

# Neural Representations of Relevant and Irrelevant Features in Perceptual Decision Making

Andrew S. Kayser,<sup>1</sup> Drew T. Erickson,<sup>1</sup> Bradley R. Buchsbaum,<sup>2</sup> and Mark D'Esposito<sup>3</sup>

<sup>1</sup>Ernest Gallo Clinic and Research Center, Department of Neurology, University of California, San Francisco, Emeryville, California 94608, <sup>2</sup>Rotman Research Institute, Baycrest Research Centre, Toronto, Ontario M6A 2E1, Canada, and <sup>3</sup>Helen Wills Neuroscience Institute, Department of Psychology and The Henry H. Wheeler, Jr., Brain Imaging Center, University of California, Berkeley, Berkeley, California 94720

Although perceptual decision making activates a network of brain areas involved in sensory, integrative, and motor functions, circuit activity can clearly be modulated by factors beyond the stimulus. Of particular interest is to understand how the network is modulated by top-down factors such as attention. Here, we demonstrate in a motion coherence task that selective attention produces marked changes in the blood oxygen level-dependent (BOLD) response in a subset of regions within a human perceptual decision-making circuit. Specifically, when motion is attended, the BOLD response decreases with increasing motion coherence in many regions, including the motion-sensitive area MT+, the intraparietal sulcus, and the inferior frontal sulcus. However, when motion is ignored, the negative parametric response in a subset of this circuit becomes positive. Through both modeling and connectivity analyses, we demonstrate that this inversion both reflects a top-down influence and segregates attentional from accumulation regions, thereby permitting us to further delineate the contributions of different regions to the perceptual decision.

## Introduction

Perceptual decision making is a fundamental cognitive ability important for everyday function. During an activity as commonplace as driving, we critically depend on this capacity to rapidly apply the brakes based on a red traffic signal, or to veer left or right based on the movement of a nearby car. Both our own (Kayser et al., 2010) and other work (for review, see Heekeren et al., 2008) implicate a network of brain regions important for such tasks, including sensory areas in posterior cortex, integrative regions in parietal cortex, and response-related regions within the frontal lobe.

Importantly, past work has found that subjects' accuracies and response times in a dot motion coherence task accord well with a decision-making model in which evidence for motion direction accumulates until a threshold is reached and a decision is made (Palmer et al., 2005; Ratcliff and McKoon, 2008). Because this model predicts that stronger stimuli lead to faster evidence accumulation, the summed neural activity reflected in the blood oxygen level-dependent (BOLD) response is less for higher motion coherence than for lower motion coherence, generating a negative parametric prediction (i.e., that higher motion coherences lead to lower BOLD responses) for accumulator regions and downstream areas (Kayser et al., 2010). However, an initially puzzling finding in our previous study was the identification of a

negative parametric effect in MT+, a region not thought to demonstrate an effect of accumulation (Gold and Shadlen, 2007). Although explanations other than accumulation, including both bottom-up and top-down (e.g., attentional) mechanisms, could be responsible, these results illustrate the more general need to identify additional factors influencing the circuit. Attention, for example, can increase the gain (Maunsell and Treue, 2006) and shift the tuning (David et al., 2008) of neural responses, possibly as a result of frontoparietal activity (Beck and Kastner, 2009; Silver and Kastner, 2009). If these association regions not only direct attention toward relevant features, but also participate in ignoring irrelevant features, we might dissociate these processes by comparing activity for attended and ignored stimuli during the decision. Specifically, attentional regions would likely show parametric activity in both attend/ignore conditions, reflecting the enhancement/suppression of relevant/irrelevant features, whereas accumulator regions would likely vary only with the (attended) feature necessary to the decision. We thus address two primary questions concerning the decision-making process. First, what effect do attentional manipulations have on sensory representations during perceptual decision making? Second, how is the ignored stimulus represented within decision areas of the perceptual decision-making network?

In this study, we cued subjects to attend to either the overall direction of motion or the predominant color of a dot stimulus while ignoring the other feature. Based on a simple model of motion processing in MT+, we predicted that the negative parametric variation with motion coherence in the BOLD signal would persist in the ignored condition if this effect were bottom-up. However, if the bottom-up population response in MT+ truly increased with increasing motion coherence, the parametric effect should invert (i.e., become positive) when attention was

Received June 18, 2010; revised Aug. 27, 2010; accepted Sept. 20, 2010.

This work was supported by start-up funds provided by the State of California (A.S.K.) and National Institutes of Health Grants NS-40813 and MH-63901 (M.D.).

Correspondence should be addressed to Andrew S. Kayser, Ernest Gallo Clinic and Research Center, Department of Neurology, University of California, San Francisco, 5858 Horton Street, Suite 200, Emeryville, CA 94608. E-mail: akayser@gallo.ucsf.edu.

DOI:10.1523/JNEUROSCI.3163-10.2010

Copyright © 2010 the authors 0270-6474/10/3015778-12\$15.00/0

withdrawn in the ignored condition. Additionally, we theorized that our attentional manipulation should dissociate upstream regions important for attention from those important for evidence accumulation, in that an attentional area, unlike an accumulation area, should demonstrate activity responsive to the ignored feature.

## Materials and Methods

**Subject training and task performance.** Five subjects (ages, 26–38; two males) participated in the study and gave written informed consent in accordance with the Committee for the Protection of Human Subjects at the University of California, Berkeley. All subjects had normal neuroanatomy as reviewed by a neurologist (A.S.K.), were right-handed, and had normal or corrected-to-normal vision. Before scan sessions, subjects were trained on the task for a minimum of nine 1.5 h sessions to reduce both the number of invalid trials and learning effects in the scanner. The last two training sessions were performed in the magnetic resonance imaging (MRI) scanner, both to acclimate subjects to the scanner environment and to provide an independent set of data for generation of regions of interest (ROIs). Once fully trained, subjects underwent six 1.5 h functional MRI (fMRI) task sessions, consisting of eight runs of 48 trials for a total of  $6 \times 8 \times 48 = 2304$  trials. Because of technical problems with scan acquisition, for two subjects two runs were discarded, leaving them with a total of 2208 trials each. Scanning a small number of highly trained subjects maximized our ability to detect parametric changes in the BOLD signal within all conditions, as in other visual studies (Lee et al., 2007; Silver et al., 2008; Amano et al., 2009). Additionally, the large number of trials per subject allowed us to obtain good fits of the diffusion model and robust single-subject activation maps (see below).

Subjects performed a visual dot motion or color proportion task on a stimulus consisting of multiple colored moving dots in which one of these two features (motion or color) was relevant to the task during a given run (see Fig. 1). For all trials, regardless of the attended feature, a subset of dots moved coherently (leftward or rightward) on a background of randomly moving dots, and an uncorrelated subset of dots of one color (blue or red) was present on a background of evenly apportioned blue and red dots. For each attend-motion trial, subjects were required to identify the direction of motion as quickly and accurately as possible. For each attend-color trial, subjects were required to identify the predominant color as quickly and accurately as possible. Importantly, the fully crossed factorial design ensured that the attended and ignored features varied parametrically in independent fashion. Thus, coherence values were balanced within sessions such that subjects viewed equal numbers of all combinations of attended condition, motion coherence, color coherence, leftward/rightward motion, and red/blue color proportion, in randomized and independent fashion. Each run began with a colored text prompt directing the subject to perform either the motion or color task for all trials within the run. The color of the text was either green or orange, counterbalanced across subjects (although consistent within subject). Additionally, the fixation cross was rendered in the same color as the text prompt to reinforce the relevant task.

A trial began with dimming of the fixation cross and appearance of the dot motion stimulus for 2500 ms. Color and motion coherence remained consistent throughout the trial. To indicate their choice, subjects made a button press with either their second or third fingers before the end of the stimulus interval. For motion, the second finger corresponded to leftward motion and the third finger to rightward motion. For color, the correspondence between finger and red/blue color was counterbalanced across subjects (although consistent within subject). After 2500 ms, the stimulus disappeared, the fixation cross brightened to its original contrast, and an interstimulus interval ranging from 4000 to 12,000 ms preceded the next trial. The stimulus persisted for the entire 2500 ms interval, regardless of the subject's response time, to avoid confounding response time and stimulus duration.

Subjects initially undertook training sessions in which coherence values taken from our previous study (0, 2, 4, 8, 16, 32, and 64%) (Kayser et al., 2010) were used for both motion and color features. To ensure that each task was well learned, we initially trained each subject on the color

and motion tasks in separate sessions. Three of the five subjects first performed the motion task in the absence of the other stimulus for three 1.5 h sessions. They were then trained on the color task in the absence of the motion stimulus for three additional 1.5 h sessions. The other two subjects were trained in the opposite order (i.e., color, then motion). Finally, all subjects were trained on the task with both color and motion features present for an additional three 1.5 h sessions. All subjects reached stable performance by the end of training, as determined by a 5% or less session-to-session change in the halfway accuracy threshold [corresponding to the coherence level at 75% accuracy as defined by the fitted diffusion model parameters (Palmer et al., 2005)].

As noted above, behavioral data from the training sessions were fit with a proportional rate diffusion model (Palmer et al., 2005), as per our previous work, both to further validate subject performance and to determine accuracy across the range of coherence values for motion and color, respectively. The diffusion model hypothesizes that decision making consists of a process of evidence accumulation for each of the alternative decisions available to a subject. When a threshold level of evidence is accumulated for one of the decisions, the subject generates a corresponding response. Importantly, the model permits one to fit both reaction time (RT) and accuracy data with a single set of parameters, thereby simultaneously constraining both RT and accuracy variables and providing a parsimonious and theoretically meaningful explanation for the data. The Palmer model, derived from the diffusion model of Ratcliff (Ratcliff and McKoon, 2008), consists of three variables: (1)  $A'$ , bearing on the decision threshold; (2)  $k$ , a constant describing the relationship between the motion coherence in the stimulus and the mean drift rate in the model; and (3)  $T_R$ , the mean residual time in seconds, representing a fixed processing duration independent of evidence accumulation (e.g., for low-level sensory processing or implementation of motor commands). Both sets of parameters derived for each subject (i.e., for the attend-motion and attend-color conditions) were determined by an iterative procedure designed to optimize the log-likelihood ( $L_p$ ) of the diffusion model fit (Palmer et al., 2005). Using these fits, we determined coherence values predicted to lead to 60, 70, 80, 90, and 100% accuracy for each subject in the attend-motion and attend-color tasks. Across subjects, the geometric means for the corresponding color coherence values were 3.5, 7.3, 11.9, 19.0, and 41.4%, respectively; for motion coherence, the values were 1.3, 3.1, 4.9, 7.8, and 22.5%. The individually calibrated coherence values were used in the task performed by subjects during the fMRI scanning session, along with a 0% coherence control.

