

# Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex

Yaara Erez<sup>1</sup> and John Duncan<sup>1,2</sup>

<sup>1</sup>Medical Research Council Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom, and <sup>2</sup>Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom

Allocating attentional resources to currently relevant information in a dynamically changing environment is critical to goal-directed behavior. Previous studies in nonhuman primates (NHPs) have demonstrated modulation of neural representations of stimuli, in particular visual categorizations, by behavioral significance in the lateral prefrontal cortex. In the human brain, a network of frontal and parietal regions, the “multiple demand” (MD) system, is involved in cognitive and attentional control. To test for the effect of behavioral significance on categorical discrimination in the MD system in humans, we adapted a previously used task in the NHP and used multivoxel pattern analysis for fMRI data. In a cued-detection categorization task, participants detected whether an image from one of two target visual categories was present in a display. Our results revealed that categorical discrimination is modulated by behavioral relevance, as measured by the distributed pattern of response across the MD network. Distinctions between categories with different behavioral status (e.g., a target and a nontarget) were significantly discriminated. Category distinctions that were not behaviorally relevant (e.g., between two targets) were not discriminated. Other aspects of the task that were orthogonal to the behavioral decision did not modulate categorical discrimination. In a high visual region, the lateral occipital complex, modulation by behavioral relevance was evident in its posterior subregion but not in the anterior subregion. The results are consistent with the view of the MD system as involved in top-down attentional and cognitive control by selective coding of task-relevant discriminations.

**Key words:** categorization; fMRI; frontoparietal network; goal-directed behavior; multivoxel pattern analysis (MVPA)

## Significance Statement

Control of cognitive demands fundamentally involves flexible allocation of attentional resources depending on a current behavioral context. Essential to such a mechanism is the ability to select currently relevant information and at the same time filter out information that is irrelevant. In an fMRI study, we measured distributed patterns of activity for objects from different visual categories while manipulating the behavioral relevance of the categorical distinctions. In a network of frontal and parietal cortical regions, the multiple-demand (MD) network, patterns reflected category distinctions that were relevant to behavior. Patterns could not be used to make task-irrelevant category distinctions. These findings demonstrate the ability of the MD network to implement complex goal-directed behavior by focused attention.

## Introduction

Essential to goal-directed behavior is the ability to allocate attentional resources to behaviorally relevant aspects of sensory input,

such as categorical membership, and to filter out irrelevant information. Evidence from nonhuman primates (NHPs) has demonstrated that neuronal activity in the lateral prefrontal cortex (LPFC), and more recently in the parietal cortex, is related to attentional control, in particular task-related visual categorization. It has been shown that neurons respond to behaviorally meaningful visual categorizations, but not to task-irrelevant distinctions (Freedman et al., 2001; Everling et al., 2002, 2006; Swaminathan and Freedman, 2012). Many other studies provide evidence for selective prefrontal activity based on behavioral relevance (Watanabe, 1986; Sakagami and Niki, 1994; Miller et al., 1996; Asaad et al., 2000; Miller

Received March 24, 2015; revised June 24, 2015; accepted July 22, 2015.

Author contributions: Y.E. and J.D. designed research; Y.E. performed research; Y.E. contributed unpublished reagents/analytic tools; Y.E. analyzed data; Y.E. and J.D. wrote the paper.

This work was supported by the Medical Research Council (United Kingdom) intramural programme (Grant MC\_A060-5PQ10). Y.E. was supported by an European Molecular Biology Organization (EMBO) Long-Term Fellowship cofunded by Marie-Curie Actions (European Commission, EMBOCOFUND2010, GA-2010-267146) and a Royal Society Dorothy Hodgkin Research Fellowship (United Kingdom). We thank Daniel Mitchell for useful comments and advice throughout this study; Benedikt Poser and the Donders Institute for sharing the 3D EPI sequence; and Andre van der Kouwe from the Massachusetts General Hospital—Health Sciences and Technology (MGH/HST) Athinoula A. Martinos Center for Biomedical Imaging for sharing the multiecho MPRAGE sequence.

The authors declare no competing financial interests.

Correspondence should be addressed to Yaara Erez, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK. E-mail: yaara.erez@mrc-cbu.cam.ac.uk.

DOI:10.1523/JNEUROSCI.1134-15.2015

Copyright © 2015 the authors 0270-6474/15/3512383-11\$15.00/0

and Cohen, 2001; Cromer et al., 2010; Kusunoki et al., 2010; Kadohisa et al., 2013).

Although single-cell studies from the NHP provide only partial brain coverage, accumulating findings suggest that comparable regions in the human brain may be involved in similar attentional and categorization processes. It is widely accepted that a network of frontal and parietal cortical regions in the human brain, the “multiple demand” (MD) network, is involved in cognitive and attentional control (Norman and Shallice, 1980; Desimone and Duncan, 1995; Miller and Cohen, 2001; Duncan, 2013). Regions within this network include the intraparietal sulcus (IPS), the anterior–posterior axis of the inferior frontal sulcus, anterior insula (AI), and adjacent frontal operculum, frontal eye field (FEF), and presupplementary motor area (pre-SMA) with adjacent anterior cingulate cortex. Neuroimaging studies in humans have demonstrated the involvement of the MD network in cognitive control across a wide range of tasks, including spatial visual attention, working memory, language, and others (Woolgar et al., 2011b; Duncan, 2013; Fedorenko et al., 2013). Damage to the MD system may be an important contributor to the broad impairments in cognitive control that follow prefrontal lesions (Milner, 1963; Luria, 1966; Roca et al., 2010; Woolgar et al., 2010).

