

Shape-specific preparatory activity mediates attention to targets in human visual cortex

Mark Stokes^{a,b,1}, Russell Thompson^b, Anna C. Nobre^a, and John Duncan^b

^aDepartment of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom; and ^bMedical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge CB2 7EF, United Kingdom

Edited by Robert Desimone, Massachusetts Institute of Technology, Cambridge, MA, and approved September 25, 2009 (received for review May 14, 2009)

The mechanisms of attention prioritize sensory input for efficient perceptual processing. Influential theories suggest that attentional biases are mediated via preparatory activation of task-relevant perceptual representations in visual cortex, but the neural evidence for a preparatory coding model of attention remains incomplete. In this experiment, we tested core assumptions underlying a preparatory coding model for attentional bias. Exploiting multivoxel pattern analysis of functional neuroimaging data obtained during a non-spatial attention task, we examined the locus, time-course, and functional significance of shape-specific preparatory attention in the human brain. Following an attentional cue, yet before the onset of a visual target, we observed selective activation of target-specific neural subpopulations within shape-processing visual cortex (lateral occipital complex). Target-specific modulation of baseline activity was sustained throughout the duration of the attention trial and the degree of target specificity that characterized preparatory activation patterns correlated with perceptual performance. We conclude that top-down attention selectively activates target-specific neural codes, providing a competitive bias favoring task-relevant representations over competing representations distributed within the same subregion of visual cortex.

top-down bias | visual attention

Our perception of the external environment is continually shaped by internal goals and expectations. In particular, the mechanisms of attention fine-tune perception to facilitate the analysis of sensory input that is most likely to be relevant for behavior. Goal-oriented attentional biases determine the neural impact of sensory stimulation, enhancing (or suppressing) the neural response to information that is relevant (or irrelevant) to current task demands (1).

Influential models of attention have proposed that attentional bias is coded directly within the baseline activation profile of behaviorally relevant perceptual representations (1–4). According to a biased competition model of attention (1), re-entrant feedback, via top-down control mechanisms, increases spontaneous firing within neural populations that code task-relevant perceptual information. At the level of single-unit neurophysiology, an attentional cue for a specific target stimulus modulates baseline activity in neurons that preferentially respond to that stimulus (5, 6). Elevated baseline activity, distributed across a specific cell assembly, could potentiate, or prime, subsequent neural processing. According to this model, specificity of attentional bias is determined by the similarity between neural populations activated by top-down preparatory attention and subsequent stimulus-driven input.

Preparatory coding across cell assemblies that are co-extensive with corresponding perceptual representations could provide a crucial link between neural coding for goals and expectations in higher-level brain areas, including prefrontal and parietal cortex, and the attentional modulations observed in perceptual cortex (7). Despite the broad potential appeal of such a model, the current literature lacks critical evidence for selective activation of distinct perceptual representations in the human brain. Func-

tional magnetic-resonance imaging (fMRI) studies of attention have shown differential increases in baseline activity across retinotopically specific subregions of visual cortex during spatial attention (8–10). More generally, when a task requires attention to one visual feature, such as color or motion (11–14), then activation levels within visual areas specialized for processing the attended feature increase. However, if preparatory attention for specific target stimuli exploits the representational structure of perceptual cortex as we predict, top-down mechanisms must also be able to activate selectively target-specific representations amongst competing non-target representations, even if they are coded across intermingled cell-assemblies within the same cortical subregion. Until recently, fMRI has been unable to differentiate activity from functionally distinct, yet spatially overlapping neural populations. However, developments in image processing now provide analytical methods that can discriminate activation patterns with a remarkable degree of selectivity (15, 16). For example, recent fMRI studies of working memory have used pattern analytic techniques to demonstrate that top-down mechanisms can selectively maintain orientation-specific patterns of activity in early visual cortex following stimulus offset (17, 18). Evidence from visual imagery (19) further demonstrates top-down generation of shape-specific activity patterns in higher-level visual shape-processing areas, including lateral occipital complex (LOC). In this experiment, we apply similar methods to investigate how top-down mechanisms may be used in attentional control. We ask where attentional control signals are implemented in the human visual system, and examine their time-course and their relationship to behavior.

Participants were cued, via an auditory tone, to attend for a specific shape in preparation for faint stimuli that were embedded within visual noise. To determine the pattern-similarity between population codes activated during preparatory attention and those activated by direct visual stimulation, we first characterized the population response in visual cortex to target stimuli during a separate pattern-localizer task. Once the target-specific perceptual representations were characterized, we then tested the extent to which the same population codes were activated following the corresponding attention cue. These cross-comparison analyses confirmed that an arbitrarily assigned attentional cue can trigger preparatory activation of target-specific representations encoded among overlapping neural populations within shape-processing regions of human visual cortex. Target-specific attentional biases were maintained for the duration of attentional allocation. Importantly, within anterior LOC, preparatory activity that more closely resembled

Author contributions: M.S., R.T., A.C.N., and J.D. designed research; M.S. and R.T. performed research; M.S. analyzed data; and M.S., R.T., A.C.N., and J.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: mark.stokes@sjc.ox.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0905306106/DCSupplemental.