Part of the subject training process also consisted of eye movement training. As in our previous study (Kayser et al., 2010), subjects were trained through verbal feedback to maintain an eye position within  $3^\circ$  of the fixation cross, and to refrain from blinking throughout the duration of the stimulus regardless of the response time. These constraints were designed to reduce the potential effect of eye movements on BOLD responses. To this end, eye movement data for three subjects was acquired at the Neuroimaging Center at the University of California, San Francisco, Medical Center using an ASL Eye-Trac 6 LRO (<http://www.a-s-l.com>). Eye movement data for the other two subjects was collected during training sessions at University of California, Berkeley, using a ViewPoint Eye Tracker (Close-Focus Camera and Illuminator; <http://www.arringtonresearch.com>). Relatively stringent criteria were used to train subjects to maintain fixation and avoid blinks during the stimulus interval. Blinks were classified as any instances in which the pupil aspect ratio was equal to zero for  $>8.3$  ms. Eye movements were defined as any period lasting  $>180$  ms in which the eye position was  $>3^\circ$  from fixation. Three of the five subjects had been previously trained [two for our previously published study (Kayser et al., 2010) using the identical motion coherence stimulus] and were not retested here. For the other two subjects, performance was well within acceptable ranges ( $<1.5\%$  of trials compromised by blinks or eye movements).

The task was programmed in Matlab using components of PsychToolbox (Brainard, 1997; Pelli, 1997) adapted from our previous code. Stimulus frames were presented within a central  $7.5^\circ$  aperture at 60 frames/s. Dot density was fixed at  $16.7 \text{ dots} \cdot \text{deg}^{-2} \cdot \text{s}^{-1}$ , and dot velocity was fixed at a single value of  $5 \text{ deg/s}$  to ensure that motion energy was uniform across levels of motion coherence. Blurring effects (in which consecutive

placements of a single dot were seen as forming a line) were avoided via the serial presentation of three interleaved subsets, with each frame containing only one of the subsets. To ensure that dots were initially placed evenly across the viewing aperture, we rejected initial dot placements that showed evidence for an unusually skewed starting configuration. Specifically, we rejected initial random dot configurations that showed a 95% or greater chance of deviating from the expected  $\chi^2$  distribution for the frequency of dots over a  $4 \times 4$  grid covering the viewing aperture (note that the grid was not displayed on the screen). Once set in motion, dots that moved outside the aperture were repositioned on the opposite side of the window to prevent them from collecting in any particular region of the aperture over time. For the color task, equiluminant hues were derived from CIE  $xyY$  space in which  $Y = 25$  and saturation was maximal. These values were then converted into RGB space, and specific red and blue values were selected that maximized the color of interest component (red or blue) while minimizing the other two components. The two selected values were as follows: red = [255 65 2] and blue = [5 137 255]. To confirm that there was no difference in the perceptual salience of these colors, they were matched behaviorally during training such that responding at the 0% coherence level gave rise to 50% red and 50% blue responses. We also examined postcalibration data for possible bias for selecting one of the two colors. Although one subject developed a response bias that reached significance in both the attend-motion (left response favored: 64% of trials) and attend-color (blue response favored: 66% of trials) conditions, the subject's performance remained well matched for accuracy across conditions (see Fig. 2). Likewise, subtraction of the subject's parametric BOLD data classified by response (blue–red, left–right) revealed no significant differences for any studied variable (voxelwise  $\beta$  values for the parametric blue–red and left–right contrasts: all uncorrected values of  $p > 0.1$ ; all ROI-derived peak amplitudes: values of  $p > 0.6$  for the effect of response key in all attended conditions by repeated-measures ANOVA).

Actual coherences for a single display frame were determined by sampling from a uniform distribution independently for both motion and color. Values from each of these distributions were chosen by thresholding the relevant distribution by the selected coherence proportion to produce an integral number of dots assigned the coherent feature. All other dots were assigned directional and color features randomly. Specifically, motion directions were sampled uniformly from 0 to  $360^\circ$ , whereas color was randomly designated as either blue or red. Thus, for 50% color coherence, for example, 50% of the dots were assigned to the selected color (e.g., red), whereas the remaining 50% of the dots were evenly divided between red and blue. Moreover, the subset of dots representing the coherent feature (motion and/or color) changed from frame to frame so that the subset of coherent dots on one frame was not the same as the subset of coherent dots on the previous frame. Likewise, for each frame the particular group of dots representing the coherent subset for one feature (e.g., color) was selected independently from the subset of dots representing the other feature (e.g., motion). Consequently, the coherency for each feature was distributed across the full dot set, independently of the other feature, preventing subjects from making accurate decisions based solely on the behavior of a single dot or set of dots. We previously demonstrated that the mean coherence across all frames for a given trial well approximates the desired coherence (Kayser et al., 2010).

**Localizer tasks.** Subjects also underwent separate motion and color localizer tasks. In the motion localizer task, subjects viewed 10 repetitions of a 40 s trial in which the motion stimulus was present for 10 s, followed by 30 s of a static dot display. When the motion stimulus was present, it consisted of 10 consecutive 1 s presentations of 100% dot motion coherence in which the directions were chosen randomly, without replacement, from the set of  $[0, 36, 72, \dots, 324^\circ]$ . The color localizer task also consisted of 10 repetitions of a 40 s trial. However, the 10 s color stimulus consisted of 10 consecutive 1 s stationary displays in which colored squares subtending  $0.57^\circ$  were shown at a density of 16.7 squares/deg $^2$ . Twenty-four fully saturated colors were selected from the  $xyY$  space, including the blue and red hues used in the main experiment. In the 30 s interval between color displays, the same type of stimulus was shown in grayscale.

**MRI scanning.** MRI scanning was conducted on a Siemens MAGNETOM Trio 3T MR Scanner at the Henry H. Wheeler, Jr., Brain Imaging Center at the University of California, Berkeley. Anatomical images consisted of 160 slices acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo protocol [repetition time (TR), 2300 ms; echo time (TE), 2.98 ms; field of view (FOV), 256 mm; matrix size,  $256 \times 256$ ; voxel size, 1 mm $^3$ ]. Functional images consisted of 24 slices acquired with a gradient echoplanar imaging protocol (TR, 1370 ms; TE, 27 ms; FOV, 225 mm; matrix size,  $96 \times 96$ ; voxel size,  $2.3 \times 2.3 \times 3.5$  mm). A projector (Avotec SV-6011; <http://www.avotec.org>) was used to display the image on a translucent screen placed within the scanner bore behind the head coil. A mirror was used to allow the subject to see the display. The distance from the subject's eye to the screen was 28 cm. Subjects made their responses via a MRI-safe fiber optic response pad (inline model HH-1x4-L; <http://www.crslltd.com>).

**fMRI preprocessing.** fMRI preprocessing was performed using both AFNI (<http://afni.nimh.nih.gov>) and FSL (<http://www.fmrib.ox.ac.uk/fsl/>). Functional images were converted to 4D Nifti format and corrected for slice-timing offsets. Motion correction was performed using the AFNI program 3dvolreg, with the reference volume set to the mean image of the first run in the series. Images were then smoothed with a 5 mm full-width at half-maximum Gaussian kernel. Coregistration was performed with the AFNI program 3dAllineate using the local Pearson correlation cost function optimized for fMRI-to-MRI structural alignment. The subsequent inverse transformation was then used to warp the anatomical image to the functional image space. Anatomical images were normalized using the FSL program fnirt to a standard volume (MNI\_N27) available from the Montreal Neurological Institute (MNI) (<http://www.bic.mni.mcgill.ca>). The same normalization parameters were later applied to native-space statistical maps as necessary for the generation of group statistical maps (see below).

**Univariate analysis.** To address a series of hypotheses, we performed a number of voxelwise fMRI statistical analyses for each subject using the general linear model framework implemented in the AFNI program 3dDeconvolve. The overall effects of both motion and color coherence were assessed by modeling the six levels of motion coherence, the six levels of color coherence, and the attended/ignored feature (motion/color) with separate regressors, each of which was derived by convolving a  $\gamma$  probability density function (peaking at 6 s) with a vector of stimulus onsets for each condition. Tests of linear trends were performed for each voxel using the appropriate contrast vectors [the relevant coherence vector transformed to zero mean and a sum of squares equal to 1 (Kayser et al., 2010)] applied to the estimated  $\beta$  coefficients computed for each motion coherence level, each color coherence level, and each attended condition. Because each trial was associated with both an attended feature coherence and an ignored feature coherence, the responses across trials could be parameterized by the coherence of either feature. Thus, the “attend-motion” and “ignore-color” analyses represented the same trials parameterized by different values. Importantly, since the attended and ignored feature coherences varied independently/orthogonally, the parametric effect across all trials would differ, depending on whether the attended or ignored coherence value was analyzed. The resulting values were subject to group level analyses, then mapped to the spatially normalized cortical surface.