Here, we sought to build on the findings from the NHP and track categorical discrimination of visual objects across the MD network in the human brain using fMRI. We predicted that neuronal activity across the MD network would reflect categorical discrimination and that this discrimination would depend on whether it is behaviorally relevant. We adapted a cued-detection task that was used previously in the NHP (Kadohisa et al., 2013) with a categorization component added into it. In the task, a cue indicated two visual categories as targets. In a subsequent visual display, participants detected whether an object from one of these two categories appeared on the screen. Importantly, categories could serve as either targets on some trials or nontargets on other trials depending on the cue, with two additional categories always serving as nontargets. Using multivoxel pattern analysis (MVPA), we tested for the modulation of categorical discrimination by behavioral relevance across the MD network and in a high-level visual region, the lateral occipital complex (LOC).

## Materials and Methods

### Participants

Nineteen healthy volunteers (12 females, mean  $\pm$  SD age, 25.8  $\pm$  5.6 years) participated in the experiment. All participants gave written informed consent and were reimbursed for their time. All participants were right-handed, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric illness. The study was approved by the Cambridge Psychology Research Ethics Committee. One participant was excluded from the analysis because of frequent pauses for rest between trials.

### Experimental paradigm

**Task.** Participants were scanned while performing a cued category-detection task (Fig. 1A). The task was organized in miniblocks of four trials. At the start of the miniblock, a cue that included names of two visual target categories was displayed on the screen for 2 s, followed by a blank period of 1 s and the sequence of 4 trials. In each trial, a visual display was presented for 150 ms and the participant decided whether a stimulus from either of the two target categories was present. The response (target category present or absent) was made by pressing a button with the right or left index finger (see below). Participants were instructed to respond as fast and as accurately as possible. After the participant's response, there was a 1 s delay before stimulus onset for the next trial. After the 4 trials were complete, there was a further delay of 1 s before onset of the cue for the next miniblock.

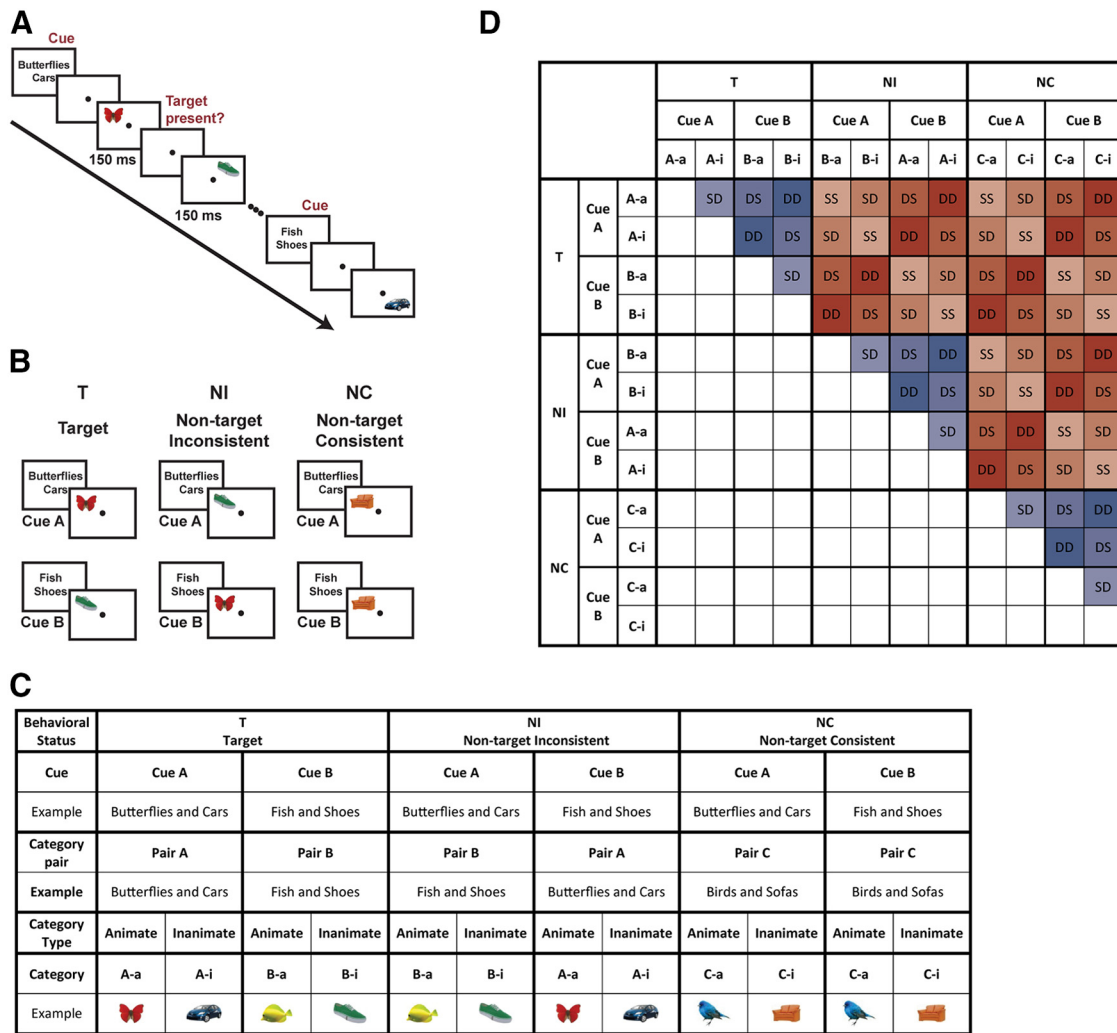
**Visual categories.** Six visual categories were used and included two category types: three animate categories (butterflies, fish, birds) and three inanimate categories (cars, shoes, sofas). For each participant, the categories were paired throughout the experiment and each pair included one animate and one inanimate category. Two of the three pairs of categories (pair A, pair B) were assigned a cue (cue A and B, respectively). Therefore, they could serve as either targets (T) or nontargets (NIs, nontargets that are inconsistent, i.e., may be targets on other trials), depending on the cue. The remaining pair of categories (pair C) was not associated with a cue; therefore, stimuli from these categories were always nontargets (NCs, nontargets that are consistent; Fig. 1B,C). This design resulted in three levels of behavioral status of the categories (T, NI, and NC). The combinations of category pairs and their association with the cues were balanced across participants. Overall, we had 18 possible combinations of categories, one combination for each of the 18 participants that were included in the analysis. We predicted that, for any two categories, the discrimination between them as assessed by MVPA across the MD network would be based on the behavioral relevance of this discrimination. Patterns of neural activity for two categories with different behavioral status (e.g., T and NI, T and NC) would be well discriminated. In contrast, patterns of neural activity for two categories with the same behavioral status (e.g., T and T, NI and NI) would not be well discriminated.

