid fit. (Bottom) Mean correct rejection (left) and mean false-alarm rates (right). Error bars: jackknife s.e.m. Color conventions are as in Fig. 1B. **B**. (Top) Average psychophysical functions showing sensitivity at different change angles. Curves: Naka-Rushton fit. (Bottom) Asymptotic sensitivity (d_{max}, left) and change-angle at half-max sensitivity ($\Delta\theta_{50}$, right). Other conventions are as in panel A. **C.** Median sensitivity (top-left), criterion (bottom-left), constant criterion bias (b_{CC}, top-right) and likelihood ratio bias (b_{LR}, bottom-right) at different locations. Error bars: s.e.m. Color conventions are as in panel A.

First, we tested whether subjects were utilizing information provided by cue regarding the location of the imminent change. We quantified the effect of cueing "raw" performance metrics, including hit, false alarm and correct rejection rates (Fig. 2A). Subjects (n=30) exhibited highest hit rates for changes at the cued location as compared to the other locations, across the range of angles tested (Fig. 2A, top; p<0.001 for difference in hit rates between cued vs. opposite and cued vs. adjacent; Wilcoxon signed rank test, Benjamini-Hochberg correction for multiple comparisons), indicating that subjects were indeed heavily relying on the cue to perform the task. Nevertheless, false-alarm rates were also highest at the cued location (Fig. 2A, bottom right; cued vs. opposite: p=0.0014; cued vs. adjacent: p<0.001).We asked whether this pattern of higher hit rates concomitantly with higher false alarm rates occurred due to a higher sensitivity at the cued location, a higher bias (lower criterion) at the cued location, or both ((19, 21); Methods), by estimating these psychophysical parameters . To distinguish between these possibilities, we analyzed these psychometric data with a 4-ADC model.

Sensitivity was consistently highest at the cued location as compared to the other, uncued locations (p<0.001; Fig. 2B, top). We fit the 4-ADC model incorporating a parametric form of the psychophysical function (the Naka-Rushton function). Asymptotic sensitivity (d_{max}) was highest for the cued location as compared to each of the uncued locations (p<0.001, bootstrap test against a null distribution of differences generated by randomly shuffling

location labels; SI Methods; Fig. 2B, bottom left) whereas change magnitude corresponding to half-max sensitivity ($\Delta\theta_{50}$) was significantly lower for the cued, as compared to each of the uncued locations ($p<0.001$, bootstrap test; Fig. 2B, bottom right). Similarly, bias was uniformly highest at the cued location compared to the uncued locations: criteria and constant criteria were lowest, and likelihood ratio bias highest towards the cued location compared to the other locations (Fig. 2C; $p<0.001$; corrected for multiple comparisons).

Next, we asked if these modulations of sensitivity and bias reflected a benefit (relative to baseline) at the cued location, a cost at the uncued locations, or both. For this, a subset of the subjects (n=10) who were tested on the cued detection task were also tested on a neutrally-cued version of the task presented in interleaved blocks (Methods). In this task, changes were equally likely (25%) at each of the four locations. Within this pool of subjects, sensitivity for the neutrally cued locations was significantly lower than that at the cued location ($p=0.004$), but only marginally significantly different from that at the uncued locations ($p=0.049$; SI Fig. S5C, top left). On the other hand, biases (b_{CC} and b_{LR}) for the neutrally cued locations were intermediate between that of the cued and uncued locations, and robustly significantly different from cued ($p=0.002$) as well as uncued ($p<0.01$) locations (SI Fig. S5C, right).

In sum, these results show that endogenous cueing of attention produced both a higher sensitivity as well as higher bias towards the cued location relative to uncued locations. The increase in sensitivity manifested primarily as a benefit at the cued location. In contrast, cueing produced both a strong benefit in bias at the cued location and a strong cost at uncued locations, relative to baseline.

Sensitivity and bias: Same or different attentional mechanisms?

Are the enhancements of sensitivity and bias by endogenous cueing mediated by the same or different attentional mechanisms? To answer this question, we evaluated two competing models (Fig. 3A). According to the “common” mechanism model, cue-induced selection biases compete for sensory processing resources to enhance sensitivity; in this model sensitivity at each location co-varies systematically with bias at that location. On the other hand, according to the “disjoint” mechanisms model bias and sensitivity modulation are decoupled and independent of each other. We sought to distinguish between these two models by examining evidence from seven different approaches.

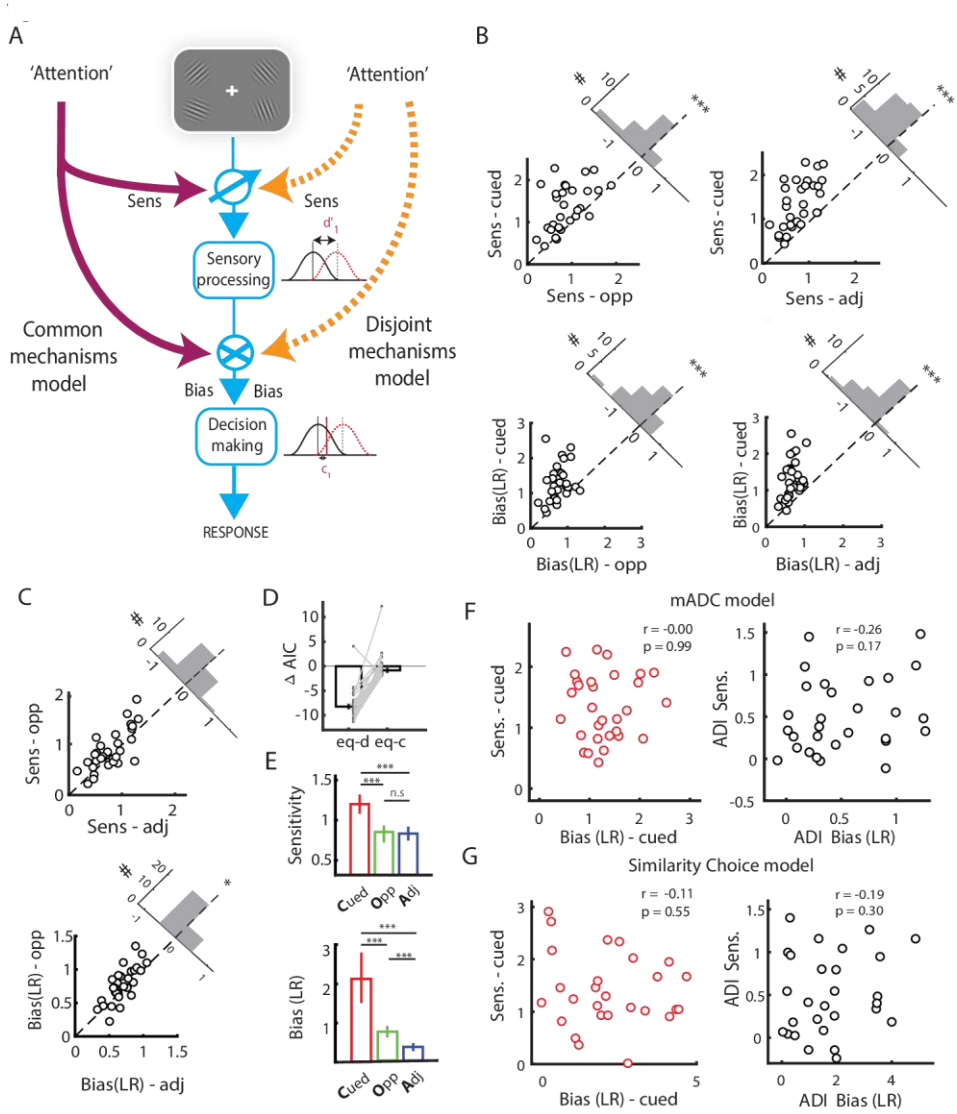


Figure 3. Common versus distinct mechanisms of sensitivity and bias changes. A. Schematic of common (left) versus disjoint (right) models of attentional modulation of sensory enhancement (sensitivity, d') and decisional gating (bias or criterion). Attention may modulate sensitivity and bias through a common mechanism (purple arrows) or through distinct mechanisms (orange arrows).**B.** Sensitivity (top) or bias (bottom) at the cued versus uncued (opposite or adjacent) locations. Upper right insets: Histograms indicating difference in parameter values. Data points: individual subjects. Dashed diagonal line: line of equality. **C.** Same as in B, but for the opposite versus adjacent locations. Other conventions are as in panel B. **D**. Model comparison analyses showing the difference in AIC score with the standard m-ADC model for two variants, one with equal sensitivities at uncued locations (d'_{Opp}= d'_{Adj}, left) and one with equal criteria at the uncued locations (c_{Opp} = c_{Adj}, right). Error bars: Standard error of the mean. **E.** Sensitivity (top) and bias (bottom) at cued, opposite and adjacent locations as estimated with the similarity choice model. Other conventions are as in Fig. 2C. **F.** (Left) Covariation between sensitivity and bias at the cued location. (Right) Covariation between attention difference indices of sensitivity and bias. Data points: individual subjects. **G.** Same as in panel F, but for parameters estimated with the similarity choice model.

First, we selected cue validity at the uncued locations to be at one of two different values (Fig. 1D, left). One-third of the change trials occurred at the uncued locations. Of these, changes were twice as likely at the cue-opposite location (16.7%) as at either of the cue-

ipsilateral or cue-contralateral ("adjacent") locations (8.3% each). We asked if these different cue validities would produce systematically differing values of sensitivity and bias at the different uncued locations. Across the population of subjects (n=30), sensitivities at the opposite location were not significantly different from sensitivities at the adjacent locations (p=0.08; Fig. 2C, top left, Fig. 3B-C upper panels). In contrast, constant criteria bias (b_{CC}) were lower for the opposite location compared to the adjacent locations (p=0.009; Fig. 2C, top right), and likelihood ratio bias (b_{LR}) was significantly higher for the opposite location compared to the adjacent locations (p=0.018; corrected for multiple comparisons). Taken together with the graded value of bias in the neutral cued condition (Fig. 2C, bottom right; Fig. 3B-C, lower panels; SI Fig. S5C, right), these results indicate that bias, rather than sensitivity, systematically modulated with endogenous cue validity.

Second, we evaluated two competing models against the standard m-ADC model. In the first model – m-$ADC_{eq\text{-}d}$ – we constrained sensitivities at the opposite and adjacent locations to be identical ($d'_{opp} = d'_{adj}$); criteria were free to vary across these locations. In the second model – m-$ADC_{eq\text{-}c}$ – we constrained criteria at the opposite and adjacent locations to be identical ($c_{opp} = c_{adj}$); sensitivities were free to vary across these locations. Model selection analysis based on the Akaike Information Criterion (AIC) revealed that the m-$ADC_{eq\text{-}d}$ model significantly outperformed both the m-$ADC_{eq\text{-}c}$ model and the standard m-ADC model (AIC: m-$ADC_{eq\text{-}d}$ = 500.1 [466.2-602.9]; m-$ADC_{eq\text{-}c}$ = 508.6 [475.5-614.2]; m-ADC = 508.7 [474.4-614.3]; p<0.001, Wilcoxon signed rank test), whereas there was no significant difference in the performance of the m-$ADC_{eq\text{-}c}$ and the standard m-ADC models (p=0.12, Fig. 3D). The significantly higher evidence in favor of a model that incorporated identical sensitivities, but distinct criteria, at all uncued locations confirms that criteria, rather than sensitivities, varied in a graded manner with endogenous cue validity.

Third, despite these trends in parameters at the uncued locations, we asked if subjects who showed higher sensitivity at the cued location would also have a higher bias (or lower criterion) at the cued location. Sensitivity and criteria were strongly positively correlated at the cued location (ρ=0.6, p<0.001, SI Fig. S6A) indicating that across observers' criteria co-varied with sensitivities, in line with a key prediction of the m-ADC model for optimal decisions in this task (SI Appendix: equation 11). On the other hand, neither constant criteria nor likelihood ratio bias were correlated with sensitivity (ρ=0.0-0.13, p>0.4; Fig. 3F, left). The uncued locations (opposite, adjacent) revealed a similar trend of positive correlation between sensitivity and criteria and no correlation (b_{CC}) or even a negative correlation (b_{LR}) between sensitivity and bias (SI Fig. S6A). These results demonstrate that although endogenous

cueing produced both the highest sensitivity and bias at the cued location, there was no evidence of a positive correlation between these quantities.

Fourth, although sensitivity and bias at the cued location did not covary, we asked if attentional modulation of sensitivity and bias, produced by endogenous cueing, would be correlated across subjects. In other words, would subjects who showed the greatest differential sensitivity at the cued location (relative to uncued locations), also show the greatest differential bias toward the cued location? To answer this question we tested whether sensitivity and bias modulations, measured either as a difference index (ADI) or a modulation index (AMI) across cued and uncued locations (Methods), were correlated. Neither bias value (constant criterion or likelihood ratio) was significantly co-modulated with sensitivity, as measured with the difference index ($p>0.17$; Fig. 3F and SI Fig. S6B), or with the modulation index ($\rho_{cc\text{-}d'}=-0.25$, $p=0.19$; $\rho_{LR\text{-}d'}=0.25$, $p=0.18$, SI Fig. S6C).

Fifth, although sensitivity and bias were not co-modulated across the population of subjects, we asked if sensitivity and bias would be co-modulated for individual subjects across experimental blocks. We divided data from each subject's experimental session into two subsets of contiguous experimental blocks. Difference indices (ADI) or modulation indices (AMI) for bias from each block were subjected to an n-way analysis of variance with the corresponding sensitivity (ADI or AMI) index a continuous predictor, and subjects as random effects. This analysis revealed that, neither measure of bias was significantly co-modulated with sensitivity, as measured by either modulation index, across blocks ($p>0.1$, for all tests).