Because we collected a large amount of data on a relatively small number of subjects, statistical power was relatively weak at the group level compared with the single-subject level. Thus, for the purposes of a group activation summary, we assessed significance using a fixed effects summary statistic with an overlap requirement (Friston et al., 1999). We computed a  $t$  statistic for the linear contrast for every voxel in the volume and divided this value by the square root of the number of subjects ( $n = 5$ ) (McNamee and Lazar, 2004), which was compared against a standard normal null distribution using an  $\alpha$  value of  $p = 0.001$  for the full group. We also required that, for a voxel to be declared significant, at least three of five subjects show a significant effect ( $p < 0.05$ ) at the single-subject level. To further ensure that this group summary map did not obscure large intersubject variability, we also evaluated parametric responses on a single-subject level (supplemental Fig. 1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). In contrast to the whole-brain group sum-



mary analyses, all statistics performed on ROI-extracted data were submitted to random-effects *t* test or repeated-measures ANOVA (see following text).

**ROI selection.** To avoid a ROI selection bias, fMRI data derived from the two training sessions performed in the MRI scanner, as well as data from the color and motion localizer tasks, were separately analyzed to generate regions of interest. ROIs were selected from those regions that showed both negative parametric effects across the attend-motion and attend-color conditions, as well as a positive main effect of task. Specifically, after single-subject maps were normalized to MNI space, local maxima were defined on the fixed-effects group map for the parametric effect of both attend-motion and attend-color, thresholded at  $p < 0.01$ , uncorrected. Each defined maximum served as the center of a sphere with a two-voxel radius ( $11.5 \times 11.5 \times 17.5$  mm). In cases in which neighboring spheres overlapped, the sphere with the lesser maximum was excluded. After reverse normalizing the ROIs to each subject's native space, we selected the 10 voxels from all voxels within each ROI that (1) demonstrated a positive main effect of task and (2) showed the maximal negative parametric variation for a regressor parameterized by both attended features (i.e., across both the attend-motion and attend-color conditions). This criterion was used to select voxels that were both activated by the task and responsive to changes in coherence for both features. There were two exceptions to this approach: a motion-sensitive region consistent with the location of MT+ was defined using the motion localizer task, whereas a color-sensitive region consistent with the location of V4 was defined using the color localizer task. In these two cases, we computed fixed-effects group maps for the localizer contrasts (motion–stationary or color–grayscale), thresholded at  $p < 0.001$ , uncorrected. After defining these common activations for both MT+ and V4, we reversed-normalized the two ROIs to each subject's native space and selected the 10 voxels in each subject that showed the strongest difference for each localizer condition. As described above, each of these sets of ROIs—whether derived from the training period or the localizers—was then applied to the primary (and independent) data set produced for the matched accuracy values.

These selection criteria were useful for a number of reasons. First, each ROI was constrained by the positive main effect of task to ensure that we were not reporting areas that deactivated during task performance (for other discussions of this issue, see Tosoni et al., 2008; Ho et al., 2009). Second, as noted above, voxel selection was based not on a significant parametric response to either motion or color alone, but to the entirety of the attended features. This choice ensured that we were not artificially favoring regions that responded only to color or to motion or were constrained in any way by the response to the ignored feature. Third, the responses of different voxels to a stimulus (e.g., to left/right motion) can vary by voxel. Although each of our voxels was much larger than a cortical microcolumn, by chance there are likely, for example, to be voxels that contain more left- and right-preferring microcolumns than others. That these voxels respond better to these two particular directions may also have physiological meaning, providing a physiological justification for focusing on these voxels (as opposed to ones hypothetically demonstrating a stronger response to the nonpresented orthogonal directions). Finally, and most importantly, these voxels were selected from an independent data set of 16 runs for our subjects. Thus, our selection criteria did not influence, and were not influenced by, the data ultimately analyzed.

**BOLD time course estimation.** Estimates of the hemodynamic response were calculated for each combination of feature and coherence within an ROI. To produce an unbiased estimate of the time course, we applied a deconvolution approach to the main data set using piecewise b-spline basis functions (Saad et al., 2006) separated by 2 s intervals for 20 s after onset using 3dDeconvolve of AFNI. Since onset times were not synchronous with the transistor–transistor logic (TTL) pulse, across the entire run we were able to sample the time course at a number of different points. To select and label the relevant time courses at each voxel, ROIs were reverse normalized to each subject's native space. The peak amplitude was defined as the first maximum in the average time course after stimulus onset, and time to peak was considered the time from onset to this maximum amplitude.

**Modeling.** To examine the role of top-down effects in MT+, we developed a model to determine the influence of attention in the attended and ignored conditions, respectively. For each subject, input to the model included the mean reaction times for all conditions [six attend-motion reaction times ( $RT^{am}$ ) and six ignore-motion reaction times ( $RT^{im}$ )]; the peak amplitudes of the MT+ time courses for those same 12 conditions ( $p^{am}$  and  $p^{im}$ ) as derived from the deconvolved hemodynamic responses; and the duration of the stimulus ( $d$ ). Thus, there were 25 total inputs to the model. For the analysis, attention was conceptualized as a multiplicative factor acting on the bottom-up input (Treue and Maunsell, 1996, 1999; Maunsell and Treue, 2006). This attentional factor was assumed to be different from baseline for the length of the response time, and to return to baseline thereafter. The bottom-up input, on the other hand, was modeled as persisting for the entire 2500 ms of the stimulus presentation. In addition to defining a value for the attentional factor in both the attend-motion ( $k^{am}$ ) and ignore-motion ( $k^{im}$ ) conditions, we established parameters representing the value of the bottom-up input for each of the six motion coherences ( $c_{1-6}$ ). These bottom-up inputs were assumed to be consistent regardless of the focus of feature-based attention. Importantly, no a priori relationship between the multiplicative factors, or between the bottom-up motion coherences, was assumed, parametric or otherwise. As a result, the 25 input values were characterized by eight parameters via the following equation, where the null values in the left matrix indicate that it is block-diagonal:

$$\begin{bmatrix} RT_1^{am}k^{am} + (d-RT_1^{am}) & & & & & \\ & \ddots & & & & \\ & & \emptyset & & & \\ & & & \ddots & & \\ RT_1^{im}k^{im} + (d-RT_1^{im}) & & & & & \\ & & & & & \\ & & & & & \emptyset \\ & & & & & & \ddots \\ & & & & & & & RT_6^{im}k^{im} + (d-RT_6^{im}) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{bmatrix} = \begin{bmatrix} p_1^{am} \\ p_2^{am} \\ p_3^{am} \\ p_4^{am} \\ p_5^{am} \\ p_6^{am} \\ p_1^{im} \\ p_2^{im} \\ p_3^{im} \\ p_4^{im} \\ p_5^{im} \\ p_6^{im} \end{bmatrix}$$

In essence, the model posits that the MT+ BOLD response represents a combination of bottom-up input and RT-dependent attentional modulation. The best-fitting model for each subject was calculated via maximum likelihood, and the resulting values for the eight parameters were assessed for statistical significance via random-effects *t* tests computed across subjects.

**Granger analysis.** Granger causality (GC) is a signal processing technique in which multivariate autoregressive (MVAR) models of a time series are used to predict upcoming time points (Roebroeck et al., 2005; Kayser et al., 2009). If the MVAR model of a time series of interest more reliably predicts upcoming time points when a second time series is incorporated, the second time series is said to be Granger causal for the first. Additionally, conditional Granger causality analyses can be performed. In this case, one hypothesizes that the influence of one region on another is actually mediated by a third area. By incorporating this third area into the MVAR model, the influence between two regions can be computed, conditional on the third.

To compute GC values, we used the same native-space ROIs defined for our univariate analyses. Realigned and smoothed images for each subject were then used to extract fMRI time series for each voxel. After detrending each time course, we averaged across all voxels in the ROI to produce mean time courses. GC values were subsequently computed between V4, MT+, middle intraparietal sulcus (mIPS), and inferior frontal sulcus (IFS) for each run, divided into smaller sets of 35 TRs in the same fashion as in our previous work (Kayser et al., 2009) to balance reliability of individual GC values with the ability to obtain multiple values for each time series. These values were initially segregated by attentional condition. Because there were no significant differences between GC values in the attend-motion and attend-color conditions, we combined these data across conditions. Statistical significance was determined within subject for each set of GC values via Wilcoxon's two-sided signed rank test ( $p < 0.05$ ). Only those connections demonstrating a

conjoint probability  $< 3.125 \times 10^{-7}$  across subjects (equal to  $0.05^5$ ) and significant in four of five subjects were considered significant at the group level, with the direction of the arrow determined by the mean Granger value across subjects. To determine the relative influence of mIPS and IFS, we also computed conditional Granger causality values; these values were evaluated in the same fashion as described above.

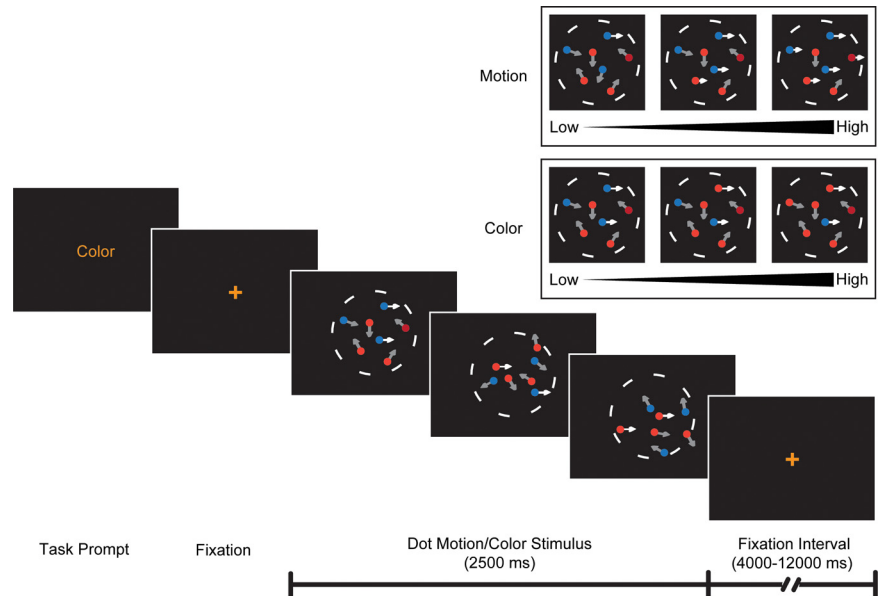
## Results

To evaluate the representation of irrelevant features during perceptual decisions, and to compare the effect of attention between attended and ignored conditions, we acquired fMRI data from five subjects making perceptual decisions about a stimulus containing both motion and color features (Fig. 1). As described in Materials and Methods, highly trained subjects were cued at the start of each run to perform a two-alternative forced-choice task in which the attended feature was either leftward/rightward dot motion (attend motion) or red/blue color proportion (attend color).