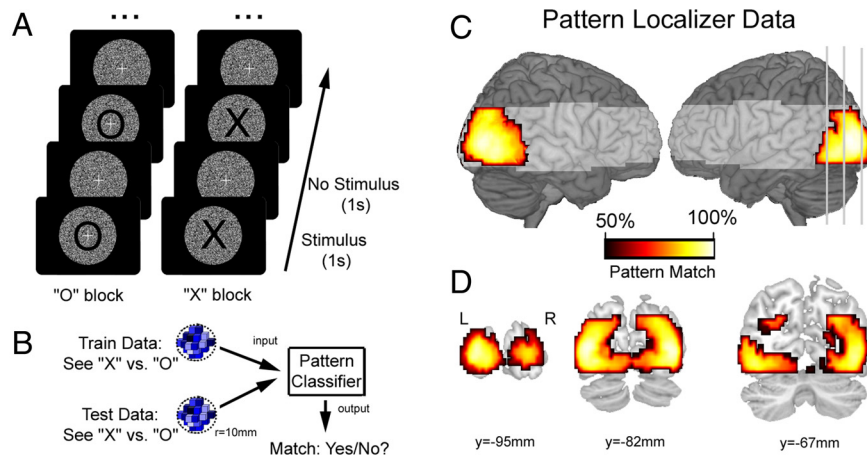


Fig. 1. The pattern-localizer task identified target-specific neural populations in visual cortex. (A) Participants alternately viewed target stimuli (X or O) presented at the centre of the visual display within a circular aperture of dynamic white noise. Participants monitored the stream of X, or O stimuli for an occasional stimulus presented in a smaller font (12.5% targets, randomly distributed). (B) We first performed pattern analysis to verify that our procedure can extract neural activation patterns specifically associated with the two target stimuli. A searchlight analysis examines each region of cortex (sphere, $r = 10$ mm): each pattern classifier was trained to discriminate between patterns for "X" or "O" estimated from a subset of training data, and classification performance was assessed on an independent test data set. Classifier output indicated whether the stimulus category of the test data matched the classifier prediction. (C) Left and right lateral views of the rendered cortical surface illustrate above-chance discrimination extending throughout left and right visual cortical areas (corrected for multiple comparisons, $P_{FDR} < 0.001$). Shading represents brain areas beyond the field of view of our data acquisition protocol, and gray lines indicate coronal slices shown in (D).

target-specific stimulus-driven activity was associated with higher detection rates for target stimuli, providing the essential link between behavior and this putative neural substrate for preparatory attention.

Results

Pattern-Localizer Task. A pattern-localizer task was used to characterize, within each participant, visual activation patterns associated with the two target stimuli used in the attention task (see Fig. 1, *Materials and Methods*, and *SI Text*). Initially, the pattern-localizer data were examined via univariate analyses to test whether target stimuli elicited different mean levels of neural activity within any brain area. Consistent with previous evidence that these letter stimuli elicit equivalent level of visual activity (19), there were no differences between activity associated with either the letter X or O in any brain areas ($P_{FDR} > 0.44$). Despite equivalent mean levels of brain activity across the two conditions, multivoxel analysis of visual activation patterns (Fig. 1B) could accurately discriminate between viewing X and O (see Fig. 1C and D). In particular, a searchlight analysis (spherical searchlight, $r = 10$ mm) revealed activation patterns throughout visual cortex that discriminated between the two perceptual conditions. These robust MVPA results suggest that the pattern-localizer task provides an effective data set from which to characterize discriminative activation patterns associated with differential population coding for the target stimuli used in the attentional task.

Attention Task. During the attention task, participants were cued to attend for either the letter X or O via auditory tones (220 Hz or 1,100 Hz for 100 ms, counterbalanced within participants; see Fig. 2A, *Materials and Methods*, and *SI Text*). During target-present trials (25% of trials), up to two sequentially presented target stimuli were displayed over the dynamic noise. During non-target trials, either non-target letters were presented (25% of trials), or no letter (50% of trials) was presented throughout the 16-s trial period. At the end of each trial, participants indicated via a button press response whether one, two, or no targets were presented. On average, participants detected 88% of all target stimuli. To avoid contamination from stimulus-

driven visual information, all data following the onset of a letter (target or non-target) stimulus were discarded from analyses of attentional bias. Similar to the localizer data, initial univariate analyses did not identify any brain regions that responded differentially during periods of attending for X or O ($P_{FDR} > 0.124$), thus confirming equivalent levels of brain activity during both attentional states. Similarly, no differences were observed between X or O averaged across the attention and localizer task (SeeX + AttendX vs. SeeO + AttendO; $P_{FDR} > 0.21$).