**Stimuli and visual display.** Stimuli were colorful images drawn from the 6 visual categories with a similar size and an average of  $4.7^\circ \times 3^\circ$ . Eight exemplars were used from each category, with similar view for all exemplars within the same category. The visual display included a white background and a black fixation dot in the center of the screen. A stimulus could appear in each of the four quadrants of the display: top-left, top-right, bottom-left, and bottom-right, counterbalanced throughout the experiment for the different conditions. Stimuli in each quadrant were centered at  $2.8^\circ$  from the fixation dot on both the  $x$  and  $y$  axes. To avoid any effects on neural representation driven by differences in visual exposure between the task conditions as a result of a saccade, stimuli were presented very briefly (150 ms only). In addition, the fixation dot appeared on the screen while the stimuli were displayed and participants were instructed to fixate. The display contained either one or four stimuli, but here we focus on the single-stimulus displays only.

**Task structure and design.** The task included four runs, with two blocks in each run. The task design was event related, and trials were organized in blocks and miniblocks. Each block comprised 33 miniblocks of four trials each, with one cue used for each miniblock. A cue was presented at the beginning of each miniblock, indicating the target categories for the four subsequent trials. The two cues (cue A, cue B) were used throughout the experiment alternately, such that each of the two cues appeared every second miniblock. A 24 s fixation dot display appeared after each block to use as the baseline and to keep the blocks as independent as possible in terms of BOLD signal.

Responses were made by pressing either the right or left button on a button box using the right and left index fingers for “target present” and “target absent” responses. The stimulus–response mapping between target present/absent and button (left/right) was changed every block and was balanced across runs. To eliminate influence of change in stimulus–response mapping on performance and neural discrimination measures, the first miniblock (four trials) in each block served as a dummy and was excluded from the analysis.

Overall, 32 miniblocks in each block were included in the analysis for a total number of 128 trials per block. Each condition appeared an equal number of times after each of the two cues within a block. Half of the trials in a block (64) included single-stimulus displays, of which half (32) were target trials (T) and half were nontarget trials, with 16 NI trials and 16 NC trials. The 32 T trials included two repetitions of each combination of cue (Cue A, Cue B), category type (Animate, Inanimate) and location (four visual quadrants). The 16 NI trials and the 16 NC trials each included one repetition of each combination. The average EPI time for each run of the task was  $12.6 \pm 0.9$  min (mean  $\pm$  SD). The task was written and presented using Psychtoolbox3 for MATLAB (The MathWorks; Brainard, 1997).



**Figure 1.** Experimental paradigm and classification approach. **A**, A cued category-detection experimental paradigm was used in which a cue (names of two target categories) is followed by a series of visual displays. In each trial, participants detected whether one of the target categories appeared or not and pressed a button accordingly. **B**, An example of visual categories and their behavioral status under different task contexts. Exemplars from visual categories could be either targets (T) or nontargets, depending on the cue. Importantly, nontargets could be either inconsistent nontargets (NI) that may be targets on other trials or consistent nontargets (NC) that are never targets, therefore creating three levels of behavioral status for the presented objects. **C**, Diagrammatic scheme of the full condition space, including example category assignments for a single participant. Each cue (Cue A, Cue B) indicated the pair of categories currently serving as targets. Pairs of categories included one animate and one inanimate category. For each participant, the six visual categories were divided into three pairs (A, B, and C). Behavioral status (T, NI, NC) was determined by the combination of the cue and presented category in each trial. **D**, Classification matrix. To assess the modulation of categorical discrimination by behavioral relevance, we used correlation-based classification (see Materials and Methods for details). Classification accuracies were computed between all possible pairs of categories and averaged across entries of the classification matrix as appropriate for each contrast. Distinction between category pairs with different behavioral status was termed behaviorally relevant (red entries), whereas distinction between category pairs with the same behavioral status was termed behaviorally irrelevant (blue entries). To assess the modulation of categorical discrimination by other aspects of the task, which are orthogonal to the behavioral decision, we averaged classification accuracies across entries depending on whether pairs of categories shared the same or different cue (Cue A, Cue B), as well as same or different category type (Animate, Inanimate), as indicated by acronyms and color coding of entries (hues of red and blue). SS, Same cue and Same category type; SD, Same cue and Different category type; DS, Different cue and Same category type; DD, Different cue and Different category type.

**Functional localizers**

We used two functional localizers to help with subject-specific feature selection for MVPA in selected regions of interest (ROIs). One localizer was used for the MD network regions and the other for LOC, with two runs for each localizer. ROIs were defined using templates (see ROIs, below) and, within each ROI template, subject-specific feature selection for the MVPA was done using the functional localizer data.