### Behavioral performance

Subjects were initially trained on stimuli in which the coherence of each feature (either motion or color) was set to predefined values (see Materials and Methods). Accuracies and response times across these coherence levels were then used to define for each subject and task the best-fitting diffusion model (Fig. 2A). This model has previously been shown to faithfully describe accuracy and reaction time data simultaneously with only three parameters: the sensitivity, which relates the stimulus coherence level to the rate of evidence accumulation; the threshold, which defines the amount of evidence required for a decision to be made; and the residual time, which accounts for nondecision processes inherent in peripheral sensory processing and implementation of motor commands (Palmer et al., 2005; Ratcliff and McKoon, 2008). To ensure that changes in accuracy could not explain behavioral or neural differences between features, we used these parameters to identify interpolated color and motion coherence levels for each subject that predicted 50, 60, 70, 80, 90, and 100% accuracy during the attended condition (see Materials and Methods). These calibrated coherence levels were used throughout the subsequent experiment.

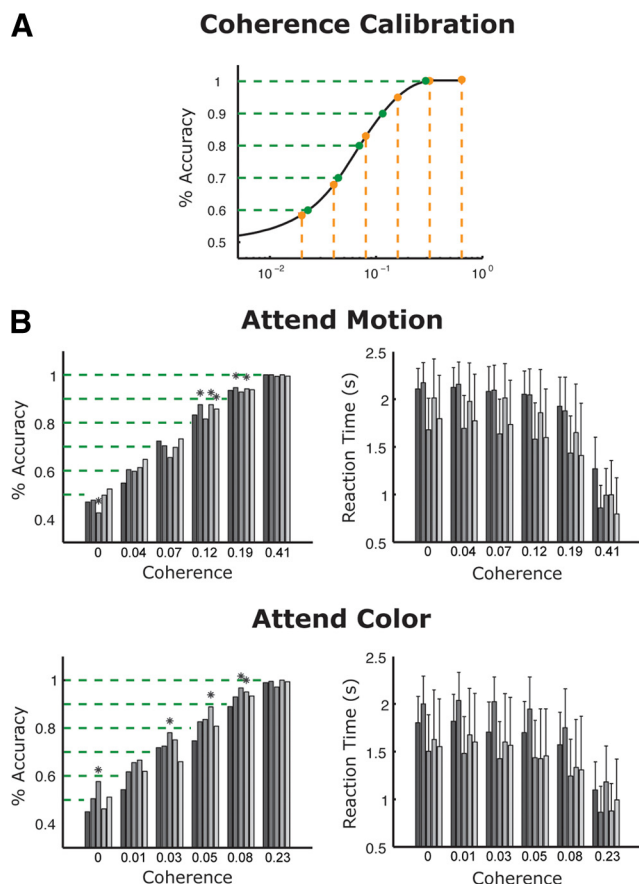
To verify that our calibrated coherence values did not appreciably change from the training phase to the experimental phase, accuracy values from the experiment were compared with the calibrated levels (Fig. 2B). For all subjects, accuracy within the experiment was very close to calibrated performance for each feature, where asterisks denote those individual subjects and coherence levels for which performance was significantly different from that predicted. Consistent with our task design, accuracy levels across attended features were not significantly different ( $F_{(1,4)} = 0.02$ ;  $p = 0.89$ ), and there was likewise no significant attended feature by coherence level interaction ( $F_{(5,20)} = 1.79$ ;  $p = 0.16$ ) to suggest differences for various coherence levels compared across attend-motion and attend-color conditions. As expected, within a task (e.g., attend-color) strong effects of feature coherence level on both accuracy and response time were



**Figure 1.** Motion/color discrimination task. Each run began with a distinctly colored text prompt defining the attended feature (motion or color) for the upcoming trials. To assist subjects in maintaining task set, the fixation cross was then rendered in the same color as the text prompt for the duration of the run. Subjects were cued to the upcoming stimulus by the dimming of the cross to a low-contrast gray, at which point the dot stimulus appeared within a central circular aperture of 7.5° diameter (dashed circle) for 2500 ms. For each trial, independent values for motion coherence and color proportion were chosen; across trials, these values varied in discrete steps from low to high (inset), in subject-specific fashion, to match predefined levels of accuracy from 50 to 100%.

observed ( $F_{(5,20)} > 19$ ,  $p < 0.00001$  for both effects). Of note, response times across subjects were reliably longer for the attend-motion than attend-color task ( $F_{(1,4)} = 14.0$ ,  $p = 0.02$ ; mean across all coherences, 0.194 s). This effect did not depend on the coherence level of the ignored feature, as no interactions between attended and ignored features for either accuracy or reaction time were observed ( $F_{(25,100)} < 1.3$ ,  $p > 0.15$  for all interactions as assessed by repeated-measures two-way ANOVA).

Since an irrelevant feature may have had a greater effect on the decision process when subjects made an error, we further segregated the behavioral data into correct and incorrect responses, excluding the upper two coherence levels (at which few errors were committed). For both attend-motion ( $F_{(1,4)} = 49.7$ ;  $p = 0.002$ ) and attend-color ( $F_{(1,4)} = 50.8$ ;  $p = 0.002$ ) conditions, there were strong effects of errors on reaction time across subjects, with error responses taking reliably longer for both conditions (attend-motion: mean difference, 0.37 s; attend-color: mean difference, 0.31 s). There was also a significant interaction between correct/error response times and coherence level for both attend-motion ( $F_{(3,12)} = 16.2$ ;  $p = 0.0002$ ) and attend-color ( $F_{(3,12)} = 3.92$ ;  $p = 0.037$ ) tasks. *Post hoc t* tests demonstrated no difference at chance (50%) performance for either task (values of  $p > 0.5$ ), but increasing RT differences with increasing feature coherence [attend-motion: 60% performance, 0.05 s;  $t_{(4)} = 2.4$ ,  $p = 0.08$  (NS); 70% performance, 0.12 s;  $t_{(4)} = 6.4$ ,  $p = 0.003$ ; 80% performance, 0.26 s;  $t_{(4)} = 6.5$ ,  $p = 0.003$ ; attend-color: 60% performance, 0.10 s;  $t_{(4)} = 3.8$ ,  $p = 0.02$ ; 70% performance, 0.19 s;  $t_{(4)} = 5.2$ ,  $p = 0.007$ ; 80% performance, 0.25 s;  $t_{(4)} = 3.7$ ,  $p = 0.02$ ]. However, neither the number of errors nor the response times varied significantly with the coherence of the ignored feature, regardless of whether motion or color was attended ( $F_{(5,20)} < 1.87$ ,  $p > 0.14$  for all comparisons). We were thus confident that our highly trained subjects did not show significant behavioral effects of the ignored stimulus.



**Figure 2.** *A*, Coherence calibration. Based on performance during training on predetermined coherence values (orange points), diffusion model fits were generated for each subject with respect to each feature (motion or color); sample gray curve shown for a hypothetical motion coherence calibration). Calibrated coherence values predicted to generate 50, 60, ..., 100% performance were extrapolated from these models (green points). *B*, Subject behavioral performance. Mean accuracy and reaction time values for motion (top) and color (bottom) tasks for each subject are shown across the calibrated coherence values. Each of the five subjects is represented by a different grayscale color; each coherence level is represented by the geometric mean of the actual coherence values across subjects. For accuracy, asterisks denote significant deviation from expected accuracy ( $p < 0.05$ ). For reaction times, error bars represent SDs.

## Summary maps

### Main effects of task and stimulus type

A whole-brain, voxelwise contrast of all conditions versus baseline identified a number of regions that were both reliably activated relative to baseline and also similar to those seen previously (Binder et al., 2004; Grinband et al., 2006; Heekeren et al., 2006; Philiastides and Sajda, 2007; Ploran et al., 2007; Thielscher and Pessoa, 2007; Kayser et al., 2010), including the bilateral IPS, IFS, frontal eye fields (FEFs), and anterior insula (aINS). To confirm that we were finding the expected modulatory effects of attention in sensory regions, we contrasted the main effects of the attend-color and attend-motion tasks using a fixed-effects analysis with an overlap requirement. As expected based on previous studies, we identified greater activity in occipital areas overlapping our MT+ ROI during attend-motion trials, and in occipital areas overlapping V4 during attend-color trials (data not shown).

### Parametric effects of coherence level

To identify those regions whose activity varied with the coherence level, we examined linear contrasts testing for a parametric effect of coherence in both color and motion tasks. Figure 3 shows parametric contrasts for both the attend-motion and

attend-color conditions, as well as for the ignored features, based on a fixed-effects analysis with an overlap requirement [ $p < 0.001$ , uncorrected, and  $p < 0.05$  in at least three of five subjects (Friston et al., 1999), shown at  $p < 0.005$ , uncorrected, for display purposes] (see Materials and Methods). In line with our previous results (Kayser et al., 2010), BOLD activity varied negatively with either attended motion or attended color coherence in the mIPS, IFS, aINS, and other regions (Fig. 3*A,B*) (for all regions, see Table 1). Posterior regions differentiated the two conditions: in the attend-motion condition, an area consistent with the location of the motion-sensitive region MT+ could be noted, whereas in the attend-color condition, more ventral, posterior regions including the human color-sensitive area V4 could be seen. When the color task was parameterized by the ignored motion coherence value, a different pattern emerged: three regions including MT+, a portion of more lateral IPS, and IFS demonstrated a significant positive rather than negative parametric variation with motion coherence (Fig. 3*A*, bottom). Thus, the parametric effect of motion coherence in these regions inverts when motion is ignored. This “inversion” of the parametric effect was not observed in these areas when the parametric effect of color coherence was compared across attended and ignored conditions (Fig. 3*B*, bottom) (see Discussion). To better understand why the ignored condition gave rise to this inverted effect, we next evaluated the time course of the BOLD response within independently defined ROIs.