Despite similar overall levels of activity, we found that we could reliably discriminate between activation patterns associated with attending for X or O using multivariate classifiers trained to discriminate between viewing X or O during the pattern-localizer task (see Fig. 2B). Accurate cross-comparison implies a significant degree of similarity between the neural ensembles activated during specific states of preparatory attention and corresponding stimulus-driven perception. Searchlight analyses revealed a highly specific cluster of significant cross-comparison results within left and right extrastriate visual cortex (peak xyz coordinates: $-45, -78, -11$; $48, -75, -8$) (Fig. 2C), corresponding closely to previously defined posterior LOC (20). Interestingly, there was no evidence for accurate cross-comparison within earlier visual areas that code for lower-level visual features, such as orientation and precise retinotopy.

Further analyses examined attentional biases within predefined regions-of-interest (ROIs) associated with visual shape processing (20): anterior and posterior LOC (aLOC, pLOC). Cross-comparison classification accuracy, averaged across the duration of the attention trial (4–20s postcue: t3, t4, ..., t10, excluding all time bins following the onset of visual stimuli), was significantly above chance (50%) within all ROIs ($P < 0.05$, one sample t -test) (Fig. 3A). Next, we performed a three-way ANOVA, with factors for ROI (aLOC, pLOC), Laterality (left, right) and Time (0- to 22-s post-cue: t1, t2, ..., t12). We also included a covariate comprising mean target detection rates for each participant (Accuracy). Main effects were observed for ROI [pLOC > aLOC: $F(1, 14) = 6.8, P = 0.020$] and Time [$F(11, 154) = 2.6, P = 0.024$], and there were no significant terms including Laterality. There was significant interaction between ROI and Time [$F(11, 154) = 2.1, P = 0.025$]. As illustrated in Fig. 3B, target-specific biases emerged within both

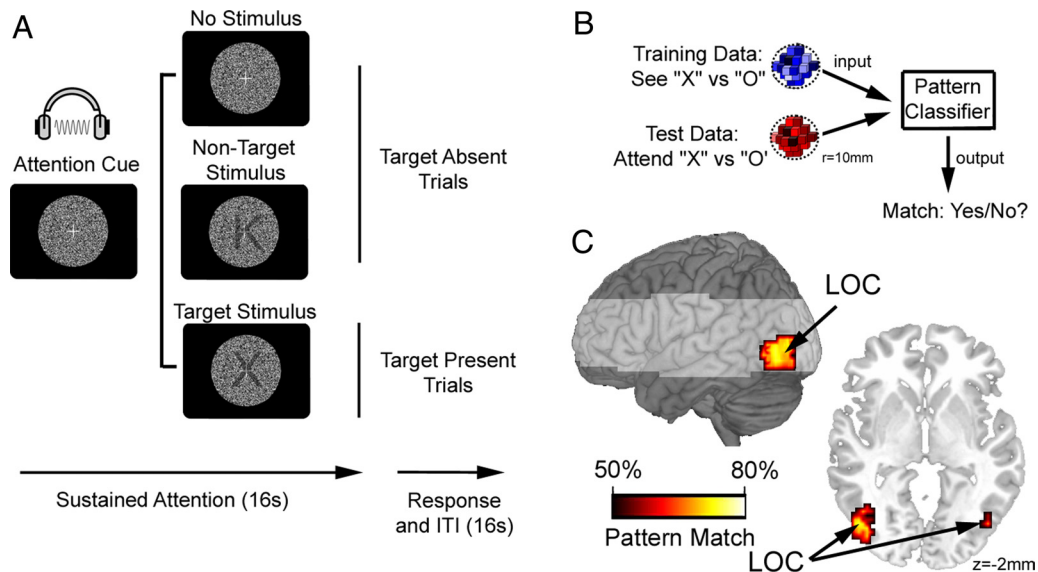


Fig. 2. Preparatory attention activates target-specific neural populations in visual cortex. (A) Participants were cued to attend for either the letter X or O via auditory tones. Target, and non-target, stimuli were semitransparent (range: 83–90% transparency) letters presented over dynamic visual noise. For illustration only, stimuli are depicted here with relative high visibility (60% transparency). Activation data following the onset of a target, or non-target, stimulus were discarded from all analyses of attentional bias. (B) Classification algorithms were trained on each 10 mm sphere of brain data to discriminate between neural patterns for "X" and "O" stimuli observed during the localizer task, and classification accuracy was tested on the neural responses measured whilst participants attended for each of the target letters. Classifier output indicated the match between perception-discriminative patterns identified in training data from the localizer task and the attention-discriminative patterns observed during the test data from the attention task. (C) Searchlight analyses identified a specific subregion of visual cortex where classifiers trained to discriminate between perceptual states for X or O could also discriminate between the cued attentional states (attend X vs. O; $P_{FDR} < 0.05$), corresponding to lateral occipital complex (LOC).

anterior and posterior LOC with a delay of 4–6 s after the cue onset, corresponding to the hemodynamic lag. Shape-specific activity then persisted until approximately 4 s after the end of the attention trial, before returning to baseline (i.e., chance at 50%). The sustained profile for attentional modulation was more pronounced in pLOC than aLOC, but the time-course in both regions contrasted with the transient profile of target-specific activity triggered by the visual presentation of target stimuli (Fig. 3C).