**MD network functional localizer.** We used a spatial working memory task to localize the MD network (Fedorenko et al., 2013). Each trial started with a 500 ms fixation display, followed by a 3 × 4 grid. Participants had to keep track of either four or eight locations that were consecutively highlighted on the grid for the Easy or Hard conditions, respectively. Then, in a two-alternative forced-choice question presented for 3 s, participants had to choose the grid with the correct highlighted locations by pressing either a left or right button. An additional 500 ms

fixation display followed each trial. Locations were highlighted on the grid for 1 s each. For the Hard condition with the eight locations, locations were highlighted on the grid in pairs. Hard and Easy conditions were blocked and balanced with four trials in each block. There were also equal numbers of left and right responses. Each participant completed two runs, with each run including six Easy blocks and six Hard blocks of 32 s each, as well as four fixation blocks of 16 s each. The total duration of each run was 7.5 min. The task was written and presented using Python (Python Software Foundation).

**LOC functional localizer.** We used a standard functional localizer that included two object categories (Objects and Scrambled-objects) presented in a blocked design. Each block lasted 12 s and included 15 stimuli, each presented for 300 ms with a 500 ms interstimulus interval. Category block order was counterbalanced within and across runs. Each run consisted of an initial 6 s fixation dummy scans, 8 blocks for each category



and 5 blocks of a baseline fixation dot (a total of 162 s). To maintain vigilance, participants were asked to perform a one-back task on the presented stimuli by pressing a button on a button box. Each participant completed two visual localizer runs. The task was written and presented using Psychtoolbox3 for MATLAB (Brainard, 1997).

Overall, the scanning session comprised a short training outside the scanner followed by testing in the scanner including four runs of the main task, two runs of the LOC localizer, and two runs of the MD network localizer. Stimuli for the main task and localizers were projected on an MRI-compatible screen inside of the scanner.

#### Data acquisition

fMRI data were acquired using a Siemens 3T TimTrio scanner with a 32-channel head coil. We used a T2\*-weighted 3D EPI fast imaging acquisition depicting the BOLD contrast (Poser et al., 2010). We used an in-plane acceleration factor (AF) of 2 and an additional factor of 2 for the slice encoding direction, resulting in a total AF of 4. Other acquisition parameters were as follows: acquisition time of 1288 ms, echo time of 30 ms, 52 slices with a slice thickness of 2.5 mm, no interslice gap, inplane resolution of  $2.5 \times 2.5$  mm, field of view of 205 mm, and flip angle of  $14^\circ$ . T1-weighted multiecho MPRAGE (van der Kouwe et al., 2008) high-resolution images were acquired for all participants (slice size 1 mm isotropic, field of view of  $256 \times 256 \times 192$  mm, TR of 2530 ms, TEs of 1.64, 3.5, 5.36, and 7.22 ms). The voxelwise root mean square across the four MPRAGE images was computed to obtain a single structural image.

#### Data analysis

The main analysis approach was MVPA to test for categorical discrimination based on behavioral relevance. This was preceded by an initial univariate analysis that was used to verify the recruitment of the MD network during task performance. Preprocessing and generalized linear modeling (GLM) were conducted using SPM8 (Wellcome Trust Centre for Neuroimaging, University College London, London). MVPA was conducted using custom MATLAB code.

**Preprocessing.** Functional images for the three tasks (main task, MD network localizer, and LOC localizer) were processed separately. For each task, functional images were realigned to the first image and resliced. Structural and functional images were normalized to a Montreal Neurological Institute (MNI) template using the normalization matrix obtained for the structural image. Slice timing correction was not required because a whole volume was acquired simultaneously.

**GLM for the main task.** A GLM was estimated for each participant for correct trials only. Separate regressors were created for each combination of behavioral status (T, NI, NC), cue (cue A, cue B), category type (Animate, Inanimate), and block within a run, ignoring location (visual quadrant). Trials were modeled as epochs lasting from stimulus onset to response (Woolgar et al., 2014) convolved with a canonical hemodynamic response function (HRF). Additional regressors were created for fixation blocks, block instructions (including the stimulus–response mapping instructions for the block), cue displays, and error trials. Movement parameters and run means were included as covariates of no interest. In the experimental paradigm, the number of T trials was double that of NI and NC conditions, to produce a 50% chance of target presence. To have an equal number of trials used to model each condition, the T trials were split into two halves while maintaining counterbalancing across all relevant parameters (cue, category type, location, etc.) and each half was modeled with separate regressors in the GLM. Spatially unsmoothed data were used.

**GLM for the MD network localizer.** A GLM was estimated for each participant, with regressors created by convolving a canonical HRF with blocks of Hard and Easy conditions. Movement parameters and run means were included as covariates of no interest. Data were smoothed with a full-width-at-half-maximum (FWHM) Gaussian kernel of 5 mm.

**GLM for LOC localizer.** A GLM was estimated for each participant, with regressors created by convolving a canonical HRF with blocks of Objects and Scrambled-objects conditions. Movement parameters and run means were included as covariates of no interest. Data were smoothed with a FWHM Gaussian kernel of 5 mm.

#### Univariate analysis

Whole-brain random-effects analysis across participants examined voxelwise BOLD response during task performance to confirm the recruitment of MD cortex. Beta estimates of the main task GLM were averaged across all single-stimulus conditions, blocks, runs, cues, and category types and were contrasted with fixation blocks to generate a *t*-contrast map. The contrast was smoothed by a kernel of 14 mm. To assess average activity for each of the conditions (T, NI, NC),  $\beta$  values were averaged across blocks, runs, cues, category type, and voxels in each ROI within the MD network template (see ROIs, below).