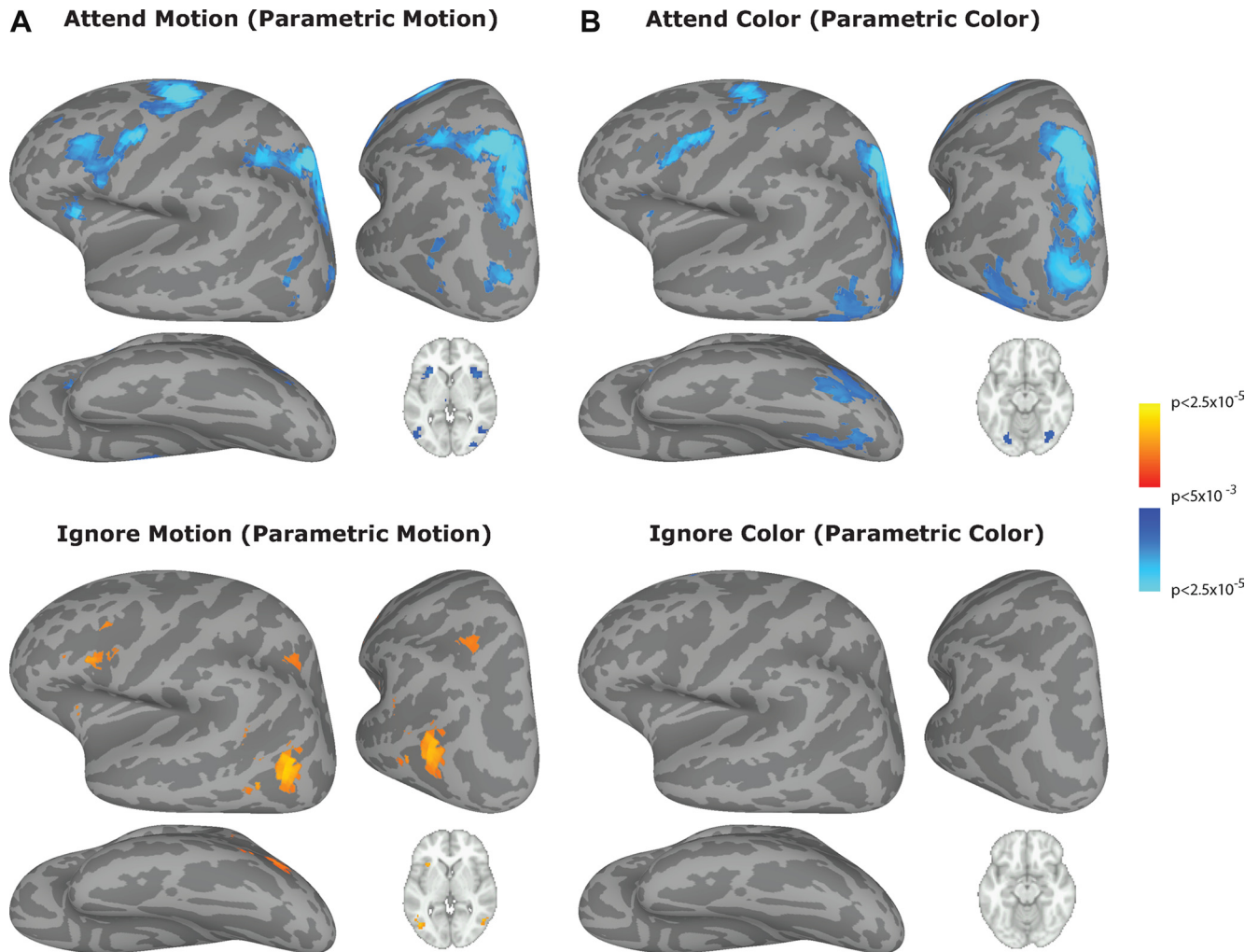
## Effects of attention

### Region of interest analyses

Our analysis of individual regions evaluated the time course of BOLD activity. As in our previous work, we examined both the amplitude of the peak response, and the time to the peak response. However, because the time to the peak response in the ignored condition is driven by the attended feature, it is less informative in this condition, and we thus focused on peak amplitude. To define ROIs, we identified those areas that showed a maximal negative parametric effect in a regression including both attend-motion and attend-color conditions during independent runs of the task acquired for each subject (see Materials and Methods). We focused our further evaluation on ROIs that were likely to be important to the task based on previous findings [MT+ (MNI coordinates,  $-46 -74 -2$ ), V4 ( $-28 -84 -22$ ), and mIPS ( $-24 -71 55$ )], and/or that demonstrated a significant parametric effect in the ignore-motion condition [MT+ and IFS ( $-42 5 29$ )], as these regions were a priori more likely to be implicated in attentional processes (see Introduction and Discussion). Nonetheless, for completeness, all ROIs demonstrating significant parametric effects of attended and ignored features can be found in Table 1.

MT+ demonstrated significant parametric effects with respect to feature coherence level in both the attended and ignored conditions. As evident in the peak amplitudes (Fig. 4), two-way repeated-measures ANOVA did not reveal a significant difference between attended features (color vs motion:  $F_{(1,16)} = 2.87$ ,  $p = 0.13$ ) but did show a significant effect of peak amplitude across the attend/ignore conditions ( $F_{(1,16)} = 38.2$ ;  $p < 0.0001$ ) as well as an interaction between feature and attention ( $F_{(1,16)} = 4.71$ ;  $p = 0.045$ ). *Post hoc t* tests revealed a significant inverse variation with motion coherence during the attend motion condition ( $t_{(4)} = -2.26$ ;  $p = 0.043$ ). However, when the BOLD activity in MT+ was parameterized by the ignored motion coherence in the attend-color condition, an effect opposite that of attend-motion was seen: a significant positive parametric varia-





**Figure 3.** Whole-brain parametric fixed-effects contrasts across the four attend/ignore motion/color conditions. All contrasts are shown at  $p < 0.005$ , uncorrected, for display purposes, after being defined at  $p < 0.001$ , uncorrected, with an additional overlap requirement (see Materials and Methods). **A**, Motion, both attended (top) and ignored (bottom). The blue color scale denotes negative parametric variation of the BOLD response with motion coherence level; the red-yellow color scale denotes positive parametric variation. As evident in the figure, a subset of brain regions that demonstrated a negative parametric variation with motion coherence in the attend-motion condition also demonstrated a positive parametric variation with motion coherence in the ignore-motion condition. (Note that the attend-motion and ignore-color conditions, as well as attend-color and ignore-motion conditions below, represent the same trials with a different parameterization.) The inset slice is taken from  $z = -4$  (MNI space). **B**, Color, both attended (top) and ignored (bottom). As for motion, the blue color scale denotes negative parametric variation with coherence level (in this case, for color), whereas a yellow color scale denotes positive parametric variation. A similar network is seen in the attend-color condition as was seen in the attend-motion condition, although with pronounced activity in V4. For the ignore-color condition, however, no significant correlations with ignored color coherence were found. The inset slice is taken from  $z = -21$  (MNI space). See also supplemental Figure S1 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

tion in peak amplitude ( $t_{(4)} = 4.55$ ;  $p = 0.005$ ). Interestingly, the MT+ ROI also demonstrated a strong negative parametric effect of attended color coherence for peak amplitude ( $t_{(4)} = -3.60$ ;  $p = 0.011$ ), in keeping with other accounts of the BOLD response to bound stimulus features (O'Craven et al., 1999; Sohn et al., 2005). Effects in the ignored color condition did not reach significance ( $p > 0.85$ ).

A similar but less pronounced effect was seen for V4. For peak amplitude, only the effect of attention reached significance ( $F_{(1,16)} = 9.3$ ,  $p = 0.0077$  via repeated-measures two-way ANOVA). *Post hoc* *T* tests demonstrated a trend effect of attention on the peak amplitude in the attend-color condition only ( $T_{(4)} = -1.67$ ;  $p = 0.085$ ). Effects in the ignored color and motion conditions did not reach significance (all values of  $p > 0.15$ ).

In mIPS, a two-way repeated-measures ANOVA revealed a significant effect of attention on peak amplitude ( $F_{(1,16)} = 37.98$ ;  $p < 0.0001$ ), but no effect of feature and no interactions between feature and attention (all values of  $p > 0.65$ ). In keeping with

these findings, peak amplitude showed significant negative parametric effects for both attend-motion and attend-color conditions by *post hoc* *t* tests (all values of  $t_{(4)} < -2.6$ ; all values of  $p < 0.03$ ), but no significant parametric effect in the ignore conditions (values of  $p > 0.4$ ).

IFS also demonstrated a significant parametric effect of attention on peak amplitude ( $F_{(1,16)} = 24.7$ ;  $p = 0.0001$ ). No significant effect of feature, or interaction between feature and attention, was seen (all values of  $p > 0.65$ ). *Post hoc* *t* tests demonstrated a significant negative parametric effect on peak amplitude ( $t_{(4)} = -2.3$ ;  $p = 0.041$ ) in the attend-motion condition, and a trend toward a negative parametric variation in the peak amplitude in the attend-color condition ( $t_{(4)} = -1.93$ ;  $p = 0.063$ ). In contrast with mIPS, the parametric effect with peak amplitude in the ignore-motion condition became positive ( $t_{(4)} = 2.23$ ;  $p = 0.045$ ).