The omnibus ANOVA also revealed an interaction between ROI and Accuracy [$F(1, 14) = 6.7, P = 0.021$]. Correlation analyses performed on classification data averaged across the duration of the attention trial (4- to 20-s post-cue: t_3, t_4, \dots, t_{10}) confirmed a significant relationship between behavior and classification accuracy in aLOC, but not in pLOC (Fig. 3D). There was also an interaction between ROI, Time and Accuracy [$F(11, 154) = 2.0, P = 0.034$], prompting us to examine how this relationship evolved throughout the trial within each region. The correlation between behavior and attentional bias was assessed using successive 4-s time windows spanning the duration of the attention trial (Fig. 3E). Interestingly, the relationship between top-down activation of target-representations and detection accuracy was most robust during the latter half of the attention trial. This profile suggests that individual differences in performance did not depend on differences in establishing a task-relevant attentional bias within visual cortex, but rather, behavioral differences were associated with differential maintenance of attentional biases within aLOC.

Discussion

The results of this experiment provide important evidence for a preparatory coding model for attentional bias in the human brain. We confirm five key predictions that follow from a general model of preparatory coding that is situated within the representational structure of perceptual cortex. Firstly, population selectivity: different neural subpopulations were selectively activated within area LOC during attentional trials. Secondly,

population specificity: cross-comparison confirmed that preparatory attention specifically activated the same neural subpopulations that were active during corresponding visual stimulation. Thirdly, we show that both selectivity and specificity were under the control of flexible top-down mechanisms by using arbitrarily assigned auditory cues to direct attention. Attentional bias, mediated via target-specific activity within LOC, was not dependent on any visual stimulation; therefore, we infer that flexible top-down attentional mechanisms are sufficient to generate shape-specific activity patterns in visual cortex. Fourthly, we also demonstrate that preparatory activity is sustained throughout the duration of attention trial. Finally, we provide crucial evidence that confirms the functional role of this sustained preparatory activity. The positive relationship between the target-specific pattern activity in aLOC and subsequent target detection demonstrates that effective maintenance of preparatory activity facilitates target processing.

Previously, neurophysiological recordings in non-human primates provided the most direct evidence for a preparatory coding model for selective attention (5, 6, 21). For example, Chelazzi et al. (5, 6) recorded from single neurons within area IT during the delay period of a cued visual search task. Neurons that responded optimally to a particular cue stimulus maintained an above-baseline firing rate until the onset of the target array. Similar modulations have also been observed in response to spatial cueing within retinotopically organized visual areas (21). In the current experiment, we asked whether similar baseline shifts are expressed at the level of population coding in human visual cortex during non-spatial selective attention. Exploiting cross-comparison MVPA to characterize population coding of attentional bias with a level of selectivity that functionally approximates the resolution of single-unit neurophysiology (15, 16), we observed evidence for preparatory activation of distinct neural subpopulations in response to specific attentional cues. In our task, preparatory activity was specific to high-level visual areas involved in shape processing (20). This functional speci-

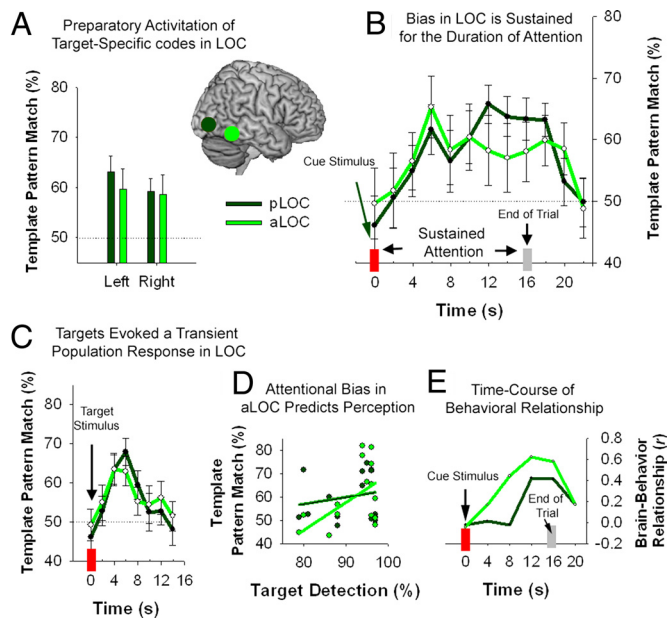


Fig. 3. Attentional bias in anterior and posterior lateral occipital complex (aLOC/pLOC). (A) Region-of-interest analyses confirmed attentional activation of shape-specific neural patterns in bilateral pLOC and aLOC. (B) Time-course analyses revealed that attentional modulation of target-specific patterns following the onset of the cue stimulus was sustained throughout the duration of the trial. In contrast with the sustained attentional modulation, (C) presentation of target stimuli resulted in a transient activation of target-specific neural populations. (D) The accuracy of the pattern match between template-specific bias observed in the attention task and stimulus-driven perception defined by the pattern-localizer within aLOC was positively correlated with detection accuracy for subsequently presented target stimuli ($r = 0.59$, $P = 0.016$), but not within pLOC ($P = 0.434$). (E) Finally, time-course analysis of the correlation between visual template activation and target detection revealed a stronger relationship during the latter portion of the attention trial. Error bars, ± 1 SEM.