#### ROIs

We used a conjunction of template ROIs and single-subject localizer data to select voxels within each ROI (feature selection) to be used in the MVPA. For each participant and ROI, we first selected voxels located within the template ROI. Then, within the ROI template, we selected the *N* voxels with the largest *t* values as derived from the relevant subject-specific localizer independent dataset and relevant contrast (MD regions: Hard > Easy contrast, *n* = 200 voxels for each ROI; LOC: Objects > Scrambled-objects contrast, *n* = 200 voxels, with *n* = 180 voxels for its anterior and posterior subregions when these were separately analyzed; see LOC template, below).

**MD regions template.** We used a template for the MD network ROIs in MNI space as defined in a separate study contrasting hard and easy conditions across seven tasks and 40 participants (Fedorenko et al., 2013; see *t*-map at <http://imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem>). The regions within this network include the AI; IPS; anterior, middle, and posterior parts of the middle frontal gyrus (MFG-ant, MFG-mid, MFG-post, respectively), pre-SMA, and FEF, symmetrically defined in left and right hemispheres (Fig. 2B).

**LOC template.** LOC was defined using a functional localizer that identified object-selective cortex (Malach et al., 1995) in an independent study with 15 participants (Lorina Naci, PhD dissertation, University of Cambridge). BOLD responses evoked by forward- and backward-masked objects were contrasted to those evoked by masks alone (Cusack et al., 2012). We further divided the LOC into an anterior part, the posterior fusiform region (pFs) of the inferior temporal cortex, and a posterior part, the lateral occipital region (LO) using a cutoff MNI coordinate of  $y = -62$  (Fig. 3).

#### MVPA

Our main pattern analysis was ROI based and hypothesis driven, testing for the modulation of categorical information by behavioral relevance in the MD network regions as well as high-level visual cortex regions. ROIs and subject-specific feature selection were used as described above in the “ROIs” section. As a complementary approach, we conducted a whole-brain searchlight pattern analysis, to identify additional regions in which representation of categorical information is modulated by behavioral relevance (Kriegeskorte et al., 2006) (see Whole-brain searchlight pattern analysis, below).

**Categorical discrimination.** We used MVPA to test for categorical discrimination and its modulation by behavioral relevance. We computed discrimination between all possible pairs of categories, as illustrated in the classification matrix in Figure 1D. Each entry in the matrix corresponds to discrimination between two categories as indicated by the combination of row and column categories. Each of the six categories was considered separately depending on the preceding cue (cue A, cue B), resulting in 12 conditions as indicated in the number of rows and columns in the classification matrix. For example, the inanimate category from pair A (A-i) was considered separately depending on the cue, with status target (T) following Cue A and inconsistent nontarget (NI) following Cue B. Therefore, this category, as well as all other categories, appears twice (following each of the two cues) in each row/column in the classification matrix.

To test for the effect of behavioral relevance on category discrimination, we considered separately pairs of categories for which the distinction between them was behaviorally relevant or not. Importantly, if for a given pair of categories, both categories had the same behavioral status (e.g., T and T, NI and NI, NC and NC), then the discrimination between

these two categories was not behaviorally relevant (Fig. 1D, blue entries). In contrast, for pairs of categories with different behavioral status (e.g., T and NI, T and NC, NI and NC), the discrimination between them was behaviorally relevant (Fig. 1D, red entries). NIs and NCS differed in their behavioral status because NIs are potential targets, whereas NCs are never targets. Therefore, NI–NC pairs were considered as a distinction that is behaviorally relevant.

We were also interested in testing for the modulation of categorical discrimination by other aspects of the task. Therefore, within behaviorally relevant/irrelevant category distinctions, we further divided the pairs of categories depending on whether they had the same/different preceding cue (Cue A, Cue B) and whether they were from the same/different category type (Animate, Inanimate) (see color-coding and entry naming conventions in the classification matrix in Fig. 1D). Both of these distinctions (the preceding cue and category type) were orthogonal to the final behavioral response.

**Pattern analysis.** We applied classification using a correlational approach on bias patterns (Stokes et al., 2009). For each pair of categories, we computed the relative activation bias associated with one category compared with another by subtracting the voxelwise  $\beta$ -estimates for the two categories. The voxelwise bias pattern was then normalized by the univariate noise estimates by computing the voxelwise  $t$ -contrast of the two categories using SPM. This procedure was repeated for each of the eight blocks across the four runs. To estimate the generalizability of the bias pattern across runs, we used a leave-one-run-out cross-validation. We correlated the average bias pattern across three runs (training set), with each of the two bias patterns in the remaining run (test set). Correlation coefficients larger than zero were considered as correct classification of the category pair bias pattern and coefficients of zero or smaller were considered as incorrect classification. We repeated this procedure four times across all possible training-test permutations, resulting in eight test iterations. The proportion of correct classification was then computed across the eight iterations. Classification accuracy scores were then averaged for each participant across entries of the classification matrix as appropriate for each contrast (see Results) and tested against chance of 50% using second-level group analysis (one-tailed  $t$  test). Group accuracy scores significantly larger than chance were interpreted as being a representation of categorical discrimination across the distributed pattern of activity. Because data for each type of T stimulus were split into two halves (equating numbers of T and N trials in each regressor; see above), each classification accuracy involving a T category was measured twice, once for each half of the data, and the two resulting values were averaged.

To check the robustness of the results, we repeated the pattern analysis using a support vector machine (SVM) classifier (LibSVM library for MATLAB). For each pair of categories, we computed the voxelwise  $\beta$ -estimates for the two categories for each of the eight blocks across the four runs. Then, we used a leave-one-run-out (two patterns) cross-validation and averaged the obtained classification accuracy across the four training-test iterations.