To summarize, the finding of a negative parametric variation in the peak amplitude of the attended motion stimulus, but a

**Table 1.** Areas showing significant univariate activity during performance of the task, in MNI coordinates

Region	Hemi	Coordinates			Tstat	Overlap
		X	Y	Z		
Attend-motion						
SFG	L	−24	−11	58	−19.76	5
mIPS	L	−24	−69	55	−16.34	5
SFG	R	30	−9	52	−15.01	5
aIPS	L	−38	−46	45	−9.61	5
SMA	M	0	1	63	−16.71	4
SMA	M	−2	12	57	−13.31	4
mIPS	R	26	−76	44	−11.54	4
PMC	L	−46	−2	41	−11.48	4
mIPS	R	22	−71	62	−11.36	4
aINS	L	−30	22	8	−10.72	4
pIPS	L	−24	−76	26	−10.21	4
aINS	R	42	18	3	−9.54	4
FEF	L	−22	3	65	−9.41	4
MFG/IFS	L	−38	23	30	−8.90	4
IFS	L	−46	5	18	−8.55	4
MT+	L	−48	−72	3	−6.07	4
mIPS	L	−20	−77	46	−11.83	3
pOCC	L	−30	−94	17	−9.70	3
PMC	R	42	5	35	−9.05	3
ACC	R	8	23	40	−8.68	3
pOCC	R	36	−93	2	−8.58	3
aIPS	R	42	−46	54	−8.45	3
Ignore-motion						
MT+	L	−42	−70	5	8.80	4
MT+	R	48	−64	5	8.65	4
IFS	L	−42	9	33	8.60	4
pMTG	L	−57	−63	14	8.73	3
STG	R	55	−44	11	8.26	3
Attend-color						
mIPS	L	−24	−71	55	−18.63	5
mIPS	R	26	−76	44	−16.38	5
dOCC	R	36	−88	17	−10.63	5
pMFG	R	36	−7	54	−10.08	5
SMA	M	0	1	63	−13.83	4
FEF	L	−24	−9	58	−13.11	4
dOCC	L	−26	−86	19	−12.91	4
pIPS	R	36	−75	24	−12.10	4
ACC	R	6	16	49	−10.06	4
aIPS	R	30	−56	49	−9.82	4
MT+	R	46	−72	−9	−8.57	4
aIPS	R	42	−48	54	−8.18	4
V4	L	−24	−71	−18	−7.93	4
vOCC	L	−40	−80	−14	−7.27	4
PMC	R	48	2	39	−12.12	3
IFS	L	−40	5	33	−10.40	3
SPL	L	−32	−61	62	−8.11	3
aINS	L	−38	16	3	−7.29	3

The top section lists those regions that demonstrated a significant parametric effect with respect to attended motion coherence; note that the direction of the parametric effect (positive or negative) is indicated by the sign of the associated *T* value. In the middle section are listed those areas showing a significant parametric effect for ignored motion coherence, and in the bottom section can be seen those regions demonstrating a significant parametric effect of attended color coherence. Each of the three sections (top, middle, bottom) only includes areas whose main effect of task was positive (i.e., that activated during stimulus presentation). “Overlap” indicates the number of individual subjects who showed significant univariate effects within each region. No regions demonstrated a significant parametric effect of ignored color coherence. SFG, Superior frontal gyrus; aIPS, anterior intraparietal sulcus; SMA, supplementary motor area; PMC, premotor cortex; pIPS, posterior intraparietal sulcus; MFG, middle frontal gyrus; pOCC, posterior occipital cortex; ACC, anterior cingulate cortex; pMTG, posterior middle temporal gyrus; STG, superior temporal gyrus; dOCC, dorsal occipital cortex; pMFG, posterior middle frontal gyrus; vOCC, ventral occipital cortex; SPL, superior parietal lobule.

positive parametric representation of the peak amplitude for the ignored motion stimulus, was observed within some, but not all, regions of the network. As the data above demonstrate, the effect was most pronounced in MT+ and IFS.

### Model of attention

To explain how the parametric changes in BOLD response in MT+ in both the attend-motion and ignore-motion conditions might vary, we created a model in which baseline stimulus-driven (“bottom-up”) activity was modulated by a multiplicative attentional (“top-down”) factor, informed by previous work on visual attention in primates (Treue and Maunsell, 1996, 1999; Maunsell and Treue, 2006) (Fig. 5A). In attempting to model the peak amplitude of the BOLD response for the different coherence values across the attend- and ignore-motion conditions, we assumed that the magnitude of feature-based attention differed from the baseline value for the duration of the reaction time, and returned to baseline thereafter; it was not constrained to be the same magnitude for the attend-motion and attend-color conditions. We also assumed that the bottom-up input was the same for a given motion coherence regardless of whether motion was attended or ignored across the full 2500 ms of the stimulus. Finally, no a priori parametric constraints were placed on the relationship between the magnitude of the bottom-up response and the different motion coherence values. We focused on motion processing in MT+ rather than on color processing in V4, because unlike MT+, which has been shown to respond parametrically to motion coherence (Britten et al., 1993), mean V4 activity should not clearly respond to variations of color proportion within a stimulus of otherwise constant color content during the ignore-color task. In addition, because of our desire to focus on motion coherence, the color stimulus was not optimized to elicit strong responses in V4 (see also Discussion).

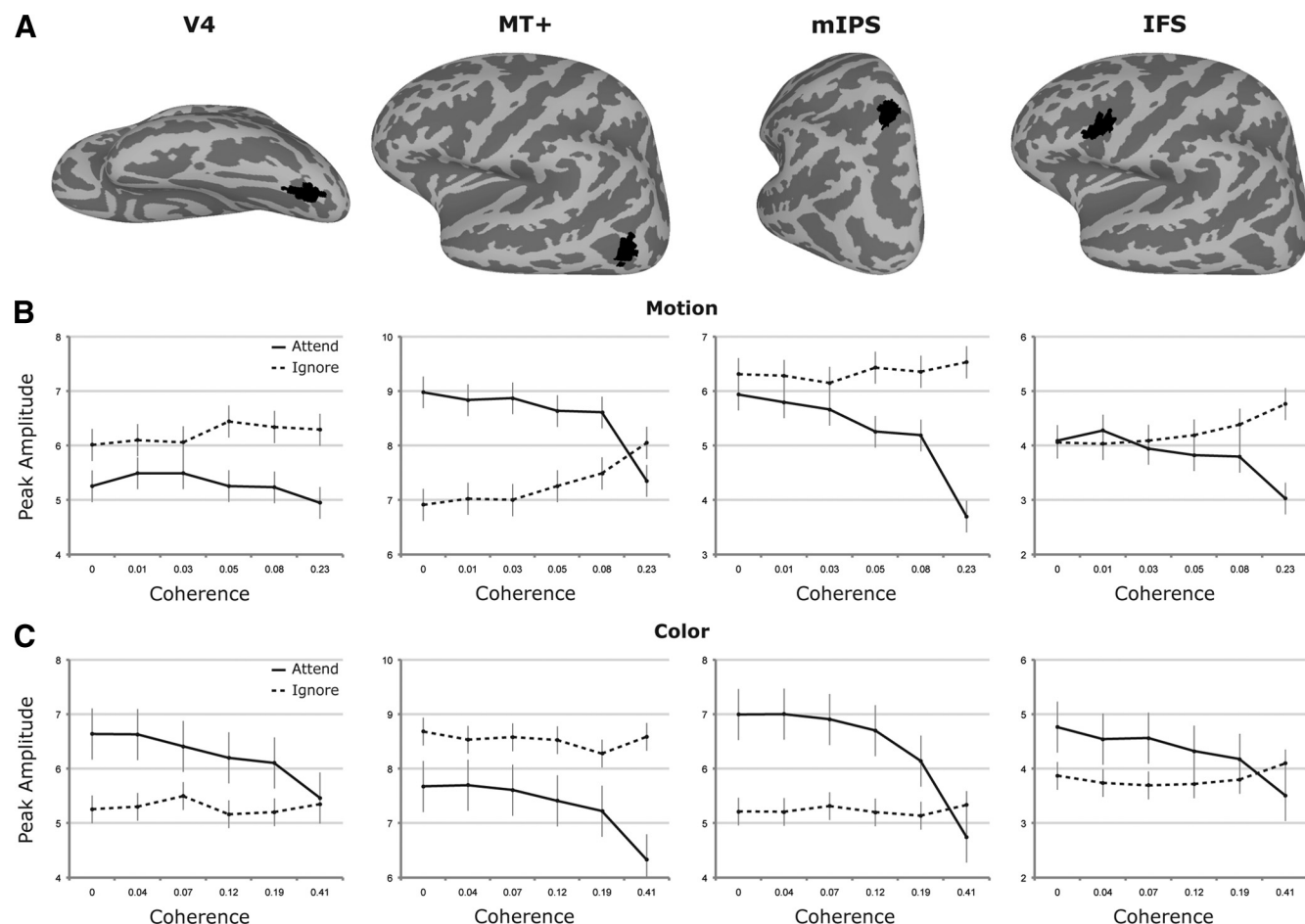
This simple model explained the data well (Fig. 5B). Specifically, the correlations between the values predicted by the model and the observed data were highly significant in four of five subjects (all values of  $p < 10^{-5}$ ), and of trend significance in the other ( $p = 0.075$ ). As shown in Figure 5C, across subjects the multiplicative factor of 1.55 in the attend-motion condition was both significantly greater than zero ( $t_{(4)} = 2.7$ ;  $p = 0.05$ ) and significantly greater than the value for the ignore-motion condition (paired  $t_{(4)} = 3.0$ ;  $p = 0.04$ ). Moreover, the bottom-up input was directly, not inversely, proportional to the motion coherence ( $t_{(4)} = 3.1$ ;  $p = 0.04$ ) (Fig. 5D). Thus, these values suggest that a direct parametric variation in the input-related activity and a multiplicative top-down factor alone can potentially explain the change in the direction of the parametric MT+ BOLD response from positive to negative when subjects switch from ignoring to attending motion.