ficacy is consistent with monkey neurophysiological evidence for shape-specific preparatory activity in IT (5, 6), but not V4 (22). Complementary evidence also suggests that attention to lower-level features such as orientation produces preparatory effects at an earlier stage of visual processing (23). These task-dependencies are consistent with the hypothesis that task context determines the primary location of preparatory bias (24). In our task, participants were instructed to respond to the cued target letter, and ignore all other non-target letters. Biasing neural subpopulations within LOC that represent the target-specific shape would optimize visual processing. If, however, attention was manipulated according to lower-level units of selection (e.g., location, orientation, color), top-down modulation would need to operate at earlier processing stages to bias stimulus-driven competition.

These results critically extend upon early neurophysiological evidence for preparatory baseline shifts (5, 6, 21), as well as more recent fMRI studies of non-spatial preparatory attention that demonstrate activation biases between feature-specific processing areas (11–13, 25). Specifically, we demonstrate that top-down attentional mechanisms can selectively activate target-specific representations coded within the same subregion of visual cortex. Moreover, adapting our experimental protocol for humans allowed us to explore attentional bias with greater flexibility than is typically afforded by non-human primate studies. In particular, using an arbitrarily assigned auditory cue to direct attention, we could demonstrate that top-down mechanisms generate, as well as maintain, target-specific biases in visual cortex. Most importantly, our results provide additional evi-

dence that top-down modulations of specific visual representations reflect the neural implementation of attentional bias. Firstly, we satisfy a necessary precondition for an attentional interpretation of the top-down activity: target-specific activation patterns were sustained throughout the duration of the attention trial. Secondly, we also found that individual participants who generated and sustained neural patterns that more closely resembled visually evoked patterns for the cued target stimulus could detect the faint target stimuli with greater accuracy than participants who showed weaker evidence for attentional bias. Together, baseline activity that scales with the duration and efficacy of selective attention provides compelling evidence that the neural response is causally related to attention (9, 10). An intriguing result is that the behavioral relationship was reliable for aLOC, but not pLOC. Although preparatory bias was evident in both regions, the more anterior region might be more closely linked to behavior.

Recent pattern analytic fMRI studies of attention have examined modulation of the population response during direct (26, 36), and indirect (27), sensory stimulation. These studies demonstrate that the population response to specific directions of motion is modulated by selective attention (26, 27), even when the driving stimulus was presented outside the measured region's receptive field (27). Similar results have also been observed using more conventional univariate analysis of fMRI (28). However, no previous fMRI experiment has demonstrated attentional activation of distinct population codes in the complete absence of visual stimulation. Recent MVPA studies of visual imagery (19) and working memory (17, 18) provide the closest evidence for top-down modulations of visual cortex that resemble our findings. Here, we asked how top-down input is used in attentional control, exploring the locus, time-course, and behavioral consequences of preparatory activity.

Overall, these results provide strong evidence that preparatory coding for attentional bias exploits the preexisting representational architecture of perceptual cortex. Perceptual representations, coded across distributed cell assemblies, are shaped throughout the history of perceptual experience, and form the storehouse of perceptual memories (29). Stimulus-driven activation of stored visual representations provides the basis of ongoing perception; however, according to a preparatory coding model of selective attention, top-down mechanisms can also selectively activate specific perceptual representations. Top-down signals coding high-level information (e.g., task goals) strategically prime behaviorally relevant representations stored amongst other perceptual representations that are currently less relevant to behavior. Similar coding schemes have also been proposed to account for more general mechanisms of perception. According to predictive coding models, for example, prior expectation, which may be expressed formally in terms of Bayesian priors, modulates specific representations stored in perceptual cortex (30). Within this framework, ongoing perception is biased by statistical regularities in the environment, which can be expressed rapidly according to higher-level information, such as foreknowledge that stimulus A always precedes stimulus B within a particular context (30).