Sixth, although average sensitivities were not significantly different between opposite and adjacent locations across the population, some observers exhibited a higher sensitivity at the opposite location than at adjacent locations ($d'_{Opp}>d'_{Adj}$, n=19/30), or vice versa ($d'_{Opp}<d'_{Adj}$, n=11/30). To test whether these differences in sensitivity were also reflected in the observers' relative biases toward these two locations, we performed the following analysis: We divided subjects into two groups depending on the relative value of their sensitivities at opposite and adjacent locations ($d'_{opp}>d'_{adj}$ or vice versa), and then tested for differences in constant criteria or likelihood ratio bias between opposite and adjacent locations within each group. This analysis did not reveal any distinct patterns in (either measure of) bias across the two groups (SI Fig. S7). Thus, inter-individual patterns of differential sensitivity toward the uncued locations (opposite vs. adjacent) did not translate into corresponding patterns of differential bias toward these same locations.

Finally, we examined the possibility that the graded variation of bias with cue validity was a consequence of parametric assumptions in the m-ADC model. Observers' false-alarm rates at each location (cued, opposite, adjacent) – a metric independent of the m-ADC model's assumptions – systematically covaried with cue validity (Fig. 2A, bottom right), indicating that the graded variation of bias with cue validity was not due to m-ADC model assumptions. Nevertheless, we sought to verify these trends with a "similarity choice model" (26), in which sensitivities and biases are estimated, not from a parametric, latent variable model, but by a factoring of the underlying probability densities. Consequently, this model does not make parametric assumptions about underlying latent variable distributions (see SI Methods for details of model fitting). Model comparison with the Akaike Information Criterion (AIC) revealed evidence overwhelmingly in favor of the more parsimonious m-ADC model (AIC median [95% CI]: choice theory=562.5 [519.6-657.1]; m-ADC=508.7 [474.4-614.3]; $p<0.0001$; paired sign rank test; SI Fig. S8A). Sensitivity and bias parameters estimated by the similarity choice model were closely correlated with the corresponding m-ADC model parameters (ρ=0.7-0.9, $p<0.001$; Fig. S8E), and also exhibited the same key trends, such as a graded variation in bias with cue-validity (Fig. 3E, bottom; SI Fig. S8B) and the lack of correlation between sensitivity and bias or their modulations (Fig. 3G; SI Fig. S8C-D).

To summarize, sensitivity enhancements were strongest at the cued location and not significantly different across uncued locations: bias modulations were more graded across locations and varied systematically with endogenous cue validity. This result suggests that the "spotlight" model of attention applies to primarily to sensitivity control, rather than to bias control, mechanisms. Moreover, modulations of sensitivity and bias by endogenous cueing were uncorrelated across subjects. These results overwhelmingly favor the hypothesis that sensitivity and bias are mediated by dissociable neural mechanisms during endogenous attention, favoring the "disjoint mechanisms" model (Fig. 3A, right).

Covariation of sensitivity and bias with decision dynamics

Attention produces systematic effects on decision dynamics, manifesting as faster reaction times (RT) when making perceptual decisions about stimuli at attended locations. To further disambiguate the "common" from the "disjoint" mechanisms models we asked if observers' RT correlated with sensitivity, with bias or with both. According to the "common mechanism" model, both sensitivity and bias would contribute in a closely similar manner towards determining behavioral reaction times. On the other hand, according to the "disjoint mechanisms" model sensitivity and bias would contribute independently to determining reaction times. In these analyses, we report correlations based on the likelihood ratio measure of bias (b_{LR}), as both measures of bias (b_{CC} and b_{LR}) showed very similar trends.

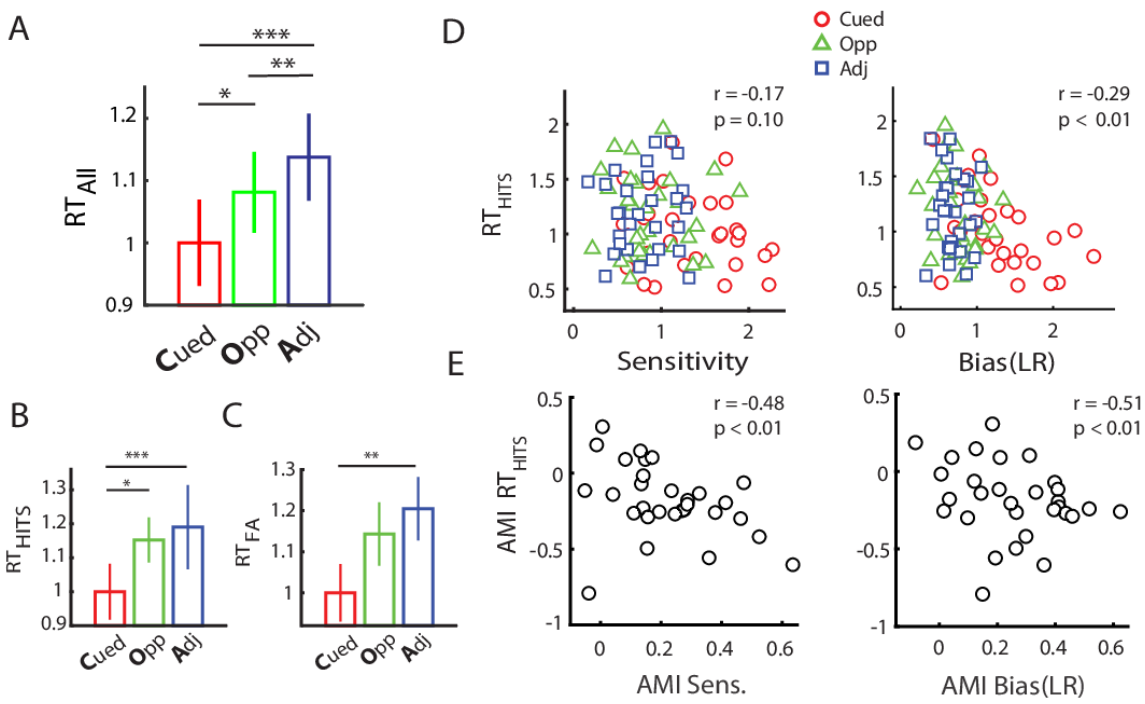


Figure 4. Covariation of sensitivity and bias with decision dynamics: A. Normalized median reaction times across subjects for each location (all trials). **B**. Same as in top panel, but for hit trials. **C**. Same as in top panel, but for false-alarm trials .**D**. Covariation of normalized median reaction times on correct trials for each subject at each location with sensitivity (left) or bias (right). Data points: individual subjects at particular locations; colors follow conventions as in Fig. 1B. **E**. Same as D, but covariation of the attention modulation indices for median reaction times with those for sensitivity (left) or bias (right). Data points: individual subjects. Other conventions are as in panel D.

First, we examined reaction times for all (change) responses made to each location (including both correct and incorrect responses). RT (normalized to cued location; SI Methods) varied in a graded fashion with cue validity; the lowest RTs occurred for changes at the cued location, followed by the opposite, and then, adjacent, locations ($p<0.05$, corrected for multiple comparisons; Fig. 4A). Similar graded trends were observed when the data were analyzed separately based on hits or false alarms (Fig. 4B - C). This graded variation with cue validity suggested a close relationship between RT and bias.

Next, we quantified the relationship between RT, sensitivity and bias, considering only correct responses (hits). RT (normalized) and bias were robustly negatively correlated (Fig. 4D, right); $\rho=-0.29$ $p=0.006$), whereas although RT showed a decreasing trend with d', their correlation was not significant (Fig. 4D, left); $\rho=-0.17$, $p=0.10$). To test if the correlation between RT and bias was rendered more robust by the latent covariation between RT and d', we performed a partial correlation analysis of RT versus bias controlling for the effect of d' (and vice versa). This analysis confirmed a significant negative partial correlation between

RT and bias, controlling for d' (ρ_p=-0.21, p=0.04); the partial correlation between RT and d', controlling for bias, was negative, but not significant (ρ_p=-0.19, p=0.07). These results indicate a strong relationship between RT and bias such that shorter reaction times occurred for decisions at locations of greater bias.

Finally, we quantified the co-modulation of RT with d' and bias across cued and uncued locations with the attentional modulation index (Methods). RT AMI was significantly negatively correlated with bias AMI (ρ=-0.51, p=0.004; Fig 4E, right), and with d' AMI (ρ=-0.48, p=0.007; Fig. 4E, left). Yet, a significant partial correlation remained between RT AMI and bias AMI, controlling for d' AMI (ρ_p=-0.44, p=0.016). Next, we performed a multilinear regression analysis of RT with change angle, average d' and bias as predictors (all predictors were identically normalized; SI Methods). The analysis revealed a significantly more negative regression coefficient for bias compared to d' ($\beta_{b\text{-}LR}$ = -0.082 [-0.21 0.05], $\beta_{d'}$ = -0.019 [-0.069 0.071] – median [95% CI]; p=0.046, bootstrap test for difference of magnitudes).

Do these correlations with RT simply reflect a motoric response bias, rather than a cueing induced choice bias, toward the cued location? According to the motoric bias hypothesis subjects provided faster and more impulsive responses to cued locations manifesting as higher false alarms and lower RTs at the cued location. The correlation of RT with bias was observed even when considering only correct response trials, arguing against this alternative hypothesis. Nevertheless, to rule out the hypothesis of motoric response bias we conducted the following analysis: We correlated false alarm rates in each of the six experimental blocks with the mean reaction time on false alarm trials in that block; this analysis was done separately, for the cued and uncued locations. A significant correlation would indicate that faster responses (due to a motor bias) produced more false alarms and, correspondingly, a higher bias. Contrary to this hypothesis, we found no significant correlation between RT and false alarm rates across blocks at any location, including at the cued location (cued: ρ=0.05, p=0.58; opp: ρ=0.13, p=0.18; adj: ρ=-0.15, p=0.10) these results were confirmed with an anova analysis with RT as the response variable and false alarms as continuous predictors with subjects as random factors. These results indicate that decisional bias induced by endogenous cue validity, rather than a motor bias, explained the systematic pattern of RT correlations.

Thus, attention modulated RT and bias in a closely similar manner. Moreover, bias variations far outperformed sensitivity variations in explaining the variation in RT across cued and uncued locations. Taken together, these results suggest that bias, rather than sensitivity,

modulations determine attention's reaction time effects, lending further support to the "disjoint mechanisms" model.

Sensitivity, bias and optimal decision metrics

Previous studies have hypothesized that attention enables optimal decision-making at the attended location (10). We evaluated this hypothesis using the m-ADC model, and tested if subject's decisional optimality would vary with sensitivity, with bias or with both.

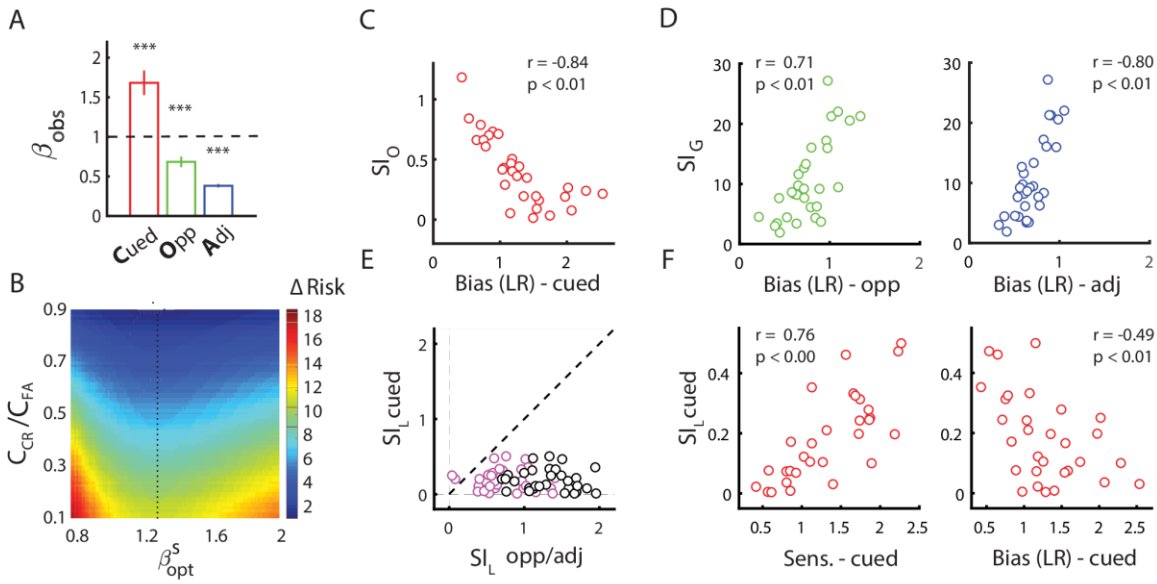


Figure 5. Covariation of sensitivity and bias with optimal decision-making. A. Median observed cost ratio (βobs) at the different locations. Dashed line: βopt value (=1) for maximizing correct and minimizing incorrect responses. Color conventions are as in Fig. 1B. **B.** Difference between the actual Bayesian risk and the optimal Bayesian risk (ΔR) for different values of β (x-axis) and different values of the cost ratio of false alarms to correct rejections (CFA/CCR; y-axis); data pooled across subjects. Warmer colors: larger values of ΔR. Black dots: β* corresponding to minimal ΔR for each CFA/CCR. **C.** Covariation of the objective sub-optimality index (SI_O) with bias at the cued location. Data points: Individual subjects. **D.** Covariation of the global sub-optimality index (SI_G) with bias at the opposite (left) or adjacent (right) locations. Other conventions are as in panel C. **E.** Covariation of the locational sub-optimality index (SI_L) at the cued (y-axis) and uncued (x-axis) locations. Black: SIL at cued versus opposite locations; magenta: SI_L at cued versus adjacent locations. Other conventions are as in panel C. **F.** Covariation of the locational sub-optimality index (SI_L) with sensitivity (left) and bias (right) at the cued location. Other conventions are as in panel E.