### Connectivity

To further support the idea that top-down signals related to attention or other processes (e.g., accumulation) influence the response in MT+, we used Granger causality to evaluate the primary direction of information flow between two posterior regions (MT+, V4) and two anterior regions (mIPS and IFS). If IFS and mIPS are engaged in top-down control, whether attentional or otherwise, of posterior cortical representations, we hypothesized that they should demonstrate significant Granger causal influences over both MT+ and V4. However, if the representation of motion coherence within IFS in the ignored condition is driven by bottom-up inputs, the influence might instead be directed from posterior to anterior. Our data demonstrate that both mIPS and IFS exert Granger causal influences on both MT+ and V4, regardless of the attended/ignored feature (Fig. 6A).

To assess the relative influence of IFS and mIPS on MT+ and V4, we performed conditional Granger causality analyses in which particular Granger causal influences were conditioned on





**Figure 4.** Peak amplitudes derived from time course data for selected ROIs. **A**, Surface maps illustrating the anatomical locations of ROIs V4, MT+, mIPS, and IFS. **B**, Peak amplitudes for each of the four ROIs  $\pm$  95% confidence intervals for the attend-motion (solid lines) and ignore-motion (dashed lines) conditions. Negative parametric trends are significant ( $p < 0.05$ ) in the attended condition for MT+, mIPS, and IFS; positive parametric trends in the ignored condition are significant for MT+ and IFS only. **C**, Peak amplitudes for each of the four ROIs  $\pm$  95% confidence intervals for the attend-color (solid lines) and ignore-color (dashed lines) conditions. Negative parametric trends are significant ( $p < 0.05$ ) for MT+ and mIPS, reaching trend significance for V4 ( $p = 0.085$ ) and IFS ( $p = 0.063$ ) for the attended condition. No parametric trends are significant for the ignored condition.

activity in other regions. For example, to assess whether the influences between mIPS, MT+, and V4 were contingent on IFS activity, one could condition on IFS. In this data set, conditioning on activity within IFS did not disrupt the influence of mIPS on MT+ and V4 (Fig. 6B). However, conditioning on mIPS activity rendered the influences between IFS, MT, and V4 no longer significant (Fig. 6C), suggesting that mIPS provided a more direct top-down effect on these sensory areas.

## Discussion

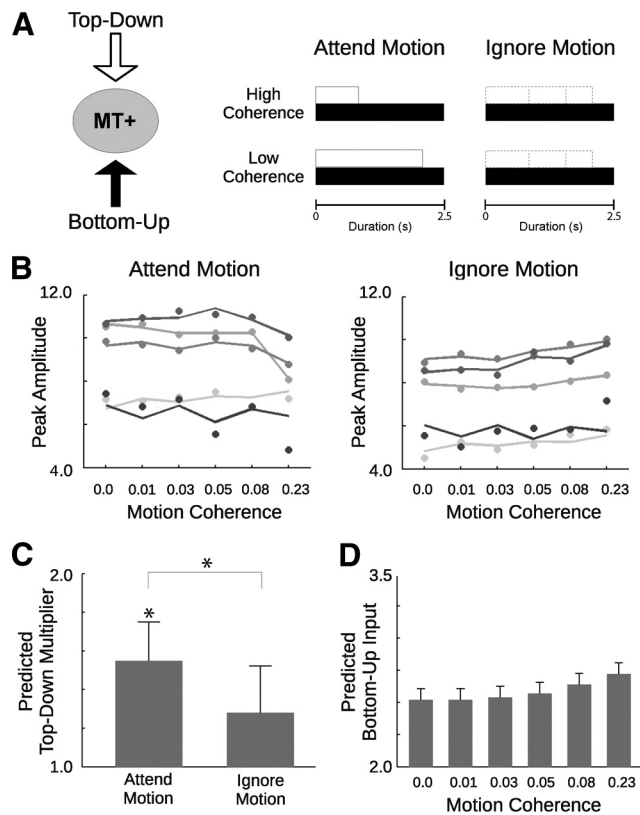
In this study, we demonstrated that regions within the perceptual decision-making network reflect divergent patterns of activity for attended and ignored stimulus features, a finding that argues that attention can profoundly affect the interpretation of fMRI responses to manipulation of an experimental variable. Indeed, the BOLD response to variations in motion coherence inverts from negatively parametric (attended motion) to positively parametric (ignored motion), depending on the locus of feature-based attention. In keeping with previous theories about sensory regions (Maunsell and Treue, 2006), modeling and multivariate analyses suggest that this “inversion” can be accounted for in MT+ by a top-down factor that acts in multiplicative fashion on bottom-up activity.

These findings also suggest that a subset of regions within the decision-making circuit devote processing resources to the ig-

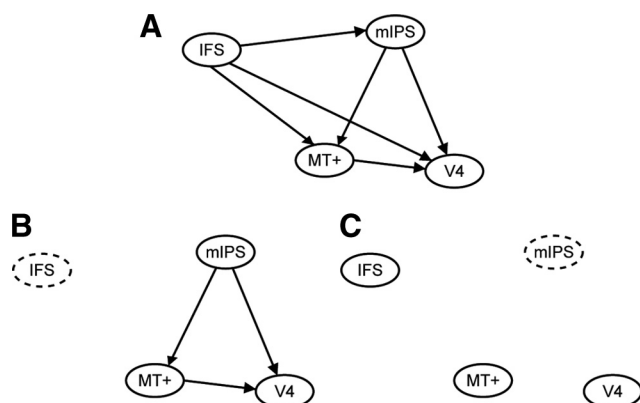
nored stimulus. In a whole-brain analysis, we noted such responses within the IFS, anterior insula, lateral IPS, and MT+; and in an independent ROI-based analysis, we confirmed these differences within IFS and MT+. In frontal areas, this activity would be consistent with a role in feature-based attention, with increasing attentional control required as the ignored sensory stimulus becomes stronger. The significant Granger causal influence from IFS to MT+ supports this model, and is less consistent with one in which anterior activity simply reflects bottom-up input. Moreover, that these regions respond parametrically to a stimulus irrelevant to the current task suggests that they are less likely to contribute to accumulation of sensory evidence.

## Sensory regions

The first aim of these experiments was to investigate the effects of the attentional manipulation on sensory representations available to decision-related areas during perceptual decision making. Our findings demonstrate that the bottom-up input provided to MT+ increases monotonically with motion coherence, but when motion coherence is attended, this relationship inverts because of the increasing influence of top-down attentional input. These results potentially reconcile our recent finding of a negative parametric effect in MT+ with a previous report showing a direct relationship (Rees et al., 2000). Based on our results, a positive



**Figure 5.** Model of MT+ activity across attend and ignore conditions. **A**, Bottom-up inputs (black) are assumed to be present for the duration of the stimulus, but modulated in multiplicative fashion by top-down inputs (white) for the duration of the reaction time. In the attend-motion condition, high motion coherence is typically associated with shorter reaction times, whereas low motion coherence is associated with longer reaction times, represented by the width of the white bars. In the ignore-motion condition, reaction time is determined by the independently varying color proportion, represented by the variable-length white bars. **B**, Fits of the model (solid lines) to the actual peak amplitudes (filled circles) for the attend- and ignore-motion conditions across all five subjects (gray-scale lines). **C**, Predicted values for the top-down multiplier in the attend-motion and ignore-motion conditions, across the five subjects  $\pm$  SEM.  $*p < 0.05$ . **D**, Predicted values for the bottom-up input across both attend-motion and ignore-motion conditions  $\pm$  95% confidence intervals.



**Figure 6.** Granger analysis results. BOLD activity within selected ROIs (V4, MT+, mIPs, IFS) was examined for Granger causal influences. **A**, The arrows indicate direction of influence for all significant connections. **B**, When conditioned on IFS, the Granger causal influences between mIPs, MT+, and V4 remained. **C**, When conditioned on mIPs, no Granger causal influences reached significance. These findings argue that mIPs exhibits a top-down influence on MT+ and V4 independently of IFS, and that the magnitude of the influence of IFS on these posterior regions may be at least partially mediated via mIPs.

parametric effect could also result for either, or both, of two reasons: (1) short stimulus presentations, or (2) attention directed to only one of multiple stimuli. If viewing times are quite short (e.g., 250 ms), attention may be applied equally throughout the presentation of all stimuli and thus does not in itself strongly differentiate attend/ignore-motion coherence conditions. In contrast, longer display times allow attention to potentially differentiate higher from lower motion coherences. Furthermore, if motion coherence is presented simultaneously in both attended and ignored fields, it is possible, perhaps via callosal connections, that the response to the ignored coherence will contribute to the BOLD response. Regardless of the etiology of this positive parametric effect, our previous study (Kayser et al., 2010) suggests that such a monotonically increasing effect would be consistent with a population response curve in MT+ whose width is greater than  $\sim 68^\circ$ —in keeping with measurements in macaque MT (Britten and Newsome, 1998). At this width, modeling suggests that the summed response of neurons/voxels to the random movements in a low motion coherence stimulus is equal to the summed responses of those few neurons/voxels tuned to the specific motion direction in a high motion coherence stimulus. Tuning curves wider than  $68^\circ$  therefore lead to a parametrically increasing response with increasing motion coherence.

Consistent with previous studies (O'Craven et al., 1999; Sohn et al., 2005; Buracas and Albright, 2009; Katzner et al., 2009), we also found that stimulus features bound to the same object (i.e., dot motion and dot color) were reflected in the BOLD signal even in regions thought to be poorly responsive to one of the features, such as MT+ to color. This effect was generally smaller than that of the preferred feature (Fig. 4), a finding supported by our modeling work. The multiplier effect for MT+, for example, was greater for attend-motion (1.55) (Fig. 5) than for attend-color (1.28). Intriguingly, these differences between attention for the preferred and bound features are also quite similar to those obtained by O'Craven et al. (1999) for a task in which subjects were required to attend to either motion, faces, or houses. In area MT/MST (averaged across both face and house responses in their Fig. 2), their BOLD data suggest that attention to motion, relative to baseline (attention to a stationary stimulus in the presence of a moving irrelevant stimulus), leads to a 1.52-fold increase in the average BOLD response as assessed by percentage signal change. Attention to a different feature of a moving object, compared with baseline, leads to a 1.29-fold BOLD increase. Thus, despite the very different paradigms, the data in the two papers are in reasonable agreement.