Although numerous influential models of attention and expectation assume that top-down modulation of baseline neural activity biases ongoing perception (1, 4, 30, 31), the precise mechanisms of top-down perceptual activation in the absence of sensory input still remain poorly understood. In particular, perceptual representations are coded across distributed neural populations (32); therefore, competition between representations stored within the same cortical area can only be resolved by selectively biasing spatially intermingled neural subpopulations. Selective preactivation of behaviorally relevant neural codes that spatially overlap with competing representations presents a significant challenge for mechanisms of attention if

they are to modulate perceptual bias from a central, anatomically distant, control centre such as a prefrontal, and/or parietal cortex. In this study, we demonstrate that the brain meets this challenge during flexibly controlled selective attention, but future research is needed to explore the precise mechanisms of top-down control over perceptual bias. An important goal for future research will be to discover the neurophysiological mechanisms that link flexible, and abstract, goal representations in frontoparietal cortex to top-down modulation of baseline activation of stable, but distributed and overlapping, representations in perceptual cortex.

Materials and Methods

Participants and Behavioral Task. Sixteen right-handed volunteers (nine female; mean age 25 years and 6 months, range 18 years and 5 months to 35 years and 3 months) participated in this experiment. All participants had normal, or corrected-to-normal, vision and no history of neurological or psychiatric illness. All participants were screened for MR contraindications and gave written informed consent before scanning. Participants received a small honorarium to reimburse them for taking part and to reward accurate performance. All experimental protocols were approved by the Hertfordshire Local Research Ethics Committee.

The experiment was conducted over six separate scanning runs, and consisted of a pattern-localizer task in addition to the main attention task. Each scanning run contained three epochs of the pattern-localizer task, and three epochs of the attention task. The order of presentation was randomized across scans and participants, and each epoch was preceded by a verbal cue informing participants of which task they were about to perform. Throughout all scanning, participants were instructed to refrain from head movements and eye movements.

All visual stimuli were viewed via a back-projection display (1024 × 768 resolution, 60 Hz refresh rate), with a black background. Visual Basic (Microsoft Windows XP; Dell Latitude 100L Pentium 4 Intel 1.6 GHz) was used for all aspects of experimental control, including stimulus presentation, recording behavioral responses (via an MR-compatible button-box) and synchronizing experimental timing with scanner pulse timing.

Pattern Localizer. The localizer task was used to characterize patterns of brain activity associated with visual presentation of the target stimuli used in the attention task: the letters “X” and “O.” All letters were presented at the centre of the visual display in black 72-point Arial font ($\approx 1.5^\circ$ of visual angle) within a circular aperture containing a dynamic visual noise pattern. The aperture was 200 pixels in diameter ($\approx 3.25^\circ$) and the noise pattern was generated by assigning a random greyscale value to each pixel within the aperture. Pixel values covered the full range of greyscale values from white to black with a uniform frequency distribution. Each pixel was randomly assigned a new value every 33.3 ms to give the impression of an animated noise pattern.

Each block of the localizer task consisted of 8 repetitions of a single letter, either X or O. Each letter was presented for 1 s and followed by a 1-s interstimulus interval. A white fixation cross (0.33°) was also visible throughout each block. All visual stimuli were clearly visible over the noise display, and participants were instructed to monitor the letter stimuli for occasional size deviations. The size deviant stimuli were slightly smaller than the standard letters (60-pt Arial font, $\approx 1.25^\circ$). In total, one-third of blocks contained no deviants, one-third contained one deviant, and the final one-third contained two deviants. The position in which the deviants occurred was randomized across blocks. At the end of each block, a response screen was presented and participants were given 4 s to indicate how many deviant stimuli they had detected. This was followed by a 12-s period of fixation after which the next block began. Each epoch of the localizer task contained a single block of each letter, yielding a total of three blocks per letter in each scanning run, and a total of 18 blocks per letter over the course of the whole experiment.

Attention Task. Each trial of the attention task began with a cue informing participants to attend for either the letter “X” or the letter “O.” Cues were auditory tones (220 Hz or 1,100 Hz) presented for 100 ms, and the association between cues and target stimuli was randomized across participants (i.e., 220 Hz → X, and 1,100 Hz → O; or 220 Hz → O, and 1,100 Hz → X). Halfway through the experiment, the cue-target mapping was also reversed for each participant to ensure that cue-related effects were specific to the contents of attention rather than any physical differences between cues. The task cue at the beginning of each epoch reminded participants of the current mapping.

A visual noise pattern identical to that used in the pattern-localizer task was also presented with the onset of the auditory cue. This noise pattern was displayed for the entire 16-s duration of each trial, during which participants were instructed to monitor the noise pattern for the occasional presentation of the current target letter (i.e., the cued letter, X or O). A white fixation cross (0.33° of visual angle) was also visible throughout each trial to help participants maintain central fixation. Trial types were divided into target-present (25% of all trials), non-target (25%), and no-letter (50%). During target-present trials, either one or two letters were presented, with at least one presentation being the currently cued target letter. Attentional cues were never invalid, therefore successful allocation of selective attention was never penalized. Non-target trials also contained up to two letter presentations, but these were randomly selected letters, excluding both of the possible target letters.