First, we measured subjects' observed cost ratio (β_{obs}), defined as the ratio of the prior odds ratio to the bias at each location (SI appendix: equation 12), and compared it to the optimal cost ratio (see SI Methods for details). The m-ADC model decision rule is based on minimizing Bayesian risk (or maximizing utility) and, in this framework, β_{opt} is equal to the ratio of the cost of correct rejections versus false alarms to the cost of hits versus misses ($\beta_{opt} = (C_{CR} - C_{FA})/ (C_{Hit} - C_{Miss})$; SI Appendix: equation 8). Therefore, subjects with a goal of

maximizing successes and minimizing errors, without assigning differential costs to different types of errors or successes, should have assumed an optimal cost ratio of unity (β_{opt}=1) at all locations (Fig. 5A, dashed horizontal line). Under these assumptions the bias at each location should match the prior odds ratio at that location. Yet, the vast majority of subjects exhibited systematic deviations from this optimum (Fig. 5A): the observed cost ratio was significantly greater than 1 at the cued location, and was significantly less than 1 at uncued locations ($p<0.001$, signed rank test; Fig. 5A), across subjects. Because the cost ratio is inversely related to bias (SI Appendix: equation 12), this implies that subjects adopted a lower bias than optimal at the cued location and a higher bias than optimal at uncued locations. We investigated the reason for this systematic pattern of sub-optimalities.

A first, potential scenario is that subjects' observed cost ratio deviated from the optimal ratio because they perceived a prior odds ratio at each location that differed systematically from the actual prior odds ratio. This could have happened because subjects failed to detect some proportion of changes, especially when the change in orientation was small. However, this explanation is not tenable for the following reason: small orientation changes would be difficult to detect at both cued and uncued locations. Hence, the perceived prior ratio would have had to be lower than the actual ratio at all locations. While this scenario can account for the lower than optimal bias at the cued location (Fig. 5A, red bar), it cannot account for the higher than optimal bias at each of the uncued locations (Fig. 5A, green and blue bars).

An alternative scenario is that subjects assumed different cost ratios (β-s) at the different locations, such that they judged errors arising from false alarms as more costly compared to errors arising from misses at the cued location ($\beta>1$), and vice versa at uncued locations ($\beta<1$). Nevertheless, no rational explanation can be readily conceived of for subjects attributing different costs to false alarms and misses at the different locations.

A third, more plausible, scenario is that each subject sought to minimize Bayesian risk by adopting a "subjective" cost ratio, one that was uniform across all locations, but differed from 1 ($\beta^{s}_{opt}\neq1$). This ratio was subjectively optimal to each observer considering her/his own estimate of the relative cost of false alarms and misses. We determined this β^{s}_{opt} for each observer as the cost ratio that minimized the difference between the actual Bayesian risk and the optimal Bayesian risk ($\Delta R^{obs-opt}$), parametrized by the ratio of false alarm to correct rejection costs (C_{FA}/C_{CR}; SI Methods). With data pooled across subjects, β^{s}_{opt} occurred at a value of 1.3, across the range of C_{FA}/C_{CR} values tested (Fig. 5B). We investigated how sensitivity and bias covaried with optimality metrics, in this, third, scenario.

First, we examined two metrics, which measured the deviation from optimality of each observer's overall performance: i) an objective sub-optimality index (SI_O) which measures the deviation of β^s_{opt} from β_{opt} (SI Methods), and ii) a global sub-optimality index (SI_G) which measures the deviation of the observed Bayesian risk from the optimal Bayesian risk ($\Delta R^{obs-opt}$). The two indices revealed complementary patterns of correlations with sensitivity and bias. SI_O was negatively correlated with bias at the cued location (ρ=-0.84, $p<0.001$; Fig. 5C) but not at the uncued locations ($p>0.07$). SI_O was not correlated with sensitivity at any location ($p>0.4$). SI_G was positively correlated with bias at both opposite and adjacent locations (ρ=0.7-0.8, $p<0.001$ Fig. 5D) but not at the cued location ($p>0.4$), and was negatively correlated with sensitivity at the opposite and adjacent locations (ρ=-0.4-0.6, $p<0.05$).

Next, we quantified the deviation from optimality of performance at each location. We defined a locational sub-optimality index (SI_L), which quantifies the how the cost ratio at each location (β_{obs}) deviated from each individual's optimal cost ratio (β^s_{opt}). Locational sub-optimality indices were significantly lower at the cued location than at either uncued location ($p<0.001$; Fig. 5E), indicating that subjects' decisions were more optimal at the cued location compared to uncued locations. Moreover, the locational sub-optimality index revealed opposite patterns of covariation with bias and sensitivity at the cued location: the sub-optimality index was negatively correlated with bias (ρ=-0.5, $p=0.006$), and was positively correlated with sensitivity (ρ=0.76, $p<0.001$; Fig. 5F).

To summarize, subjects who exhibited a greater bias toward the cued location, and lower bias toward uncued locations, made more optimal decisions overall in this attention task (Fig. 5C-D). Decisional optimality was highest at the cued location and showed opposite patterns of correlations with sensitivity and bias, as measured with the locational sub-optimality index (Fig. 5E-F). This dissociation between sensitivity and bias in terms of decision optimality further substantiates the "disjoint mechanisms" model.

Discussion

We demonstrate that perceptual (sensitivity) and decisional (bias) components of endogenous visuospatial attention, are not under the control of single, unitary mechanism,

but operate through dissociable mechanisms. Bias modulated systematically with endogenous cue validity (prior odds), was uncorrelated with sensitivity modulations, and correlated strongly with key decisional metrics, including reaction times and decisional optimality indices. On the other hand, sensitivity enhancements were strongest at the cued location, and not significantly different across uncued locations. The results suggest that the "unitary spotlight" model of attention is relevant chiefly to the sensitivity, rather than the bias, component of attention, and are consistent with the hypothesis that sensitivity control mechanisms allocate sensory processing resources in an all-or-none manner, whereas bias control mechanisms apportion decisional weights in a more graded manner.

Our findings have important implications for neuroscience studies of attentional mechanisms. Previous studies have revealed a plethora of attention's effects on neurons, including multiplicative, additive or no changes in neural firing; sharpening or widening of neural tuning functions; increases, decreases or no changes in neural correlations; increases, decreases or no change in neural synchrony; several of these effects are not readily reconciled (3, 5, 27–36). A potential explanation for these conflicting reports is that endogenous cueing of attention engages not a single, unitary process, but multiple processes, each of which potentially carries with it different behavioral or neural signatures. Our results show that at least two components of endogenous attention – sensitivity and bias modulation – are mediated by dissociable mechanisms, prompting an investigation into the distinct neural correlates of these sensory and decisional components of attention.

These insights are developed with a Bayesian decision model, the m-ADC model that quantifies changes in sensitivity as distinct from changes in bias induced by endogenous cueing. Perhaps the most significant contribution of our Bayesian m-ADC modeling approach is to resolve a critical misunderstanding, widely prevalent in literature, concerning the application of signal detection models to attention tasks. Previous studies of attention (8, 12, 23) model the subject's final decision as arising from a conjunction of multiple one-dimensional decisions; these are, in turn, based on the value of the decision variable relative to criterion, at the different locations. Nevertheless, this is not an appropriate formulation for the conventional attention task (shown in Fig. 1B) because: a) it assumes that the change/no change decision at each location is independent of the decisions at other locations; b) does not permit modeling interactions among the decision variable distributions at the different locations; and critically, c) does not specify a decision strategy for trials in which decision variable values exceed criteria at multiple (more than one) locations. Other studies have sought to sidestep this limitation by employing simpler models, but have had to factorize, *ad hoc*, some responses in the contingency table (e.g. false alarms), based on assumptions

that are not readily justified (1, 8). Our model corrects these pitfalls and develops a decision rule for minimizing Bayesian risk (or maximizing utility) across all stimulus response contingencies. The resultant decision surface is a multidimensional decisional manifold, which cannot, in principle, be reduced to multiple independent, one-dimensional decision models (i.e. independent criteria at each location). Consequently, the entire contingency table of responses is utilized, without ad hoc simplifications, when estimating sensitivities and criteria. The m-ADC model, therefore, represents a normative, and fundamentally novel approach for quantifying sensitivity and bias (or criterion) changes induced by attention.

Our findings resolve several decades of debate in a rich literature exploring attention's effects on sensitivity and bias.

Bachinski and Bacharach (1) were among the earliest to test the effects of spatial cueing of attention on sensitivity and bias, using dot stimulus detection task with probabilistic cueing (80% valid cues). They reported a benefit in sensitivity on the cued side, with little corresponding cost on the uncued side. Surprisingly, they found that bias was not different between cued and uncued locations. In direct contradiction to these findings, Shaw (6) reported that, in luminance detection tasks, focusing attention produced bias (criterion) changes without concomitant changes of sensitivity. Along similar lines, Muller and Findlay (8) demonstrated results similar to those of Shaw (6), showing that attention primarily produced changes of bias at the cued (relative to uncued) location in luminance detection tasks.

These contradictions can be readily explained by shortcomings in the psychophysical models used to analyze these three-alternative task designs. The analysis of Bashinski and Bacharach (1) suffered from at least two critical flaws: First, the authors analyzed a three-alternative detection task with two one-dimensional models and incorrectly grouped misidentification (mislocalization) responses with misses (SI Fig. S2). Second, they either pooled the false-alarm rates across cued and uncued locations, or partitioned the false-alarm rates according to an incorrect, ad hoc rule, a pitfall highlighted by other studies as well (2, 8). Both these factors likely resulted in incorrect estimates of sensitivity and bias. Similarly, Muller and Findlay employed a two-stage signal detection model, which was unable to take into account all 9 stimulus-response contingencies in their three-alternative task. Specifically, their estimates of d' were based only on hit and false-alarm contingencies; misses and misidentification responses were ignored. Nevertheless, misses and misidentification responses constitute a significant proportion of overall responses for each

event type (Fig.1E; SI Fig. S2). In general, analyses that ignore any category of response can produce inaccurate estimates of sensitivities and biases (SI Fig. S2).

To overcome the pitfalls of analyzing a three-alternative design with one-dimensional SDT models, Luck et al (2) employed a post-hoc response probe paradigm (following a protocol previously adopted by Downing (20)). In this paradigm subjects were cued to attend to one of four locations, and following a brief presentation of the target a response probe appeared. Subjects had to indicate whether or not a target had appeared at the location of the response probe. Cue validity was determined as the proportion of trials in which the attentional cue matched the location of the response probe. Despite multiple potential stimulus locations, the introduction of a response probe rendered this a 2-AFC (Yes/No) design that is amenable to analysis with a one-dimensional signal detection model. With this model, the authors found that endogenous cueing (central cueing) of attention produced systematic changes in sensitivity at cued versus uncued locations. Surprisingly, endogenous cueing did not reliably produce changes of bias. In contrast, the m-ADC model revealed a graded variation of bias across cued and uncued locations, with the highest bias at the endogenously cued location (Fig. 2C). Which result is correct?

These contradictory results can be readily explained within the framework of Bayesian decision theory, based on the respective task designs. In the Luck et al paradigm, also adopted by later studies (12, 15, 37) subjects' decision need be based only on sensory evidence at the response probe location, and there is no need to evaluate sensory evidence at any location (e.g. cued location) against evidence at any other location (e.g. uncued location). Moreover, because bias was calculated based on the likelihood ratio corresponding to the yes-no criterion, and because the prior probability of target appearance at each probed response location was 50% (cued or uncued), the likelihood ratio measure of bias was close to 1 at all locations (e.g. SI Appendix: equation 12), and not significantly different across locations. On the other hand, in the m-ADC paradigm with probabilistic cueing (Fig. 1A) subjects must directly compare sensory evidence at the cued and uncued locations when formulating their decision (e.g. SI Appendix: equation 14). Thus, the cued m-ADC paradigm, unlike the post-hoc response probe paradigm permits measuring competition for decisional bias, which is essential for measuring and quantifying a change of bias with cueing.

The need for an appropriate psychophysical model and task design is also highlighted by neuroscience studies that have sought to identify the neural correlates of sensitivity and bias modulation. For example, Chanes et al (23) tested the roles of different frequencies of brain

oscillations with rhythmic TMS over the right FEF, and reported frequency specific effects on sensitivity (at high-beta frequencies) and criterion (at gamma frequencies). They adopted a three-alternative task design in which subjects had to detect and report the location of a Gabor grating with one of three button presses (‘left’, ‘right’, ‘neither’). Again, because of the lack of an appropriate psychophysical model for analyzing this three alternative task, mislocalizations were entirely removed from the analysis as “error” responses, a pitfall that could lead to incorrect estimates of sensitivity and bias. Along similar lines, Luo and Maunsell (12) sought to determine if attentional modulation of neural activity in area V4 was correlated with sensitivity or bias changes while the monkey performed either a conventional three-alternative attention task (their Fig. 1A) or a delayed matched to sample task. The authors concluded that neural activity signatures, including an increase in the firing rate of V4 neurons, were correlated with sensitivity (signal-to-noise ratio) changes, but not with criterion changes. As before, sensitivity and bias were quantified with a combination of one-dimensional models. Interestingly, Baruni et al (9) reported results in direct contradiction to these findings. These authors manipulated absolute and relative rewards across different stimulus locations, and reported that neural modulation of V4 activity did not reflect the action of signal to noise (sensitivity) mechanisms. Again, these contradictions could perhaps be reconciled if the m-ADC model were employed to analyze animals’ behavior in these attention tasks.