**Whole-brain searchlight pattern analysis.** To identify additional regions that might show modulation of categorical information by behavioral relevance, we conducted a whole-brain searchlight pattern analysis (Kriegeskorte et al., 2006). This technique also allows testing for information coding in more spatially restricted regions. Searchlight analysis was computed for each participant and followed by a second-level group analysis. First, a group mask was computed based on the individual normalized brain masks of all of the participants. The group mask was composed of only voxels that were included in all of the individual brain masks. Then, for each participant, data were extracted from spherical ROIs (20 voxels) centered on each voxel in the brain mask. Searchlight ROIs near the surface of the brain were extended laterally to keep the number of voxels in each ROI constant. A classification matrix was computed for each participant and each spherical ROI across the brain using the correlational approach that was used for the main ROI-based pattern analysis. For each voxel, the contrast “behaviorally relevant discrimination” versus “behaviorally irrelevant discrimination” (red vs blue entries in the classification matrix, respectively) was computed, generating individual whole-brain classification accuracy maps. The individual contrast

maps were subsequently smoothed by a 10 mm FWHM Gaussian kernel and entered into a second-level random-effects group analysis (SPM 8).

### Task difficulty level and classification results

Activity in the MD network is known to be modulated by task difficulty. Therefore, we tested for the possibility that discrimination effects may be related to differences in difficulty level of the task conditions, as reflected in differences in reaction time (RT). Importantly, in the GLM of our main analysis, trials were modeled as epochs lasting from stimulus onset to response, thus controlling for lengthened BOLD response as a result of longer response time (Woolgar et al., 2014). We further tested for the potential contribution of RT to our classification results using two separate control analyses.

In the first analysis, for each participant, we correlated classification accuracy across the MD network for all pairs of categories with the absolute difference in RT between them. Correlations were computed separately for category pairs for which the distinction between them was behaviorally relevant and for category pairs for which the distinction was behaviorally irrelevant, as indicated by red (48 pairs) and blue (18 pairs) entries in the classification matrix (Fig. 1D), respectively. Correlation coefficients were averaged across participants and tested against 0 using a one-tailed  $t$  test. Furthermore, we used regression of classification scores over the absolute difference in RT, computed separately for each participant and for behaviorally relevant and irrelevant distinctions between category pairs. For each regression, the intercept indicates the predicted accuracy score when there is no difference in RT/difficulty level. An intercept larger than chance (50%) implies that accuracy score is above chance even when there is no difference in difficulty level, therefore suggesting that discrimination is not driven by task difficulty. We computed intercepts for pairs of categories with behaviorally relevant and irrelevant distinctions separately across all participants and tested for the average intercept against 50% using a one-tailed  $t$  test.

In the second control analysis, we regressed out RT at the GLM level and then repeated the classification analysis. In the GLM, we used an additional regressor for all events (trials) with fixed duration (0.5 s) and with parametric modulation by a quadratic polynomial expansion of RT to regress out both linear and nonlinear effects of RT (Todd et al., 2013; Woolgar et al., 2014). The rest of the events were modeled with fixed duration (0.5 s). Then, we computed the voxelwise bias patterns as before and repeated the classification analysis.

### Control for ROI size

To be able to compare MVPA results between regions, we controlled for ROI size by using feature selection with a fixed number of voxels. To verify that the results that we obtained are robust and not limited to a specific choice of ROI size, we repeated the analysis with a varying size of ROI (100, 150, 250, and 300 voxels).

## Results

### Behavioral results

Mean RTs were  $702 \pm 31$ ,  $860 \pm 49$ , and  $702 \pm 34$  ms for the T, NI, and NC conditions, respectively. RTs for the three conditions were significantly different (one-way repeated-measures ANOVA:  $F_{(2,34)} = 42.3$ ,  $p < 0.001$ ), with larger RT for NI than T and NC (paired  $t$  test for T vs NI and NI vs NC, respectively:  $t_{(17)} = 6.9$ ,  $p < 0.001$ ;  $t_{(17)} = 7.5$ ,  $p < 0.001$ ) and no difference in RT for T vs NC (paired  $t$  test:  $t_{(17)} = 0.1$ ,  $p > 0.9$ ). The percentage of correct trials was  $94.6 \pm 0.6\%$ ,  $92.1 \pm 1\%$ , and  $98.1 \pm 0.4\%$  for the T, NI, and NC conditions, respectively. The proportion of correct responses was significantly different across the three conditions (one-way repeated-measures ANOVA:  $F_{(2,34)} = 26.1$ ,  $p < 0.001$ ), with the highest success rate for NC and the lowest for NI (paired  $t$  test:  $t_{(17)} = 2.9$ ,  $p = 0.009$  for T vs NI;  $t_{(17)} = 5.4$ ,  $p < 0.001$  for T vs NC;  $t_{(17)} = 6.2$ ,  $p < 0.001$  for NI vs NC). As expected, the best performance overall was for NC stimuli, the status of which as nontargets was fixed throughout the test session.