During both target-present and non-target trials, letter presentations began at any time between 0.5 and 14.5 s after the auditory cue. As in the pattern-localizer task, letter stimuli were presented at central fixation in black 72-pt Arial font. To increase the difficulty of target detection, and to encourage attentional allocation, the letter stimuli were rendered semitransparent. Each letter was initially presented as 100% transparent (i.e., invisible), and the degree of transparency was gradually decreased over the course of a 1-s presentation. The final transparency of each letter stimulus was randomly allocated a value between 90.2 and 86.3%, corresponding to a contrast of approximately 19–26% between the mean luminance of pixels within the letter stimulus and the mean luminance of the dynamic white noise.

At the end of each trial, a response screen was presented, and participants were given 4 s to indicate how many target stimuli they had detected. This was followed by a 12-s period of fixation after which the next trial began. Each epoch of the attention task contained two trials for each cue type, including one target-present and one target-absent (non-target or no letter) trial. The order in which trials were presented was randomized in each epoch.

fMRI Data Acquisition. All brain imaging was performed using a Siemens 3T Tim Trio scanner with a 12-channel head coil. Functional volumes (T2*-weighted echo planar images) consisted of 16 slices (64×64 voxels per 3-mm slice; $3 \times 3 \times 3$ mm voxel resolution with a 25% gap between slices), acquired in descending order [TR = 1 s; TE = 30 ms; flip angle = 78°]. The near-axial acquisition matrix was positioned to capture occipital and temporal cortex while avoiding the ocular orbits. To avoid T1 equilibrium effects, the first eight volumes were discarded from all functional analyses. Finally, high-resolution ($1 \times 1 \times 1$ mm) structural images were acquired for each participant for standard co-registration and spatial normalization procedures.

Univariate Data Analysis. Initially, data were analyzed using a conventional univariate approach in SPM 5 (Wellcome Department of Cognitive Neurology, London, U.K.). The purpose of these analyses was to verify that the conditions of interest each evoked equivalent levels of activity. Specifically, we tested whether X or O letter stimuli differentially activated visual cortex during the pattern localizer. Similarly, during the attention task, we applied univariate analyses to test whether attending for “X” or “O” elicited differential mean responses in brain activity. In the absence of mean differences in regional activity between conditions, significant differentiation based on MVPA implies subtle, yet reliable, differences in the pattern of neural activity.

For all univariate analyses, functional images were passed through the following preprocessing steps: realignment, unwarping, slice-time correction, spatial normalization (MNI template), smoothing (Gaussian kernel: 8-mm full-width half-maximum) and high-pass filtering (cut-off = 128 s). For the localizer task, data from individual participants were analyzed using separate fixed effects analyses. Each of the six scanning runs were modeled using two explanatory variables: viewing X or O. These were derived by convolving the onsets for each letter presented during the localizer task with the canonical hemodynamic response function. First-level analyses of the attention task were performed using a finite impulse response (FIR) basis set to model the time-course of neural activity following the presentation of an attention cue for “X” or “O.” Specifically, we modeled the time-course of activity from the onset of each cue type until 12 s after the end of the attention trial. This period was modeled using 15 separate time bins, each with a duration of 2 s. To ensure that the attention-related activity was uncontaminated by letter-specific visual input, time-bins occurring after the onset of either target or non-target stimuli were excluded from the estimate of cue-related activation patterns. The response to target stimuli was examined separately, using an FIR model time-locked to the onset of target stimuli during target-present trials. For all three design matrices, constant terms were included to account for session effects. Serial autocorrelations were estimated using an AR (1) model

with prewhitening, and second-level analyses were based on beta parameter estimates for each condition, averaged across scanning runs.

Multivoxel Pattern Analysis. The full details of the MVPA procedure used in this paper are presented in *SI Text*. All pattern analyses were performed on minimally preprocessed data. Functional images were spatially aligned, unwarped and slice-time corrected, but not spatially normalized or smoothed. The time-series data from each voxel were high-pass filtered (cut-off = 128 s) and data from each session were scaled to have a grand mean value of 100 across all voxels and volumes. Using a searchlight procedure (33), neural classifications were performed at each voxel location, based on the pattern of activity observed within the surrounding cortical volume (a radius = 10-mm sphere, containing ≈ 90 voxels). Each classification was performed using a correlation-based approach (19, 34, 35) (see *SI Text*). Data from the localizer task were analyzed using a leave-one-out cross-validation procedure. For each iteration, data were divided into test and train sets, and discriminative patterns for X vs. O were derived for each set by subtracting the activation pattern for X from the activation pattern for O across all voxels within the current searchlight focus. A voxel-wise correlation was then calculated to assess the similarity between train and test patterns. Correlation coefficients above zero were coded as correct classifications, whereas coefficients of zero or below were recorded as incorrect classifications. The overall percent of correct classifications was calculated over six train-test permutations. We used a cross-comparison variant of the correlation approach to assess the attention data. First, we trained classifiers within each searchlight sphere to discriminate between stimuli presented during the localizer task, and then we tested for a pattern match using data from the attention task. We used the FIR model described above to characterize the degree of pattern match over 12 separate

time points following the attention cue. For comparison, we also examined the stimulus-driven pattern response using data from 0–14 s after the onset of target stimuli.