It is increasingly clear that cognitive processes, like attention or decision-making, are not unitary constructs. Psychophysical models, like the m-ADC models, are essential to tease apart component mechanisms associated with these cognitive phenomena. Such models are increasingly relevant for mechanistic analyses of how cognitive processes emerge from neural activity and influence behavior, both in health and in disease.

Materials and Methods

Participants. 37 subjects (21 males, age range 19-60 yrs; median age 22yrs) with no known history of neurological disorders, and with normal or corrected-normal vision participated in the experiment. All participants provided written informed consent, and all experimental procedures were approved by the Institute Human Ethics Committee at the Indian Institute of Science, Bangalore. Seven subjects were excluded from analysis because of improper gaze fixation (see section on eye tracking). Data from 30 subjects were included in the final analysis.

Task. Participants were tested on a cued, five-alternative change detection task. The subject was seated in an isolated room, with head positioned on a chin rest 60 cm from the center of a contrast calibrated visual display (22-inch LG LCD monitor, X-Rite i1 Spectrophotometer). Stimuli were programmed with Psychtoolbox (version 3.0.11;(38)) using MATLAB R2014b (Natick, MA). Responses were recorded with an RB-840 response box (Cedrus Inc.). Subjects were instructed to maintain fixation on a central fixation cross during the experiment. Fixation was monitored with a 60 Hz infrared eyetracker (Gazepoint GP3).

Subjects began the task by fixating on a fixation cross at the center of a grey screen (0.5° diameter). Following 200 ms, four full contrast Gabor patches appeared, one in each visual quadrant (Fig. 1A) at a distance of 3.5° from the fixation cross (Gabor s.d.= 0.6°; grating spatial frequency, 2 cycles/degree). The orientation of each Gabor patch was drawn from a uniform random distribution, independently of the other patches, and pseudorandomized across trials. After another 500 ms, a central cue (directed line segment, 0.37° in length) appeared. After a variable delay (600 ms-2200 ms, drawn from an exponential distribution), the stimuli briefly disappeared (100 ms), and reappeared. Following reappearance either one of the four stimuli had changed in orientation, or none had changed. The subject had to indicate the location of change, or indicate “no change”, by pressing one of five buttons on the response box (configuration in SI Fig. S2).

We term trials in which a change in orientation occurred in one of the four patches as “change” trials, and trials in which no change in orientation occurred as “catch” or “no change” trials. 25% of all trials were no change trials, and the remaining 75% were change trials. We term the location toward which the cue was directed as the “cued” (C) location, the location diagonally opposite to the cued location, as the “opposite” (O) location, and two other locations as “adjacent-ipsilateral” (A-I) or “adjacent-contralateral” (A-C) locations, depending on whether they were in the same visual hemifield or opposite visual hemifield to

the cued location (Fig. 1 D-E). Changes occurred at the cued location on two-thirds of the change trials, at the opposite location on one-sixth of the change trials, and at each of the adjacent locations on one-twelfth of the change trials. Thus, the cue had a conditional validity of 67% on change trials, and an overall validity of 50%. The experiment was run in six blocks of 48 trials each (total, 288 trials per subject), with no feedback. In an orienting session prior to the experiment, subjects completed 96 trials (two blocks) with explicit feedback provided at the end of each trial about the location of the change and the correctness of their response. Data from these "training" blocks were not used for further analyses.

Ten out of the thirty subjects were tested on a version of the task that incorporated neutrally cued blocks. In this task, subjects were tested on a total of 8 experimental blocks (48 trials each; total 384 trials), with 4 blocks comprising predictively cued trials (as before) and the remaining 4 blocks comprised neutrally cued trials. On neutrally cued trials, the cue was made up of four directed line segments, each pointing toward one of stimuli in each of the four quadrants (Fig. 1A) and changes were equally likely at all four locations. Subjects were informed by on-screen instructions before the beginning of each block as to whether it was a predictive cueing or neutral cueing block, and the order of blocks were counterbalanced and pseudorandomized across subjects. All other training and testing protocols remained the same as before.

mADC psychophysical model: The goal of our psychophysical model was to dissociate sensory from decisional components of attention by analyzing each subject's 5x5 contingency table of responses obtained from the multi-alternative attention task (Fig. 1E). A description and mathematical derivation of the model are provided in the Supporting Information.

Data analysis.

Contingency tables. Subjects' responses in the task were used to construct 5x5 stimulus-response contingency tables, with change locations relative to the cue on the rows and response locations relative to the cue on the columns; no change events and responses were represented in the last row and last column respectively. Thus, each contingency table comprised five categories of responses: hits, misses, false-alarms, mislocalizations/misidentifications and correct rejections (SI Fig. S2). Since six values of orientation changes were tested at each location, the contingency table contained 100

independent observations (24 hits, 72 misidentifications and 4 false-alarms; the last category of responses do not depend on orientation change magnitude).

Psychometric and psychophysical functions. To compute the psychometric function (percent correct as a function of orientation change angle), we calculated the proportion of trials in which the subject detected and localized the change accurately; this was computed separately for each location (relative to the cue). Percent correct values, across all angles of orientation change, were fitted with a three parameter sigmoid function to generate the psychometric function (Fig. 2A, top). Psychometric functions in 24/30 subjects were estimated with a set of six change angles spanning 2° to 90° ([2,10,15,25,45,90]°), presented in an interleaved, pseudorandomized order across locations. As the psychometric function tended to saturate around 45°, in the remaining 6/30 subjects psychometric functions were estimated with a set of six change angles spanning 2° to 45°. The analyses were repeated by excluding this latter set of 6 subjects, who were tested on the more limited range of angles, and in all cases we obtained results similar to those reported in the main text. A single, combined psychometric curve was generated by pooling contingencies across all subjects and computing the above metrics for the pooled data (Fig. 2A, top). False alarm and correct rejection rates were calculated based, respectively, on subjects' incorrect and correct responses during no-change trials.

Model fitting and prediction analyses. To compute the psychophysical function (sensitivity as a function of orientation change angle), individual subjects' response contingencies were fitted with the m-ADC model described previously. We estimated sensitivities and criteria with maximum likelihood estimation (MLE), using a procedure described previously (24).The sensitivity is expected to change depending on the magnitude of stimulus strength (change angle value); hence, different sensitivity values were estimated for each change angle tested. On the other hand, the criterion at each location was estimated as a single, uniform value across change angles. As change angle magnitudes at each location were distributed pseudorandomly it is reasonable to expect that the subject could not anticipate and, hence, alter her/his decision criteria for different values of change angles. Thus, the model estimated 28 parameters (6 d' values for each of the 4 locations, and 4 criteria) from 100 independent observations in the contingency table.

We simplified the estimation further using the following approach: As the probability of change was identical across the adjacent-ipsilateral and adjacent-contralateral locations (Fig. 1B), we compared the parameters (sensitivities, criteria) estimated for these two locations. Sensitivities and criteria were strongly correlated (opp. vs. adj: d': ρ=0.58,

p=0.002; c: ρ=0.74, p=0.0001; SI Fig S4), and not significantly different between these locations (difference of ipsi. vs. contra medians: Δd'=0.07, Δc=0.03, p>0.5, Wilcoxon signed rank test). Thus, responses from cue ipsilateral and contralateral sides were averaged, and treated as a single "adjacent" contingency in the contingency table. Now, the m-ADC model estimated 21 parameters (6 d' values for each of the 3 locations – cued, opposite and adjacent – and their corresponding 3 criteria) from 63 independent observations in the contingency table. In all of the subsequent analyses we report psychophysical parameters calculated for these three locations relative to the cue, treating the two cue-adjacent locations (A-I, A-C) as a single "adjacent" location (A).

For the behavior prediction analysis, (Fig. 1F-G) model fitting was done with only subsets of contingencies (hits and false-alarms, hits and misses or missed and false-alarms), as described in the main text. The predicted contingency table was compared with the observed contingency table with a goodness-of-fit (randomization) test (see SI Methods: *Statistical tests, correlations and goodness-of-fit.*).

In addition, we computed two measures to quantify attentional modulation of these parameters induced by endogenous cueing: a difference index and a modulation index. The attention difference index for each parameter was computed as the difference between the values of the respective parameter at the cued and uncued locations (e.g. $d'_{ADI} = d'_{cued} - d'_{uncued}$) whereas the attention modulation index was computed as the ratio of the difference of the parameters at the cued and uncued locations to their sum (e.g. $d'_{AMI} = (d'_{cued} - d'_{uncued})/(d'_{cued} + d'_{uncued})$).

Computing psychophysical parameters and their modulations. After fitting the m-ADC model to the contingency table to estimate sensitivity and criteria we computed, at each location, the mean sensitivity (d'_{av}; average value across change angles), as well as two measures of bias: the constant criterion ($b_{CC} = c\text{-}d'_{av}/2$), and the likelihood ratio (b_{LR}; SI Appendix: equation 12). Both b_{CC} and b_{LR} are candidate measures for quantifying bias (19) and both exhibited similar patterns of variation in several of the analyses reported here. Occasionally, parameters are reported as a single value for uncued locations; these were based on their average values across uncued locations.

Psychophysical function fits. The psychophysical function was generated by fitting the sensitivity values across angles at each location with a three-parameter Naka-Rushton function allowing asymptotic sensitivity (d_{max}) and orientation change value at half-max ($\Delta\theta_{50}$) to vary as free parameters, keeping the slope parameter (n) fixed at 2; a value determined

from pilot fits to the data (39). As before, a single combined psychophysical curve was generated by pooling contingencies across all subjects and computing the above metrics for the pooled data (Fig. 2B, top). Values of d_{max} and $\Delta\theta_{50}$ reported in the main text correspond to pooled psychophysical fits with jackknife error bars (Fig. 2B, bottom left and right). Tests of significant differences in d_{max} and $\Delta\theta_{50}$ values across locations were performed with a bootstrap approach: by comparing the true differences in values against a null distribution of differences calculated from contingency tables obtained by randomly shuffling, 1000 times, the labels across stimulus and response locations (cued, opposite, adjacent).

Further details regarding eye-tracking, goodness-of-fit tests, robust correlations (40), model comparison analysis, analysis of reaction times and optimal decisions are presented in SI Methods.

Acknowledgments

The authors would like to thank Nicholas Steinmetz, Behrad Noudoost and Kelsey Clark for their comments on a preliminary version of this manuscript, and Subbulakshmi Sankaranarayanan for help with data collection. This research was funded by a Wellcome Trust-Department of Biotechnology India Alliance Intermediate fellowship, a Science and Engineering Research Board Early Career award, a Pratiksha Trust Young Investigator award, a Department of Biotechnology-Indian Institute of Science Partnership Program grant, and a Tata Trusts grant (all to DS).

References

1. Bashinski HS, Bacharach VR (1980) Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Percept Psychophys* 28(3):241–248.

2. Luck SJ, Hillyard SA, Mouloua M, Hawkins HL (1996) Mechanisms of visual–spatial attention: Resource allocation or uncertainty reduction? *J Exp Psychol Hum Percept Perform* 22(3):725–737.

3. McAdams CJ, Maunsell JHR (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19(1):431–441.

4. Ungerleider LG, Kastner S (2000) Mechanisms of visual attention in the human Cortex. *Annu Rev Neurosci* 23(1):315–341.

5. Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12(12):1594–1600.

6. Shaw, ML (1984) Division of attention among spatial locations : A fundamental difference between detection of letters and detection of luminance increments. *Atten Perform X Control Lang Process*: 109–121.

7. Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science*, 229, pp 342-345.

8. Baruni JK, Lau B, Salzman CD (2015) Reward expectation differentially modulates attentional behavior and activity in visual area V4. *Nat Neurosci* 18(11):1656–1663.

9. Müller HJ, Findlay JM (1987) Sensitivity and criterion effects in the spatial cuing of visual attention. *Percept Psychophys* 42(4):383–399.

10. 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.

11. Krauzlis RJ, Bollimunta A, Arcizet F, Wang L (2014) Attention as an effect not a cause. *Trends Cogn Sci* 18(9):457–464.

12. Buschman TJ, Kastner S (2015) From behavior to neural dynamics: An integrated theory of attention. *Neuron* 88(1):127–144.

13. Luo TZ, Maunsell JHR (2015) Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron* 86(5):1182–1188.

14. Hawkins HL, Hillyard SA, Luck SJ, Cathryn J (1990) Visual attention modulates signal detectability. *J Exp Psychol Hum Percept Perform* 16(4):802–811.

15. Pestilli F, Carrasco M, Heeger DJ, Gardner JL (2011) Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. *Neuron* 72(5):832–846.

16. Carandini M, Churchland AK (2013) Probing perceptual decisions in rodents. *Nat Neurosci* 16(7):824–831.

17. Green, DM, & Swets JA (1966) *Signal detection theory and psychophysics.* (John Wiley and Sons, New York).

18. Ashby FG (1992) Perceptual and decisional seperability. *Multidimensional models of perception and cognition* (Lawrence Erlbaum Associates, Inc, Hillsdale, NJ).

19. Macmillan NA CC (2005) *Detection theory: A user's guide* (Lawrence Earlbaum Associates., Mahwah, NJ).

20. Downing CJ (1988) Expectancy and visual-spatial attention: effects on perceptual quality. *J Exp Psychol Hum Percept Perform* 14(2):188–202.

21. Steinmetz NA, Moore T (2014) Eye movement preparation modulates neuronal responses in area V4 when dissociated from attentional demands. *Neuron* 83(2):496–506.

22. Carrasco M (2011) Visual attention: The past 25 years. *Vision Res* 51(13):1484–1525.

23. Chanes L, Quentin R, Tallon-Baudry C, Valero-Cabré A (2013) Causal frequency specific contributions of frontal spatiotemporal patterns induced by non-invasive neurostimulation to human visual performance. *J Neurosci* 33(11):5000–5005.