An additional consideration relates to V4, which did not demonstrate a parametric effect in the ignored color condition. We suspect that this finding derives from two related issues. Unlike MT+, which has been shown to respond parametrically to motion coherence (Britten et al., 1993), it is not clear that during the ignore-color task overall V4 activity should respond parametrically to variations of color proportion within a stimulus of otherwise constant color content. In addition, because of our desire to focus on motion coherence, the color stimulus was not optimized to elicit strong responses in V4 [as, for example, was the Mondrian-like stimulus used in our color localizer (Zeki et al., 1991)].

### Attention and accumulation in the frontal and parietal lobes

Interestingly, only a subset of the other regions within the decision-making circuit showed a parametric response to the ignored feature. The IFS region was strongly responsive to both

attended and ignored motion, consistent with previous work implicating the lateral prefrontal cortex (Curtis and D'Esposito, 2003; Miller and D'Esposito, 2005; Ranganath and D'Esposito, 2005) and, specifically, inferior frontal regions (Aron et al., 2004) in attentional and other control processes. However, in selecting regions of interest that were strongly (and negatively) parametrically modulated in both the attend-motion and attend-color conditions, we identified a ROI in mIPS (MNI coordinates,  $-24 -71 +55$ ) that did not show a significant positive parametric response to ignored motion but that was near a more ventrolateral parietal region in the surface map that did (Fig. 3) (MNI coordinates,  $-28 -58 +45$ ). One possible explanation for this dissociation is that IPS regions responsible for attentional control or representation of multiple features, and those IPS regions responsible for other processes, such as evidence accumulation, may be segregated within the dorsal parietal lobe. As noted above, attentional regions would potentially be active in both attend/ignore conditions, reflecting the presence of activity important for enhancement/suppression of relevant/irrelevant features, whereas accumulator regions [or those sensitive to other aspects of the decision, such as decision ambiguity (Hampshire et al., 2008)] would be expected to vary only with the attended feature, because only this feature is relevant to the decision. Our conditional Granger causality analyses argue that mIPS has a significant influence on both MT+ and V4 in either case [i.e., whether related to attention or to other factors (e.g., accumulation of sensory evidence)] consistent with other measures of timing between IPS and occipital regions (Lauritzen et al., 2009). Additionally, the interactions between IPS and mIPS are consistent with the multiple demand network hypothesis of Duncan and colleagues (Duncan, 2006), in keeping with the role of this circuit in flexible representations within multiple task contexts.

Previous studies also suggest that finding a representation of the ignored feature in nearby parietal regions should perhaps be unsurprising. Dorsal parietal lobe is thought to be part of a frontoparietal system that provides top-down attentional signals (Corbetta and Shulman, 2002; Duncan, 2006; Silver and Kastner, 2009). In the above scenario, the positive parametric variation in this area might reflect increasing attentional demands as the coherence of the ignored feature increases. In complementary fashion, work in object recognition has demonstrated that up to four objects, on average, can be retained in parietal cortex (Xu and Chun, 2009), whereas recent work suggests that the intraparietal sulcus maintains multiple features in working memory (Xu, 2007). Todd and Marois (2004), for example, identified a region within the intraparietal sulcus (Talairach coordinates,  $[+23 -59 +45; -22 -65 +42]$ ) later investigated by Xu (2007) that was responsive to multiple features. This area is very close to the lateral IPS region we have identified that is responsive to the ignored feature; it is thus possible that the function of this lateral IPS region could also be to maintain representations of the multiple features present in the color-motion stimulus, and to use them to influence temporo-occipital activity.

## Conclusions

In summary, we have demonstrated that attention has profound effects on the operation of a perceptual decision-making circuit. In addition to using mechanisms that modulate the BOLD activity of an attended stimulus, regions within this network also respond to an irrelevant stimulus in a manner distinct from the concurrent decision making process. The suspected utility of this activity lies in the possibility that irrelevant stimuli may subsequently merit attentional resources, despite the potential costs—

computational, metabolic, or otherwise—of encoding them. As well as supporting the significant effect of attention on neural activity during the selection of relevant stimuli, our results also argue for the role of attention in evidence accumulation independent of other functions, such as feature representation. Interesting future work might further explore the capacity of the decision-making system (Xu, 2007; Xu and Chun, 2009) and dissociate the effects of additional variables [such as confidence (Kiani and Shadlen, 2009) and previous knowledge (Liston and Stone, 2008)] on the nodes of the circuit.

## References

- Amano K, Wandell BA, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J Neurophysiol* 102:2704–2718.
- Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. *Trends Cogn Sci* 8:170–177.
- Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Res* 49:1154–1165.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7:295–301.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Britten KH, Newsome WT (1998) Tuning bandwidths for near-threshold stimuli in area MT. *J Neurophysiol* 80:762–770.
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1993) Responses of neurons in macaque MT to stochastic motion signals. *Vis Neurosci* 10:1157–1169.
- Buracas GT, Albright TD (2009) Modulation of neuronal responses during covert search for visual feature conjunctions. *Proc Natl Acad Sci U S A* 106:16853–16858.
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.
- Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci* 7:415–423.
- David SV, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron* 59:509–521.
- Duncan J (2006) EPS Mid-Career Award 2004: brain mechanisms of attention. *Q J Exp Psychol (Colchester)* 59:2–27.
- Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ (1999) Multisubject fMRI studies and conjunction analyses. *Neuroimage* 10:385–396.
- Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574.
- Grinband J, Hirsch J, Ferrera VP (2006) A neural representation of categorization uncertainty in the human brain. *Neuron* 49:757–763.
- Hampshire A, Thompson R, Duncan J, Owen AM (2008) The target selective neural response—similarity, ambiguity, and learning effects. *PLoS One* 3:e2520.
- Heekeren HR, Marrett S, Ruff DA, Bandettini PA, Ungerleider LG (2006) Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proc Natl Acad Sci U S A* 103:10023–10028.
- Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. *Nat Rev Neurosci* 9:467–479.
- Ho TC, Brown S, Serences JT (2009) Domain general mechanisms of perceptual decision making in human cortex. *J Neurosci* 29:8675–8687.
- Katzner S, Busse L, Treue S (2009) Attention to the color of a moving stimulus modulates motion-signal processing in macaque area MT: evidence for a unified attentional system. *Front Syst Neurosci* 3:12.
- Kayser AS, Sun FT, D'Esposito M (2009) A comparison of Granger causality and coherency in fMRI-based analysis of the motor system. *Hum Brain Mapp* 30:3475–3494.
- Kayser AS, Buchsbaum BR, Erickson DT, D'Esposito M (2010) The functional anatomy of a perceptual decision in the human brain. *J Neurophysiol* 103:1179–1194.
- Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324:759–764.
- Lauritzen TZ, D'Esposito M, Heeger DJ, Silver MA (2009) Top-down flow



- of visual spatial attention signals from parietal to occipital cortex. *J Vis* 9:18.1–14.
- Lee SH, Blake R, Heeger DJ (2007) Hierarchy of cortical responses underlying binocular rivalry. *Nat Neurosci* 10:1048–1054.
- Liston DB, Stone LS (2008) Effects of prior information and reward on oculomotor and perceptual choices. *J Neurosci* 28:13866–13875.
- Maunsell JH, Treue S (2006) Feature-based attention in visual cortex. *Trends Neurosci* 29:317–322.
- McNamee RL, Lazar NA (2004) Assessing the sensitivity of fMRI group maps. *Neuroimage* 22:920–931.
- Miller BT, D'Esposito M (2005) Searching for “the top” in top-down control. *Neuron* 48:535–538.
- O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. *Nature* 401:584–587.
- Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J Vis* 5:376–404.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Philiastides MG, Sajda P (2007) EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *J Neurosci* 27:13082–13091.
- Ploran EJ, Nelson SM, Velanova K, Donaldson DI, Petersen SE, Wheeler ME (2007) Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. *J Neurosci* 27:11912–11924.
- Ranganath C, D'Esposito M (2005) Directing the mind's eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Curr Opin Neurobiol* 15:175–182.
- Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput* 20:873–922.
- Rees G, Friston K, Koch C (2000) A direct quantitative relationship between the functional properties of human and macaque V5. *Nat Neurosci* 3:716–723.
- Roebroeck A, Formisano E, Goebel R (2005) Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* 25:230–242.
- Saad ZS, Chen G, Reynolds RC, Christidis PP, Hammett KR, Bellgowan PS, Cox RW (2006) Functional imaging analysis contest (FIAC) analysis according to AFNI and SUMA. *Hum Brain Mapp* 27:417–424.
- Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. *Trends Cogn Sci* 13:488–495.
- Silver MA, Shenhav A, D'Esposito M (2008) Cholinergic enhancement reduces spatial spread of visual responses in human early visual cortex. *Neuron* 60:904–914.
- Sohn W, Chong SC, Papathomas TV, Vidnyánszky Z (2005) Cross-feature spread of global attentional modulation in human area MT+. *Neuroreport* 16:1389–1393.
- Thielscher A, Pessoa L (2007) Neural correlates of perceptual choice and decision making during fear–disgust discrimination. *J Neurosci* 27:2908–2917.
- Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428:751–754.
- Tosoni A, Galati G, Romani GL, Corbetta M (2008) Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. *Nat Neurosci* 11:1446–1453.
- Treue S, Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382:539–541.
- Treue S, Maunsell JH (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neurosci* 19:7591–7602.
- Xu Y (2007) The role of the superior intraparietal sulcus in supporting visual short-term memory for multifeature objects. *J Neurosci* 27:11676–11686.
- Xu Y, Chun MM (2009) Selecting and perceiving multiple visual objects. *Trends Cogn Sci* 13:167–174.
- Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. *J Neurosci* 11:641–649.