For all searchlight analyses, classification accuracy within each sphere was recorded at the central voxel. Repeating this procedure across all voxels produced a 3-D accuracy map. Accuracy maps for each pattern analysis were then spatially normalized to the MNI template, and assessed via a random-effects group analysis. We also examined our results within predefined regions of visual cortex associated with visual processing of shapes (20). Mapping category-specific subregions of extrastriate cortex, Spiridon et al. (2006) identified separate coordinates for anterior and posterior lateral occipital complex (LOC). Regions-of-interest (ROIs) were defined by a sphere ($r = 10$ mm) centered the peak activation co-ordinates for each of these regions (averaged across hemispheres; aLOC: $\pm 37, -51, -14$; pLOC: $\pm 45, -82, -2$). For the attention data, an overall summary value was also calculated for each subject by averaging the classification data across 8 time bins (i.e., 16 s), beginning 4 s after the cue onset. This time-averaged summary value was used to assess the results of the searchlight analysis (Fig. 2C) and initial ROI analyses (Fig. 3A), and to examine the relationship between behavior and attentional bias in LOC (Fig. 3D). The time-course of classification accuracy was also down-sampled by averaging together classification accuracy scores obtained for the two time bins within each 4-s epoch. This more stable estimate was used to examine the time-course of the brain-behavior relationship (Fig. 3E).

ACKNOWLEDGMENTS. We thank C. Summerfield and M. Peelen for valuable comments and suggestions. This work was supported by Medical Research Council (UK) Intramural Program Grant U.1055.01.001.00001.01 (to J.D.), and a Junior Research Fellowship, St. John's College, Oxford (to M.S.).

- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222.
- Driver J, Frith C (2000) Shifting baselines in attention research. *Nat Rev Neurosci* 1:147–148.
- Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the Human cortex. *Annual Review of Neuroscience* 23:315–341.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Chelazzi L, Miller EK, Duncan J, Desimone R (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363:345–347.
- Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol* 80:2918–2940.
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annual Review of Neuroscience* 27:611–647.
- Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22:751–761.
- Ress D, Backus BT, Heeger DJ (2000) Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci* 3:940–945.
- Silver MA, Ress D, Heeger DJ (2007) Neural correlates of sustained spatial attention in human early visual cortex. *J Neurophysiol* 97:229–237.
- Fannon SP, Saron CD, Mangun GR (2007) Baseline shifts do not predict attentional modulation of target processing during feature-based visual attention. *Front Hum Neurosci* 1:7.
- Chawla D, Rees G, Friston KJ (1999) The physiological basis of attentional modulation in extrastriate visual areas. *Nat Neurosci* 2:671–676.
- Giesbrecht B, Weissman DH, Woldorff MG, Mangun GR (2006) Pre-target activity in visual cortex predicts behavioral performance on spatial and feature attention tasks. *Brain Res* 1080:63–72.
- Shibata K, et al. (2008) The effects of feature attention on prestimulus cortical activity in the human visual system. *Cerebral Cortex* 18:1664–1675.
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.
- Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7:523–534.
- Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci* 20:207–214.
- Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature*.
- Stokes M, Thompson R, Cusack R, Duncan J (2009) Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J Neurosci* 29:1565–1572.
- Spiridon M, Fischl B, Kanwisher N (2006) Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum Brain Mapp* 27:77–89.
- Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* 77:24–42.
- Chelazzi L, Miller EK, Duncan J, Desimone R (2001) Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex* 11:761–772.
- Haenny PE, Schiller PH (1988) State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. *Exp Brain Res* 69:225–244.
- Duncan J (1998) Converging levels of analysis in the cognitive neuroscience of visual attention. *Philos Trans R Soc Lond B Biol Sci* 353:1307–1317.
- McMains SA, Fehd HM, Emmanouil T-A, Kastner S (2007) Mechanisms of feature- and space-based attention: Response modulation and baseline increases. *J Neurophysiol* 98:2110–2121.
- Kamitani Y, Tong F (2006) Decoding seen and attended motion directions from activity in the human visual cortex. *Curr Biol* 16:1096–1102.
- Serences JT, Boynton GM (2007) Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55:301–312.
- Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. *Nat Neurosci* 5:631–632.
- Fuster JM (2001) The prefrontal cortex - An update: Time is of the essence. *Neuron* 30:319–333.
- Summerfield C, et al. (2006) Predictive codes for forthcoming perception in the frontal cortex. *Science* 314:1311–1314.
- Ungerleider LG, Courtney SM, Haxby JV (1998) A neural system for human visual working memory. *Proc Natl Acad Sci USA* 95:883–890.
- Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. *Nat Rev Neurosci* 1:125–132.
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103:3863–3868.
- Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Williams MA, Dang S, Kanwisher NG (2007) Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci* 10:685–686.
- Peelen M, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460:94–97.