24. Sridharan D, Steinmetz NNA, Moore T, Knudsen EI (2014) Distinguishing bias from sensitivity effects in multialternative detection tasks. *J Vis* 14(9):16.

25. Sridharan D, Steinmetz NA, Moore T, Knudsen EI (2017) Does the superior colliculus control perceptual sensitivity or choice bias during attention? Evidence from a multialternative decision framework. *J Neurosci* 37(3):480–511.

26. Luce RD (2005) Signal detectibility theory. *Individual choice behavior: A theoretical analysis* (Courier Corporation).

27. Spitzer H, Desimone R, Moran J (1988) Increased attention enhances both behavioral and neuronal performance. *Science* 240(4850), 338-340.

28. Connor CE, Gallant JL, Preddie DC, Van Essen DC (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. *J Neurophysiol* 75(3):1306–1308.

29. Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14(9):744–751.

30. Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26(3):703–714.

31. Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315(5820), 1860-1862.

32. Zénon A, Krauzlis RJ (2012) Attention deficits without cortical neuronal deficits. *Nature* 489(7416):434–437.

33. Moore T, Zirnsak M (2017) Neural mechanisms of selective visual attention. *Annu Rev*

Psychol 68(1):47–72.

34. Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27(1):611–647.

35. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nat Rev Neurosci* 2(8):539–550.

36. Cohen MR, Maunsell JHR (2011) Using neuronal populations to study the mechanisms underlying spatial and feature Attention. *Neuron* 70(6):1192–1204.

37. Wyart V, Nobre AC, Summerfield C (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc Natl Acad Sci* 109(9):3593–3598.

38. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.

39. Herrmann K, Montaser-Kouhsari L, Carrasco M, Heeger DJ (2010) When size matters: attention affects performance by contrast or response gain. *Nat Neurosci* 13(12):1554–1559.

40. Pernet C, Wilcox R, Rousselet G (2013) Robust correlation analyses: false positive and power validation using a new open source Matlab Toolbox. *Front Psychol* 3:606.

Supporting Information

SI – Methods

Model description. Signal detection theory provides a rigorous framework for distinguishing changes in the quality of perceptual representations (perceptual sensitivity) from changes in the relative weighting of sensory evidence at different locations during decisions (spatial choice bias). Sensitivity is quantified with the index of discriminability (d'), that measures the overlap of signal and noise distributions at each location (here, change versus no-change distributions), whereas bias is quantified with a decision threshold or criterion (c), that indexes the amount of evidence that must be available at a particular location before the subject decides that the event of interest (here, the change) occurred at that location. Conventional, one-dimensional SDT models do not suffice to quantify sensitivity and bias in multi-alternative attention tasks; the reasons are summarized in the Introduction (Fig. 1B-C) and elaborated in the Discussion (also see (24)).

We recently developed a multidimensional extension to SDT to quantify sensitivity and bias in attention tasks involving multiple alternative detection/change-detection (m-ADC tasks). Here, we extend this previous model by deriving, *de novo*, an optimal decision rule for task designs that employ the method of constant stimuli. In these tasks performance is measured by presenting stimuli at various, unpredictable strengths (e.g. contrast, or magnitude of orientation change) at each location, for example, attention tasks in which the full psychometric function is measured at both cued and uncued locations. The decision rule is constructed from optimal decision theory for minimizing Bayesian risk, or maximizing Bayesian utility, and the resultant optimal decision manifold is shown in Figure1C. Here, we present an intuition for how the model estimates sensitivity and bias from the 5x5 contingency table.

In our task, the subject has to detect an orientation change that can occur at any one (or none) of four locations, and indicate which of the five possible events occurred. This entails making a decision by comparing perceptual evidence for the change event at each of the four locations as well as the no change event. Five categories of stimulus-response contingencies occur in the 5x5 contingency table (SI Fig. S2): i) hits, wherein the subject correctly localizes the change; ii) mislocalizations, wherein the subject correctly reports a change event, but localizes it to an incorrect location; iii) misses, wherein the subject misses the change event and reports no-change; iv) correct rejections, wherein the subject correctly reports a no-change event and v) false alarms, wherein the subject indicates change at a location, when no changes occurred at any location. In fitting the m-ADC model, we compute

the proportion of these different categories of responses for a subject, for each location. The proportion of hits at each location would be indicative of the subject's perceptual sensitivity to changes at that location. Nevertheless, an accurate estimate of sensitivity would require factoring in the inherent, as well as cueing-induced, biases a subject might adopt as a part of her/his decision-making strategies. The proportion of false alarms at each location is indicative of these biases. The m-ADC model provides formal equations relating sensitivities and criteria at each location to the proportions of each type of response. The m-ADC model parameters, which include perceptual sensitivities at and decision criteria, toward each location are estimated from the 5x5 contingency table with maximum likelihood estimation. Biases are computed from these sensitivities and criteria using constant criterion and likelihood ratio measures described in the section on *Computing psychophysical parameters and their modulations*.

Eye-tracking. Subjects' gaze was binocularly tracked and the deviation in their gaze from the fixation cross was recorded and stored in degrees. Trials in which the eye-position deviated by more than 2 degrees from fixation either in the x- or in the y-direction, from the onset of the Gabors until the final response, were removed from further analysis. All of our subjects were South Asian and several exhibited dark pigmentation of the iris, rendering it nearly indistinguishable from the pupil. Hence, the contrast of the pupil (relative to the iris) was weak, and the tracker occasionally lost the location of the pupil; trials in which this occurred for more than 100 ms continuously were also excluded from the analysis. Finally, we excluded all subjects for whom the combined rejection rate (from eye deviation and lost tracking) was more than 30% (7/37 subjects). Among the set of subjects who remained (n=30), the rejection rate was 8.6% [4.2-17.7%]. We also confirmed with a randomization test, based on the chi-squared statistic (see below), that these rejected trials did not significantly alter the distribution of responses in the contingency table for any subject (p-values for distributions before and after rejection: 0.99, median across subjects).

Justification for the method of constant stimuli versus using a single, constant stimulus strength. In this study, we chose to estimate sensitivity, bias (criteria) and their modulations using the method of constant stimuli, by presenting different (psuedorandomized) values of orientation changes across trials. Other studies (1,13) have employed paradigms where the stimulus is presented at a single fixed strength determined from a preliminary staircasing procedure that ensures a uniform level of performance across subjects. Nevertheless, while such staircasing is possible for 2-AFC tasks, it is not readily available for 2-ADC tasks. In 2-AFC tasks, percent correct covaries directly with sensitivity and can be used as a proxy for adaptively increasing or decreasing stimulus strength based

on performance. On the other hand, in 2-ADC tasks, percent correct varies with both sensitivity and criteria (their Fig. 1G), and staircasing requires MLE estimation of sensitivity from a running contingency table in real time, a significant technical challenge. In addition, adopting the method of constant stimuli enabled us to generate the entire psychometric function, and provided sufficient variation in the sensitivity and criterion parameters to measure their correlations.

Statistical tests, correlations and goodness-of-fit. Unless otherwise stated, pairwise comparisons of different parameter values (e.g. mean d' or bias) across locations was performed with the non-parametric Wilcoxon signed rank test with FDR (Benjamini Hochberg) correction for multiple comparisons. Correlations across different parameter values or modulation indices were typically Pearson correlations. For the optimal behavior, choice theory and reaction time analyses, robust correlations ("percentage-bend" correlations) were computed to prevent outlier data points dominating the correlations. Goodness-of-fit of the model to the data was assessed using a randomization test based on the chi-squared statistic; the procedure is described in detail elsewhere (25). A small p-value (e.g. $p<0.05$) for the goodness-of-fit statistic indicates that the observations deviated significantly from the model.

To test if there was covariation between sensitivity and bias within each subject's data, we adopted the following procedure: Two contingency tables were constructed for each subject, with responses drawn from the first half of her/his respective session (first one half of the trials) and the second half (last one half of trials). Psychophysical parameters were estimated from these two subsets of data yielding two measures of each psychophysical parameter per subject. An n-way anova analysis was performed with each measure of bias (b_{CC}, b_{LR}) as the response variable and sensitivity as a continuous predictor, with subjects as random effects.

Model comparison analysis. We compared the m-ADC model against other candidate models. Two models (m-ADC$_{eq\text{-}d}$, m-ADC$_{eq\text{-}c}$ and m-ADC$_{obl}$) were modified from the m-ADC model but incorporated simpler or more complex assumptions. We also compared the m-ADC model with a similarity choice theory model. In each case, models were compared with the Akaike Information Criterion (AIC) that represents a trade off between model complexity (the number of fitted model parameters) and goodness-of-fit, (based on the log-likelihood function); a lower AIC score represents a better candidate model. Similar results were obtained when using a Bayesian Information Criterion (BIC).

i. Comparison with models assuming identical parameter values at uncued locations.
We compared the goodness of fit of our default model with two other models, one that assumed equal sensitivities at all uncued locations (m-ADC$_{eq-d}$ model; $d'_{opp} = d'_{adj}$) and one that assumed equal criteria at all uncued locations (m-ADC$_{eq-c}$ model; $c_{opp} = c_{adj}$). Because d' was estimated separately for each of the six change angles, the number of fitted parameters in the m-ADC$_{eq-d}$ model reduced to 15 (from 21 for the standard m-ADC model) and to 20 in the m-ADC$_{eq-c}$ model.

ii. Comparison with a model that differed in the orientation of oblique decision surface. The standard m-ADC model that we fit to our data assumes an oblique decision boundary -- for distinguishing signals (changes) between every pair of locations -- that is oriented at 45 degrees to each decision variable axis (Fig. 1C, thick grey line). Such a decision boundary is exactly optimal when sensitivities at every location are identical with each other, but also provides an excellent approximation to the non-linear optimal decision boundary, as shown in SI Fig. S1B. On the other hand, when sensitivities are significantly different across locations (e.g. cued versus uncued), the tilt of the optimal decision boundary reflects the relative values of the sensitivity at the two locations (25) and increases as the sensitivities at the two locations become more disparate (SI Fig. S1B, dashed lines). We term this the m-ADC$_{obl}$ model. We tested, using AIC measures, whether our subjects' behavior followed the default decision boundaries in the standard m-ADC model or were better fit with the m-ADC$_{obl}$ model. No consistent differences were found between the two model types across the population of subjects (AIC: m-ADC=508.7, m-ADC$_{obl}$=507.3; p=0.18). Thus, there was no clear evidence favoring either model, and all analyses reported here were based on fitting the standard m-ADC model.

iii. Comparison with a similarity choice theory model. A model based on similarity choice theory was used to concurrently estimate sensitivity and bias from the contingency tables. The model is based on factorization of the probability densities into sensitivity and bias components. Specifically, the joint probability densities are factored as:

$$p_{ij} = \eta_{ij}\, \beta_j \,/\, \Sigma_k \eta_{ik}\, \beta_k$$

where p_{ij} denotes the proportion of responses to location j when an event (e.g. change) occurred at location i, η_{ij} is a symmetric measure of similarity between signals at two locations (i,j) ($\eta_{ij} = \eta_{ji}$) and β_j represents the choice bias for location j. These quantities are defined on a ratio scale such that $\eta_{ii} = 1$ and, β_j is a choice bias relative to its value for the no change response ($\beta_o = 1$). To compare the values of these parameters against the m-ADC model (Fig. S9D), we defined: i) a sensitivity parameter (d_j) for each location and angle as

being inversely related to the respective similarity parameter η_{io} on a logarithmic scale (d_i=-log(η_{io})); ii) a bias parameter (β_j) defined as the geometric mean of the bias parameter across all angles; and iii) a constant criterion parameter defined as the negative of the logarithm of the bias (cc_i=-log(β_i)). Thus, the choice theory model was fitted with a total of 39 parameters (6 unique η_{ij}-s for each of the six angles, and 3 β_j parameters).Parameter estimates were strongly correlated between the m-ADC and choice theory models (SI Fig. S8E). Nevertheless, the significant difference in the number of parameters between the models resulted in the m-ADC model far outperforming choice theory in model comparison analyses based on AIC (SI Fig. S8A).

Analysis of reaction times*.* Reaction times were computed as the time from change onset to the time of response; no limit was placed on permitted response intervals, but subjects were asked to respond as quickly as possible. For each subject, trials in which RTs fell outside three standard deviations from the mean RT was considered outliers and excluded from further analysis. Each subjects' mean RT at each location was normalized by dividing by the median RT at the cued location across the population, and correlated with psychophysical parameter estimates at the respective location. Partial correlations were performed to identify dependencies between RT and measures of bias while controlling for the effects of d', and vice versa. We also performed multilinear regression analysis with RT as the response variable and change angle, mean sensitivity and bias at each location as predictors; all predictors were scaled to zero mean and unit variance before the analysis. Regression coefficient magnitudes for sensitivity ($\beta_{d'}$) and bias ($\beta_{b\text{-}LR}$) were compared with a bootstrap test; as our hypothesis was that the regression coefficient for bias would be higher than that for d' because of the similar graded variation of bias and RT across locations, we constructed a null distribution of regression coefficient differences by shuffling the location labels for the d' and bias values randomly and independently across subjects. In order to discount the hypothesis of motor bias, robust correlations were computed between mean RTs and mean FA rates for each experimental block after subtracting the mean RT and FA values subject-wise to account for subject specific effects. An n-way anova was applied, as before, treating RT as the response variable with FA rates as continuous predictors, and subjects as random factors.