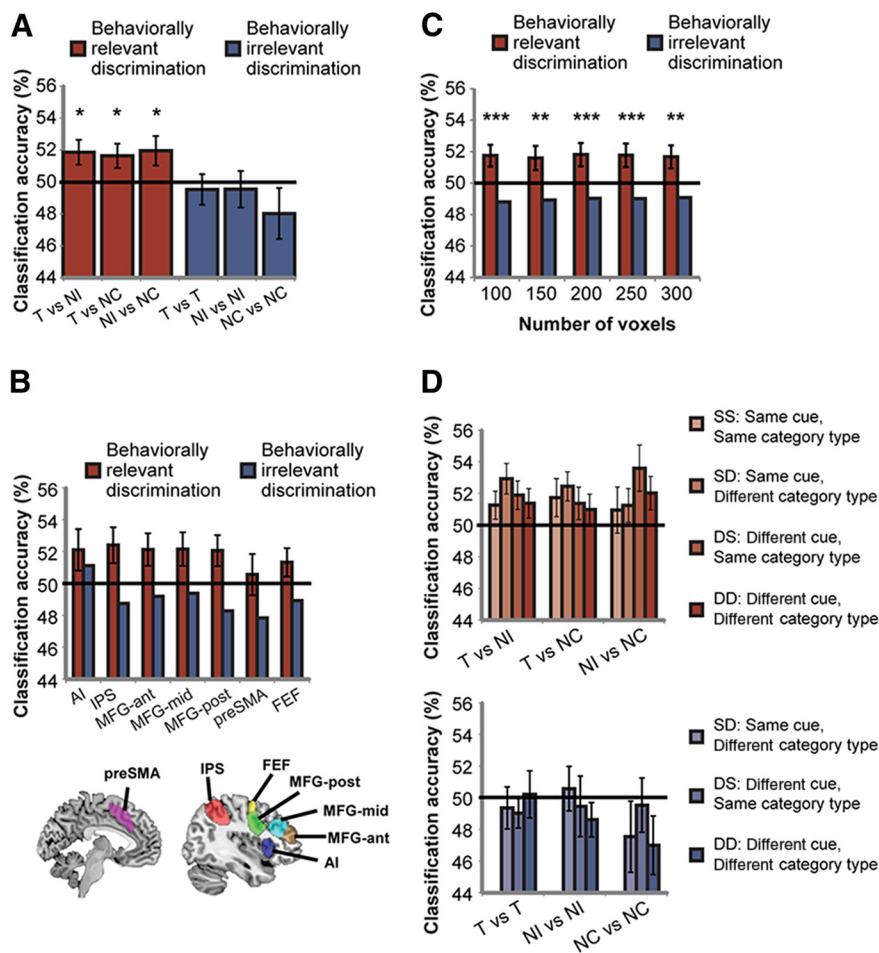
### Univariate neural activity

To examine the recruitment of the MD network during performance of the task, we first conducted univariate random-effects analysis across participants, comparing mean activity across all trial types to fixation blocks. Although no voxels survived correction for multiple comparisons across the whole brain, at lower threshold ( $t = 1.5$ ), a pattern of MD activity was revealed, including the MFG, FEF, IPS, AI, and pre-SMA. An additional typical visual component in the extrastriate cortex was evident. We further examined differences in the average univariate activity between the task conditions within the MD network ROIs, as defined by the template (see ROIs, above). Within each MD region, average  $\beta$  estimates were similar for the T, NI, and NC conditions (one-way repeated-measures ANOVAs:  $F_{(2,34)} < 1$ ,  $p > 0.4$ ). Similarly, there were no differences between conditions in LOC ( $F_{(2,34)} < 1$ ,  $p > 0.7$ ).

### Modulation of category discrimination by behavioral relevance in the MD network

Consistent with the role of the MD network in attentional and cognitive control, we predicted that categorical discrimination, as measured using MVPA, would be modulated by the relevance of the categorical distinction to behavior. In the present task, a distinction between categories that have a different behavioral status, such as a target and a nontarget, is behaviorally relevant. In contrast, a distinction between categories that share the same behavioral status, such as animate and inanimate targets, is behaviorally irrelevant. Therefore, we expected that the multi-voxel pattern of activity across the MD network would reflect the behavioral relevance of the categorical discrimination, with accuracy of pattern discrimination above chance for two categories with different behavioral status (red entries in the classification matrix in Fig. 1D), but not for two categories with the same behavioral status (blue entries in the classification matrix in Fig. 1D).

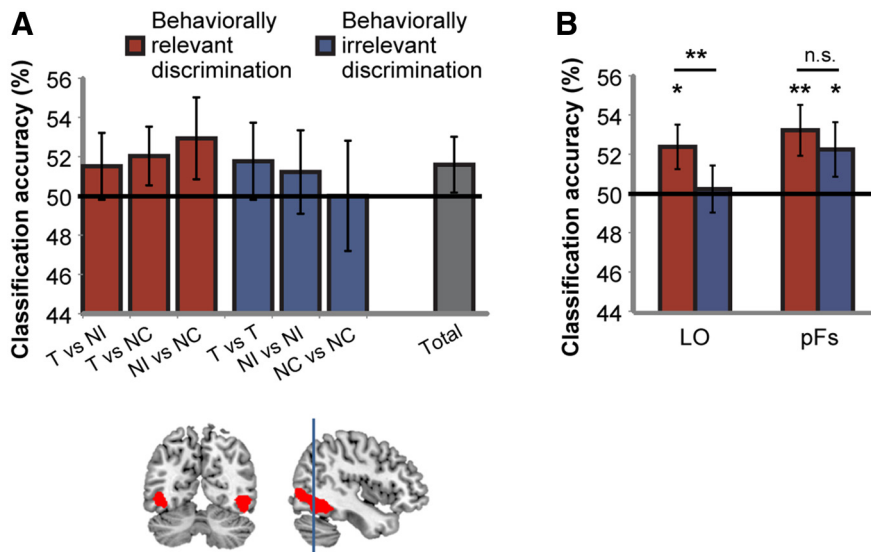
For each participant and ROI, category discriminability was computed using MVPA for all possible category pairs, sorted according to same/different behavioral status (T, NI, NC), cue (Cue A, Cue B), and category type (Animate, Inanimate) (see classification matrix in Fig. 1D; see Materials and Methods for more details). To test for category discrimination when it is behaviorally relevant, we examined T versus NI, T versus NC, and NI versus NC discriminations separately. For each of these cases and for each participant, we averaged the separate accuracy values obtained for each of the 16 contributing category pairs (entries with red hues in the classification matrix; Fig. 1D). To assess T versus NI discrimination, for example, we averaged the 16 separate values from the top-middle block in Figure 1D. The analysis