Analysis of optimal decisions*.* Optimal decisions, in the m-ADC model framework, seek to minimize Bayesian risk (SI Appendix: equation 3). Under certain assumptions (e.g. $C^l_k = C^l_m$, $m \neq l$ or for a given stimulus event type, the cost of all types of response error are identical) the optimal decision surfaces comprise hyperplanes in the multidimensional decision space. Our model fitting analysis suggests that such a model provides an excellent fit to observers'

behavior in this multialternative attention task (Fig. 1F, bottom insets), suggesting that observers decision criteria were closely in line with this optimal family of decision surfaces (as defined in SI Appendix: equation 12,13 and 14).

One potential definition of optimality is when the subject seeks to maximize the number of correct responses and minimize the number of errors. In this case, cost ratio at location j, $\beta_{obs\text{-}j} = (C^0_0 - C^0_j) / (C^j_j - C^j_0) = (C_{CR} - C_{FA}) / (C_{Hit} - C_{Miss}) = 1$ (where C^x_y is the cost of responding to location x when the event occurred at location y). Yet, we noticed that β_j, calculated as $\beta_j = (p_j / p_i)/b_{LR\text{-}j}$ (SI Appendix: equation 8, 10 and 11), systematically deviated from 1 across the population of subjects (Fig. 5A), suggesting that subjects did not assume an equal relative cost for the two kinds of error responses (false alarm vs. correct reject and misses vs. hits). We also noticed that β_{obs} was different for the different locations, being predominantly greater than 1 at the cued location and less than 1 at uncued locations (Fig. 5A). Since there is only one type of correct rejection response, and since the model assumes that the cost for false alarms to each location is identical (SI Methods, assumption in SI Appendix: equation 7), one possibility is that subjects assumed a different relative cost of hits to misses at the different locations (cued, opposite, adjacent), a somewhat implausible assumption; in the optimality analysis section of the Results we evaluate this, and other scenarios. Rather, we propose that the reason for the difference in β_{obs} across the different locations is because the subject assumed a single, common cost ratio (β^s_{opt}) at all locations, but deviated from this (subjectively) optimal ratio at some locations where she/he did not consider it necessary to perform optimally. Note that the family of decision surfaces under the m-ADC model remains unchanged under these assumptions (SI Appendix: equation 13 and 14).

To determine this subject-specific β^s_{opt} we tested different values of β and selected the one that minimized the deviation of the actual Bayesian risk from the expected Bayesian risk assuming a single β^s_{opt} at all locations ($\Delta Risk^{obs\text{-}opt}$); again this deviation is zero if the subject did not deviate from β_{opt} at any location ($\Delta Risk^{obs\text{-}opt} = \Sigma_i \Sigma_j C^i_j p_{obs}{}^j_i - p_{opt}{}^j_i$). Following some algebra, it is clear that C^i_j for hits and misses (or misidentifications) is a function of β_{opt}, C_{FA} and C_{CR} (see SI Appendix: equations 6, 8 and 9, along with the following assumptions: a) $C^j_j = C^0_0/\beta_{opt}$; b) $C^j_0 = C^0_j/\beta_{opt}$; also $C^j_i = C^j_0$ for all $i \neq j$) whereas $p_{obs}{}^j_i$ is a function of β_j and d_i, and $p_{opt}{}^j_i$ is a function of β_{opt} and d_i.

For each subject three indices of optimal performance were computed. Two indices quantified the optimality of overall performance: i) an objective sub-optimality index (SI_O)

defined as the deviation of the subject's cost ratio from an objectively optimal cost ratio (β_{opt}=1) for maximizing correct responses, and computed as the magnitude of the logarithm of the ratio of β^{s}_{opt} to β_{opt} (SI_O= |log(β^{s}_{opt} /β_{opt})|)and ii) a global sub-optimality index (SI_G), defined and computed as the magnitude of the deviation of the optimal Bayesian risk from the actual Bayesian risk (SI_G = $\Delta Risk^{obs-opt}$). A third, locational sub-optimality index was defined as the deviation of the cost ratio at each location from the subject's own optimal cost ratio, and computed for each location(SI_L) as the magnitude of the logarithm of ratio of β_{obs} (observed cost ratio) at that location and the subject's own β^{s}_{opt} (SI_L = |log (β_{obs}/ β^{s}_{opt})|). These sub-optimality indices were correlated with sensitivities, biases and their modulations using robust correlations.

Supporting figures:

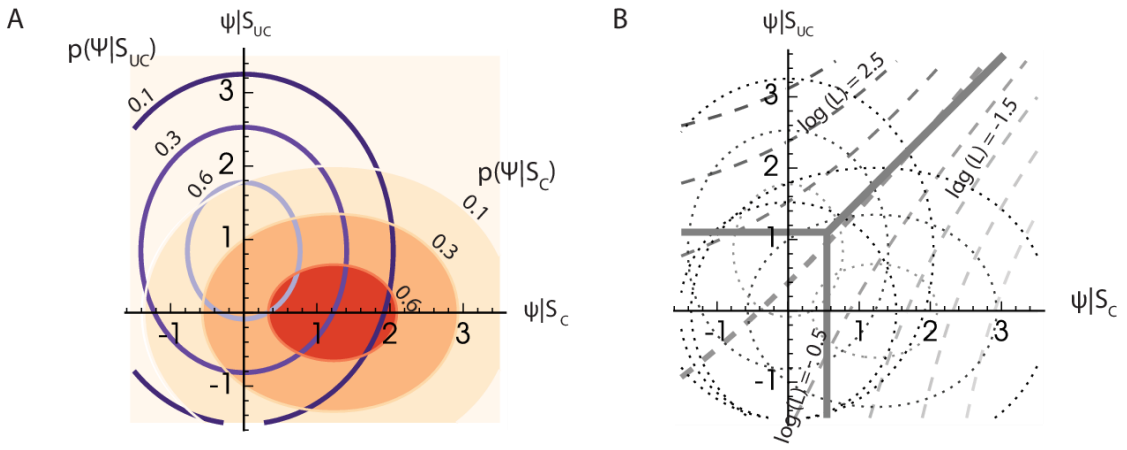
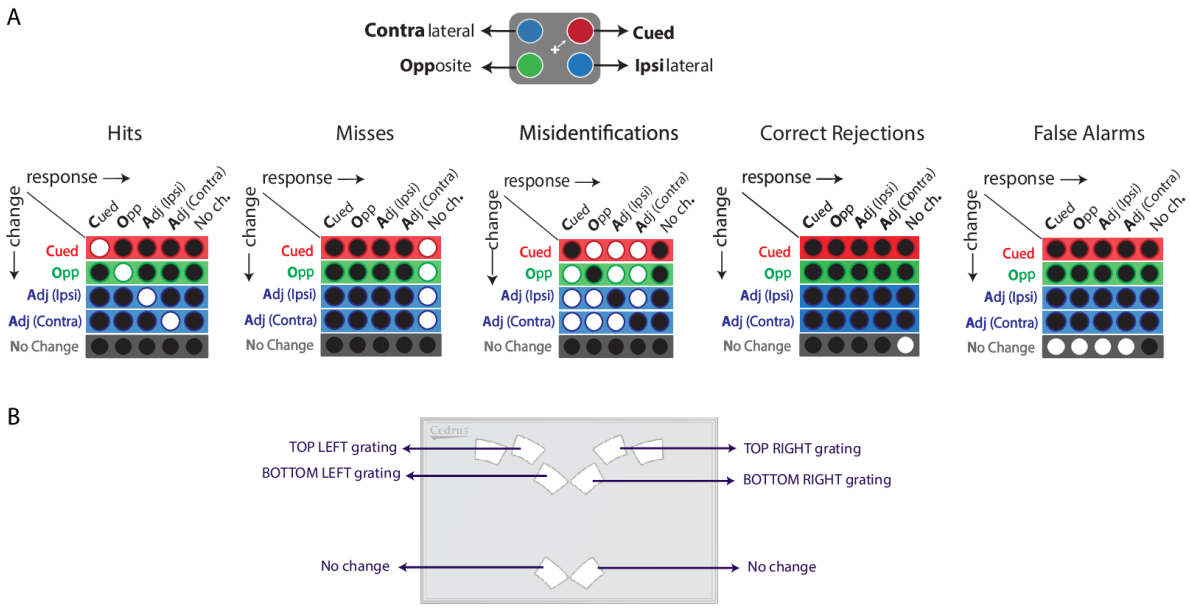


Figure S1: Decision variable distributions and decision surfaces in the m-ADC model. A. Contours of bivariate decision variable distributions (joint probability density function, pdf) in a task employing the method of constant stimuli in which stimuli or events (e.g. changes) can occur at one of multiple different strengths at the cued location (x-axis) or at an uncued location (y-axis). In the task shown in Fig.1A (main text) this corresponds to different magnitudes of orientation changes, that are employed to measure psychometric functions as shown in Fig. 2A (main text). Red, filled contours: pdf contours for changes at the cued location; blue, unfilled contours: pdf contours for changes at an uncued location. Plots are based on average psychophysical functions calculated from experimental data (Fig. 2B; main text). **B**. Optimal decision surfaces in the m-ADC model. Dashed lines: Contours of constant log-likelihood ratio (log (L)) of the two decision variable distributions shown in panel A. Thick solid lines: Planar decision surfaces in the m-ADC model. The oblique decision surface closely approximates the optimal decision contours of constant log-likelihood. Thick dashed line: Representative optimal decision contour (log(L) = 0.5). Dotted lines: Decision variable pdf contours from panel A, shown in the background, for reference. Plots are based on average criteria calculated from experimental data (Fig. 2B; main text).



A. Figure S2: A. Stimulus-response contingencies in a five-alternative attention task involving change detection. Schematics showing the different stimulus response contingencies: left to right - 'Hits' – response indicating the correct location of change; 'misses': no-change response on change trials; 'misidentifications': response indicating one location when the change happened at another location; 'correct rejections': correct no-change response, and 'false alarms': response indicating some location of change during no-change trials. **B**. Schematic showing the arrangement of buttons in the response pad (Cedrus – RB 830). The four top inner buttons correspond to the four possible locations of change. Either of the bottom two buttons can be pressed to indicate no change.

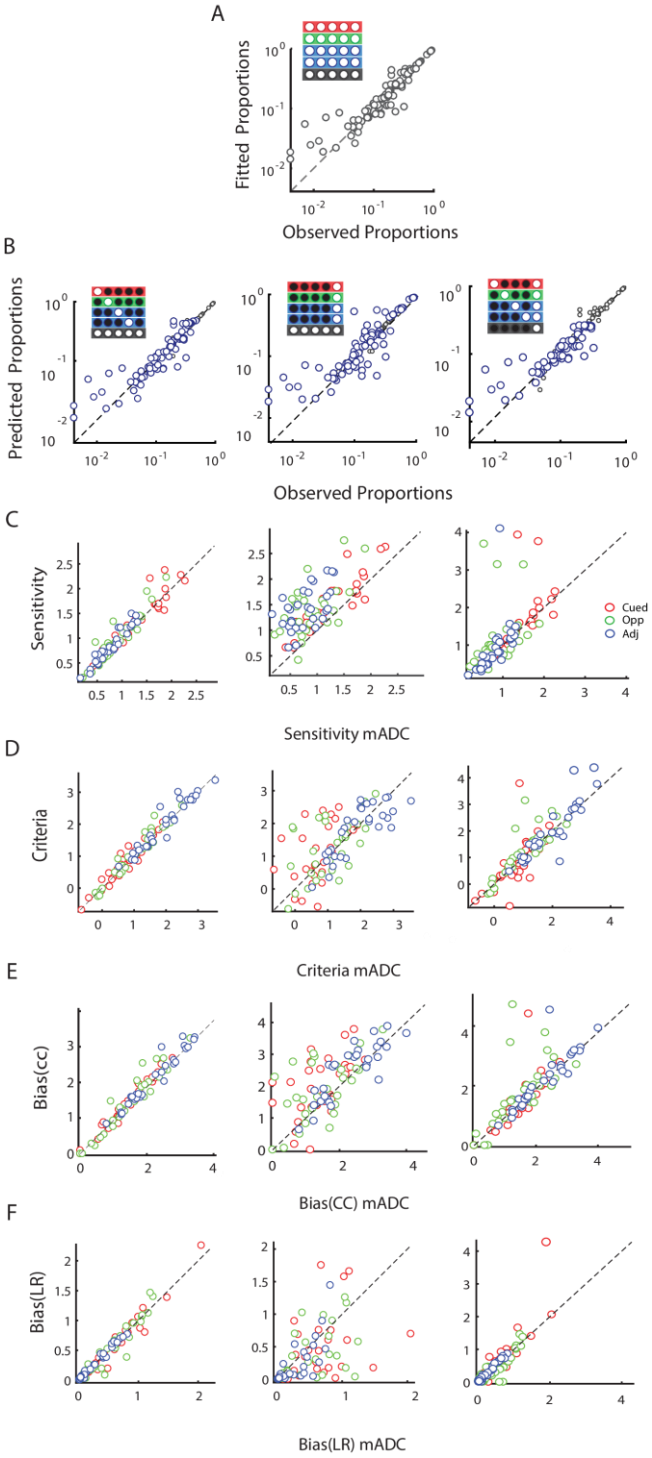


Figure S3: Comparison of parameter estimates for fitting versus prediction analyses. A. Model fit as in figure 1 F**. B.** (left to right) Model fit and predictions as in Fig 1G (main text), but with conditional probabilities. **C.** Comparison between sensitivity estimated by fitting m-ADC model using a subset (ordinate) of hits, false alarms and correct rejections model (left); subset of misses, false alarms and rejections (middle); and subset of hits, misses and correct rejections with sensitivity estimated by using all response contingencies (abscissa). Color conventions same as Figure S1.**D.** Same as A for Criteria**. E.** Same as A for Constant Criteria Bias**. F.** Same as A for Likelihood Ratio Bias.

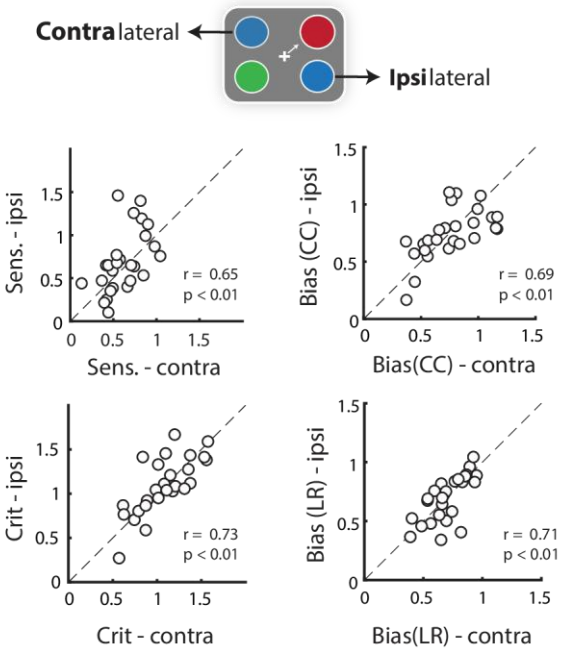


Figure S4: Comparison of parameters at cue-ipsilateral and cue-contralateral locations. Correlations between sensitivity (top left), criteria (bottom left), CC bias (top right) and LR bias (bottom right) values for ipsilateral and contralateral sides. Data points represent subjects, and diagonal shows line of equality.

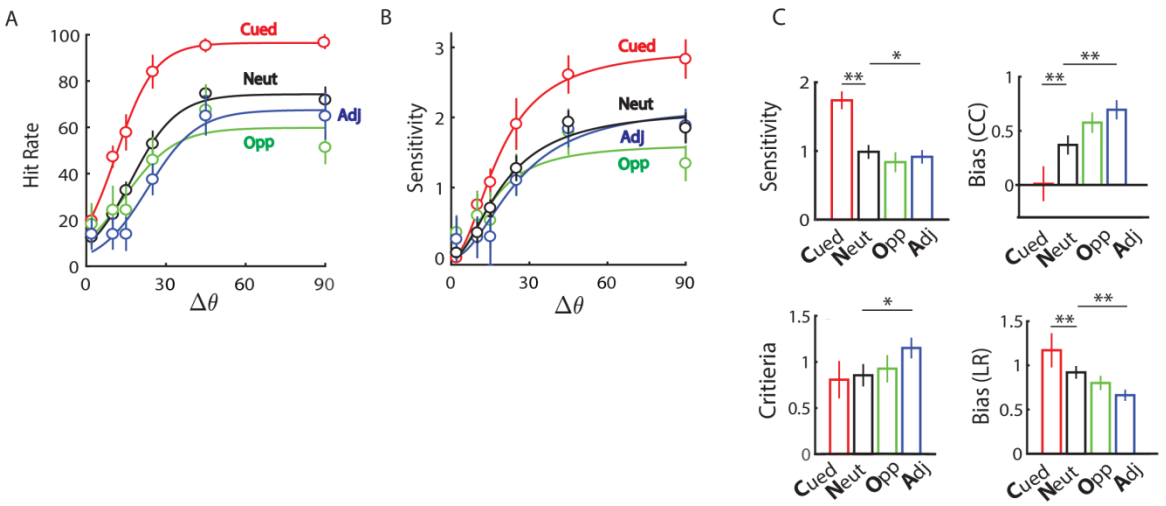


Figure S5: Sensitivity and bias effects in neutrally-cued and predictively-cued blocks. A. Average psychometric function (n=10 subjects, who performed both the neutrally cued and predictively cued blocks) showing hit rates at different change angles (Δθ) at the cued, opposite and adjacent locations and for neutrally cued trials across all locations (black). Circles, curves and error bars follow the same convention as Figure 2 A. **B.** Same as in panel A, but psychophysical functions showing sensitivity at different change angles. Circles, curves and error bars follow the same convention as Figure 2 B. **C.** Median sensitivity, criterion, constant criterion (bias-CC) and likelihood ratio (bias-LR) measures of bias at cued, opposite and adjacent locations and for neutrally cued trials across locations. Other conventions are same as Figure 2C (main text).

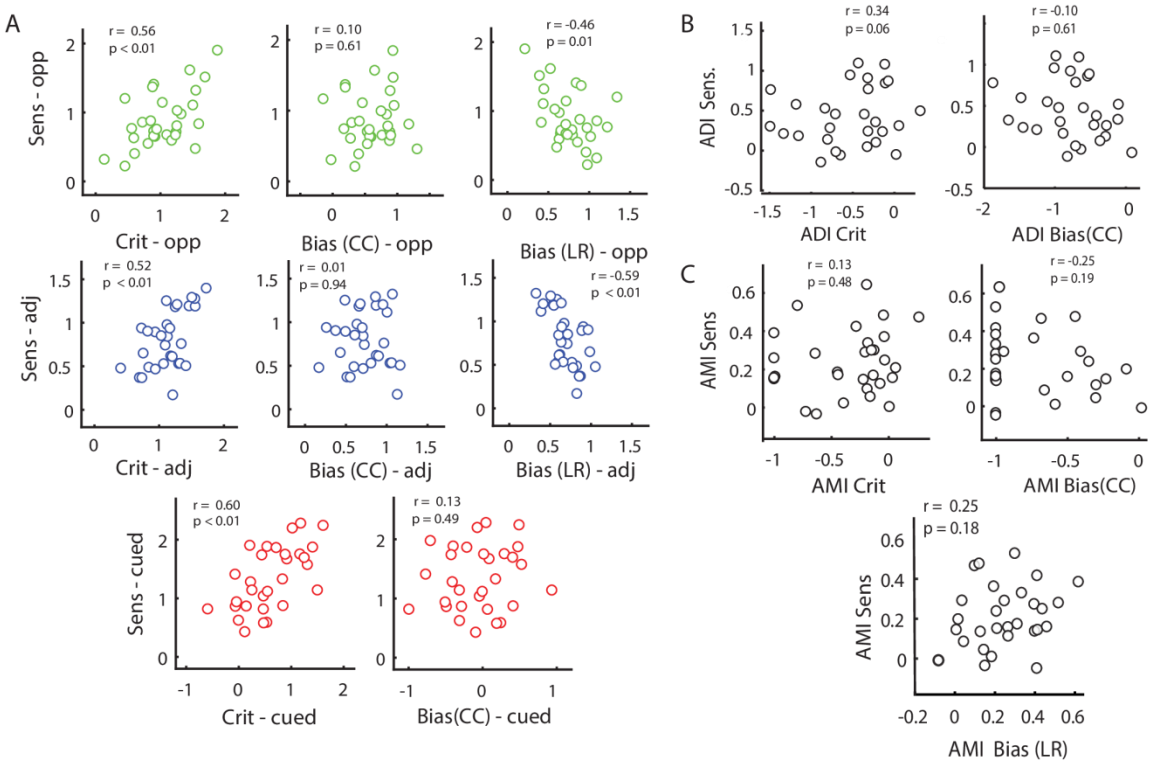


Figure S6: Covariation among psychophysical parameters: A. (Top row) Covariation of sensitivity with criterion, bias-CC and bias-LR at opposite location. (Middle row) Same as top row, but for adjacent location. (Bottom row) Covariation of sensitivity with criterion and bias-CC at cued location. **B.** Covariation of difference indices (ADI) of sensitivity with those of criteria, Constant Criteria bias. **C.** Covariation of modulation indices (AMI) of sensitivity with that of criteria, Constant Criteria bias and Likelihood ratio bias

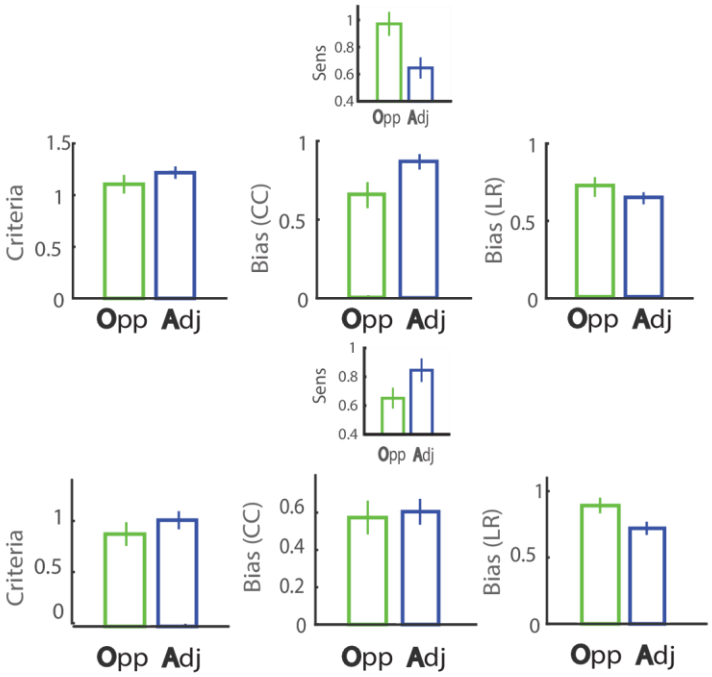


Figure S7: Comparison of bias values for subjects who showed different trends in sensitivity at cue-opposite and cue-adjacent locations. (Top row) Median criterion, CC bias and LR bias for subjects whose sensitivity is higher at the opposite location than at the adjacent location ($d'_{Opp} > d'_{Adj}$). (Inset) Sensitivity for this group of subjects. (Bottom row) Same as top row, but subjects with sensitivity lower at the opposite location than at the adjacent location ($d'_{Opp} < d'_{Adj}$).

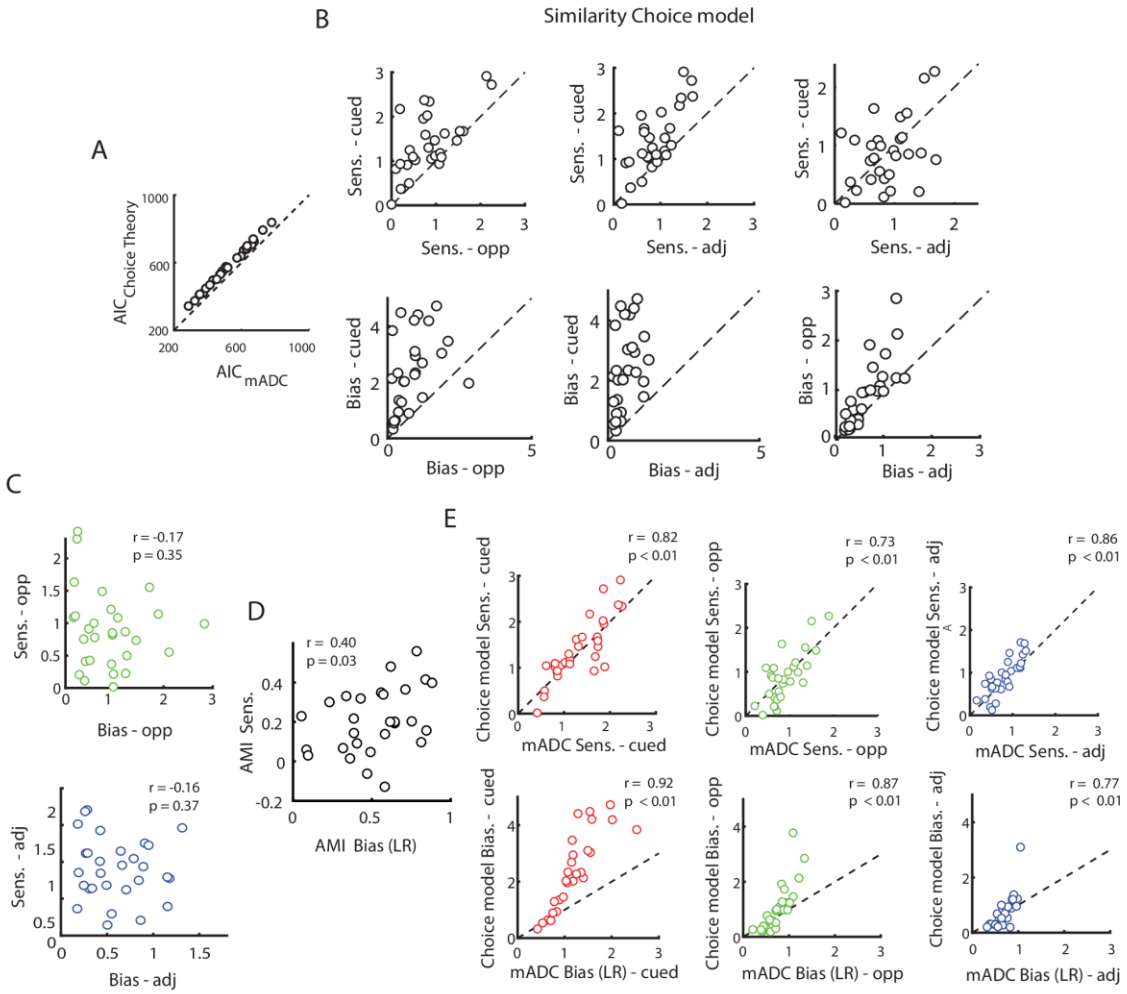


Figure S8: Similarity choice model estimates of sensitivity and bias. **A.** Comparison of AIC values obtained for m-ADC model vs. that from the choice model. **B**. Same as Figure 3 B, but for parameters estimated using similarity choice model. **C.** Same as Figure 3 G (left), but for opposite and adjacent locations. **D.** Covariation of modulation indices of sensitivity and bias estimated using choice model. Data points: individual subjects. **E** (Top) Covariation of choice model sensitivity with m-ADC sensitivity for cued (left), opposite (centre) and adjacent (right) locations. (Bottom) Same as left, but for Bias-LR. Data points: individual subjects and locations; color convention is same as Figure 1.

SI Appendix - Model derivation

mADC decision rule for a multialternative attention task employing the 'method of constant stimuli'

The conventional attention task shown in Fig.1A involves 'M+1' stimulus events: changes at one of 'M' locations or no change.

We denote the change event at each location as $S_1, ..., S_M$, and the no change event as S_0. Similarly, we denote the event (or response) locations with indices $1, 2, ..., M$ and denote the no change event (or response) with the index 0 (or ϕ). Finally, we denote p_k as the prior probability of change at location k and p_0 as the prior probability of no change.

Extending one dimensional signal detection models, we define a multivariate decision variable, $\mathbf{\Psi}$, whose components Ψ_j denote the scalar decision variable at each location j where $j \in \{1, 2, ..., M\}$.

We derive the optimal decision rule for the case where the change at each location k can occur at one of N_k different strengths (change magnitudes). The derivation is based on Bayesian decision theory for the optimal detection of signals in noise (Middleton and van Meter, 1955).

The observer makes a decision/response γ_k to one of M locations k or no response 0. This is denoted by the variable $\delta(\gamma_k|\mathbf{\Psi})$, such that

$$\begin{aligned} \delta(\gamma_k|\mathbf{\Psi}) &= 1 \qquad \text{if the observer chooses location k} \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{1}$$

$\sum_{k=0}^{M} \delta(\gamma_k|\mathbf{\Psi}) = 1$, since the observer can choose only one of the m locations (or 0 in case of no change).

In this case, the average Bayesian risk is given by:

$$R = \int \sum_{k=0}^{M} \Big[C_k^k \delta(\gamma_k|\mathbf{\Psi}) + \sum_{\substack{j=0 \\ j \neq k}}^{M} C_j^k \delta(\gamma_j|\mathbf{\Psi}) \Big] p_k w_k d\mathbf{\Psi} \tag{2}$$

where C_j^k is the cost of indicating a response to location j when the change occurred at location k, where $k, j = 0, 1, 2, ..., M$ and $w_k = p(\mathbf{\Psi}|S_k)$, i.e. the conditional probability density of the decision variable given that a change occurred at location k. Grouping terms by $\delta(\gamma_k|\mathbf{\Psi})$,

$$R = \int \sum_{j=0}^{M} \Big[\sum_{k=0}^{M} C_j^k p_k w_k \Big] \delta(\gamma_j|\mathbf{\Psi}) d\mathbf{\Psi} \tag{3}$$

Since only one of the $\delta(\gamma_j|\mathbf{\Psi}) \neq 0$, our objective is to find the j that minimizes the sum $\sum_{k=0}^{M} C_j^k p_k w_k$.

Let us consider the N_k magnitudes of changes associated with events S_k at each location k, and posit that each occurs with prior probability p_k^n and with conditional density w_k^n. Thus,

$$p_k w_k = \sum_{n=1}^{N_k} C_j^k p_k^n w_k^n \tag{4}$$

We assume also that costs are not different for the different magnitudes of change at each location. The sum to be minimized can then be written as:

$$\sum_{k=0}^{M}\sum_{n=1}^{N_k} C_j^k p_k^n w_k^n \tag{5}$$

Let us consider choosing between two alternatives i & j. The choice between i & j depends on the relative average risks, given by:

$$\begin{aligned} &C_j^0 p_0 w_0 + C_j^1 p_1 w_1 + ... + C_j^i p_i w_i + ... + C_j^j p_j w_j + ... + C_j^M p_M w_M \overset{i}{\underset{j}{\gtrless}} \\ &C_i^0 p_0 w_0 + C_i^1 p_1 w_1 + ... + C_i^i p_i w_i + ... + C_i^j p_j w_j + ... + C_i^M p_M w_M \end{aligned} \tag{6}$$

where the notation $\overset{i}{\underset{j}{\gtrless}}$ means choose i if LHS>RHS, and choose j otherwise.

We assume

$$C_j^l = C_i^l \quad \forall l \neq i, j \tag{7}$$

implying that the cost of making an incorrect response to any location (i or j) is the same for changes at a given location (l). Generally, we assume that for every event type S_k, the cost of making an incorrect response is the same, although this cost may be different across different event types.

With this simplification, equation (6) reduces to:

$$(C_j^j - C_i^j) p_j w_j \overset{i}{\underset{j}{\gtrless}} (C_i^i - C_j^i) p_i w_i$$

$$\frac{p_j w_j}{p_i w_i} \overset{i}{\underset{j}{\lessgtr}} \frac{(C_i^i - C_j^i)}{(C_j^j - C_i^j)}$$

where, the reversal of inequality happened because $(C_j^j - C_i^j)$ is defined to be non-zero and negative because the cost for making correct response is less than cost for making an incorrect response.

The left hand side of the previous equation is the generalized likelihood ratio, or the posterior odds ratio (product of the prior odds and the liklihood ratio). Introducing new notation for the cost ratio term on the right hand side,

$$\frac{p_j w_j}{p_i w_i} \overset{i}{\underset{j}{\lessgtr}} \beta_{ji}, \text{ where } \beta_{ji} = \frac{(C_i^i - C_j^i)}{(C_j^j - C_i^j)} \tag{8}$$

Substituting (4), this equation becomes

$$\frac{\sum_{n=1}^{N_j} p_j^n w_j^n}{\sum_{n=1}^{N_i} p_i^n w_i^n} \overset{i}{\underset{j}{\lessgtr}} \beta_{ji}, \text{ where } i, j \in \{0, 1, 2, ..., M\} \tag{9}$$

In our experiment, $N_i = N_j = N$ (6 change angles in our task, Fig. 1A) and $p_i^n = \frac{p_i}{N}$, $p_i^n = \frac{p_i}{N}$ $\forall i, j \in 0, 1, 2, ..., M$ (prior probabilities are evenly divided among the change angles). The only exception to this formula is for the no-change event for which $N_0 = 1$ (no change events do not occur at different change magnitudes).

We first examine the case for making a decision regarding change at location j vs no change.

$$\frac{\frac{p_j}{N}\sum_{n=1}^{N} w_j^n}{p_0 w_0} \underset{\phi}{\overset{j}{\gtrless}} \beta_{j0} \quad \text{or} \quad \frac{\sum_{n=1}^{N} w_j^n}{w_0} \underset{\phi}{\overset{j}{\gtrless}} \beta_{j0}\frac{p_0}{p_j}N \tag{10}$$

where ϕ denotes a no change response. We assume that w_j^ns are multivariate Gaussians with mean $[0, 0, ..., d_j^n, ..., 0]$ (an $m \times 1$ vector and identity covariance matrix. Again, as a special case $d_0^n = 1\ \forall n$; no change distribution has zero mean. Thus,

$$w_j^n \sim \mathcal{N}^M(d_j^n, I) = \prod_k e^{-(\Psi_k - \mu_k^n)^2}$$

Now, equation (10) above simplifies to

$$\frac{1}{N}\sum_{n=1}^{N} e^{\Psi_j d_j^n - \frac{(d_j^n)^2}{2}} \underset{\phi}{\overset{j}{\gtrless}} \beta_{j0}\frac{p_0}{p_j} \tag{11}$$

Therefore, optimal decision surfaces are hyperplanes of constant Ψ_j even if, in this case there is no evident closed form solution for Ψ_j. These hyperplanes correspond to the criteria, $\Psi_j = c_j$. Consequently, the generalized likelihood ratio measure of bias for choices to location j is defined as:

$$b_j = \frac{1}{\frac{1}{N}\sum_{n=1}^{N} e^{c_j d_j^n - \frac{(d_j^n)^2}{2}}} = \frac{p_j/p_0}{\beta_j} \tag{12}$$

which is the harmonic mean of the biases calculated from individual sensitivity values for each change angle (d_j^n).

We next examine the case for making a decision at location j vs location i.

$$\frac{\frac{p_j}{N}\sum_{n=1}^{N} w_j^n}{\frac{p_i}{N}\sum_{n=1}^{N} w_i^n} \underset{i}{\overset{j}{\gtrless}} \beta_{ji} \tag{13}$$

Dividing the numerator and denominator by $p_0 w_0$ & after some algebra,

$$\frac{\sum_{n=1}^{N} e^{\Psi_j d_j^n - \frac{(d_j^n)^2}{2}}}{\sum_{n=1}^{N} e^{\Psi_i d_i^n - \frac{(d_i^n)^2}{2}}} \underset{i}{\overset{j}{\gtrless}} \frac{\beta_{ji}}{(p_j/p_i)} \tag{14}$$

Thus, the optimal decision surface for choosing change at location i versus change at location j are not hyperplanes. However, planar decision surfaces provide excellent approximations to these optimal surfaces, as shown in Supplementary Fig. S2B, and as derived next.

Demonstration that optimal decision surfaces for the choice between locations i & j can be closely approximated by planar surfaces:

We apply the logarithm on both sides of equation (14):

$$log\left(\frac{\sum_{n=1}^{N} e^{\Psi_j d_j^n - \frac{(d_j^n)^2}{2}}}{\sum_{n=1}^{N} e^{\Psi_i d_i^n - \frac{(d_i^n)^2}{2}}}\right) \underset{i}{\overset{j}{\gtrless}} log\left(\frac{\beta_{ji}}{(p_j/p_i)}\right) \tag{15}$$

Since the numerator and denominator are identical in form, except for the location indices, we consider the numerator term alone ($log\big(\sum_{n=1}^{N} e^{\Psi_j d_j^n - \frac{(d_j^n)^2}{2}}\big)$), without loss of generality. We consider two cases, a) for small Ψ_j and b) for large Ψ_j.

a) First, consider small Ψ_j. $\lim_{\Psi_j \to 0} e^{\Psi_j d_j^n} = (1 + \Psi_j d_j^n)$, where, we have ignored terms $O(\Psi_j^2)$ and higher in the expansion for $e^{\Psi_j d_j^n}$. Thus, the numerator simplifies to:

$$\text{Nr.} \approx log\Big(\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}.(1 + \Psi_j d_j^n)\Big) \tag{16}$$

Further simplifying,

$$\text{Nr.} \approx log\Big(\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}} + \sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}} \Psi_j d_j^n\Big)$$

$$= log\left(\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}} \left(1 + \frac{\Psi_j \sum_{n=1}^{N} d_j^n e^{-\frac{(d_j^n)^2}{2}}}{\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}}\right)\right)$$

$$= log\left(\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}\right) + log\left(1 + \frac{\Psi_j \sum_{n=1}^{N} d_j^n e^{-\frac{(d_j^n)^2}{2}}}{\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}}\right)$$

$$\approx log\left(\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}\right) + \Psi_j \left(\frac{\sum_{n=1}^{N} d_j^n e^{-\frac{(d_j^n)^2}{2}}}{\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}}\right)$$

where, we have used the approximation $log(1 + x) \approx x$ for small x. This has the form,

$$\text{Nr.} = B_j + \Psi_j A_j \tag{17}$$

$$\text{where, } A_j = \frac{\sum_{n=1}^{N} d_j^n e^{-\frac{(d_j^n)^2}{2}}}{\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}} \quad B_j = log\left(\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}}\right)$$

An identical form holds for the denominator term in equation (15) for small Ψ_i.

b) Next, consider the numerator term from equation (15) for large Ψ_j

$$\text{Nr.} = log\big(\sum_{n=1}^{N} e^{-\frac{(d_j^n)^2}{2}} e^{\Psi_j d_j^n}\big)$$

$$= log\left(e^{\Psi_j d_j^N}\left[\left(\sum_{n=1}^{N-1} e^{-\frac{(d_j^n)^2}{2}} e^{\Psi_j (d_j^n - d_j^N)}\right) + e^{-\frac{(d_j^N)^2}{2}}\right]\right)$$

We consider d_j^ns to be sorted such that $d_j^N > d_j^{N-1} > ... > d_j^1$, which represent to the sensitivities corresponding to different change magnitudes (d_j^N, corresponding to the largest, and d_j^1 to the smallest change magnitude).

As $\lim_{\Psi_j \to \infty} e^{\Psi_j (d_j^n - d_j^N)} \to 0$ ($\because d_j^n - d_j^N < 0,\ \forall n < N$), this simplifies to

$$\text{Nr.} \approx log\big(e^{\Psi_j d_j^N} e^{-\frac{(d_j^N)^2}{2}}\big) \tag{18}$$

$$= \Psi_j d_j^N - \frac{(d_j^N)^2}{2}$$

$$= \Psi_j A_j^{'} + B_j^{'} \tag{19}$$

$$\text{where, } A_j^{'} = d_j^N \quad \& \quad B_j^{'} = -\frac{(d_j^N)^2}{2}$$

Again, an identical form holds for the denominator term in equation (15) for large Ψ_i.

Finally, we compute the decision surfaces for the choice between locations i and j for the different combinations of Ψ_i and Ψ_j magnitudes.
For small Ψ_i, large Ψ_j, the decision surface are given by (equations (17) and (19)):

$$\Psi_j A_j^{'} + B_j^{'} - \Psi_i A_i - B_i = log\eta_{ji} \tag{20}$$

Similarly, for large Ψ_j, large Ψ_i, the decision surfaces are given by (equation (19)):

$$\Psi_j A_j^{'} + B_j^{'} - \Psi_i A_i^{'} - B_i^{'} = log\eta_{ji}, \quad \text{where, } \eta_{ji} = \frac{\beta_{ji}}{(p_j/p_i)} \tag{21}$$

Decisions surfaces for the other two combinations (small Ψ_j, large Ψ_i and small Ψ_j, small Ψ_i) may be similarly derived. In each case, the decision rule is approximated by a planar decision surface (decision hyperplane).