**Figure 2.** Behavioral relevance modulates categorical discrimination across the MD network. **A**, Categorical distinction that is behaviorally relevant is represented across the MD network, as measured by distributed neural pattern of activity (red bars). Distinction between categories that is behaviorally irrelevant is not well represented (blue bars).  $*p < 0.05$ . **B**, A similar pattern of activity is evident across all regions within the MD network. MD regions template is shown on sagittal and coronal planes. Error bars indicate SEM of the difference between behaviorally relevant and irrelevant discriminations in each region. **C**, Modulation of category discrimination by behavioral relevance in the MD system was robust across a large range of ROI sizes. Error bars indicate SEM of the difference between behaviorally relevant and irrelevant discriminations.  $**p < 0.01$ ,  $***p < 0.001$ . **D**, Categorical discrimination is not modulated by aspects of the task that are orthogonal to the behavioral decision, such as cue and category type. Color coding of bars follows the classification matrix in Figure 1D. Error bars indicate SEM. AI, anterior insula; IPS, intraparietal sulcus; MFG-ant, middle frontal gyrus— anterior part; MFG-mid, middle frontal gyrus—middle part; MFG-post, middle frontal gyrus—posterior part; preSMA, presupplementary motor area; FEF, frontal eye field.

was conducted separately for each of the MD ROIs and the results were then averaged across ROIs. As predicted, category discrimination across participants was above chance (50%) for behaviorally relevant distinctions (one-tailed  $t$  test: T vs NI:  $t_{(17)} = 2.39$ ,  $p = 0.014$ ; T vs NC:  $t_{(17)} = 2.16$ ,  $p = 0.022$ ; NI vs NC:  $t_{(17)} = 2.12$ ,  $p = 0.025$ ). Also consistent with our prediction, discrimination for categorical distinctions that are behaviorally irrelevant was not above chance (entries with blue hues in the classification matrix; one-tailed  $t$  test:  $t_{(17)} < 1.24$ ,  $p > 0.16$ ). Discrimination was significantly larger for behaviorally relevant than for behaviorally irrelevant categorical distinctions (one-tailed paired  $t$  test:  $t_{(17)} = 3.76$ ,  $p < 0.001$ ). Importantly, a similar pattern of results was evident across all regions within the MD network (Fig. 2B). A  $7 \times 2$  repeated-measures ANOVA with MD ROI and behavioral relevance of distinction (relevant, irrelevant) as factors revealed a significant main effect of behavioral distinction ( $F_{(1,17)} = 8.45$ ,  $p = 0.01$ ), no effect of ROI ( $F_{(6,102)} = 0.91$ ,  $p = 0.49$ ), and no significant interaction ( $F_{(6,102)} = 1.54$ ,  $p = 0.17$ ). For all of the





**Figure 3.** Categorical discrimination and behavioral relevance in LOC. **A**, In LOC, categorical discrimination is similar for both behaviorally relevant and irrelevant category distinctions (red and blue bars, respectively). Overall, the classification accuracy across all category pairs is shown on the right (gray bar). **B**, In the LO subregion, categorical discrimination is modulated by behavioral relevance. In the pFs subregion, categorical discrimination is similar and above chance for both behaviorally relevant and irrelevant category distinctions.  $*p < 0.05$ ,  $**p < 0.01$ , n.s., not significant. Bottom, LOC template shown on sagittal and coronal planes, with vertical blue line indicating division into posterior (LO) and anterior (pFs) subregions. Error bars indicate SEM.

ROIs, similar results were obtained for the two hemispheres [ $2 \times 2$  repeated-measures ANOVA for each ROI with hemisphere and behavioral relevance of distinction (relevant, irrelevant) as factors: no interaction,  $F_{(1,17)} < 1.8$ ,  $p > 0.19$ ]. The modulation of categorical distinction by behavioral relevance was robust across a large range of ROI sizes (100, 150, 200, 250, 300 voxels), with average classification above chance for behaviorally relevant discriminations (one-tailed  $t$  test:  $t_{(17)} > 2.87$ ,  $p < 0.006$ ), but not for behaviorally irrelevant discriminations (classification accuracy  $< 50\%$ ). Classification accuracy for behaviorally relevant discriminations was larger than accuracy for irrelevant discriminations across the range of ROI sizes (one-tailed paired  $t$  test:  $t_{(17)} > 4.2$ ,  $p < 0.002$ ; Fig. 2C). To ensure that our results were consistent across classification methods, the analysis was repeated using SVM, yielding accuracy scores and results similar to the correlation-based pattern analysis. Across MD regions, the average classification was above chance for behaviorally relevant discriminations (one-tailed  $t$  test:  $t_{(17)} = 3.4$ ,  $p = 0.002$ ), but not for behaviorally irrelevant discriminations (classification accuracy  $< 50\%$ ), with the former larger than the latter (one-tailed paired  $t$  test:  $t_{(17)} = 5.2$ ,  $p < 0.001$ ). These results suggest that behavioral relevance modulates categorical discrimination throughout the MD network.

Our design also allowed us to test for the modulation of category discrimination by two additional factors. For each of the three behaviorally relevant category distinctions (T vs NI, T vs NC, and NI vs NC) and three behaviorally irrelevant category distinctions (T vs T, NI vs NI, and NC vs NC), we further divided the pairs according to whether they had the same or different preceding cue (Cue A, Cue B) and similarly by whether they had the same or different category type (Animate, Inanimate). This resulted in four combinations of category pairs: “Same cue and Same category type” (SS), “Same cue and Different category type” (SD), “Different cue and Same category type” (DS), and “Different cue and Different category type” (DD) (see acronyms and different hues of blue and red entries in the classification

matrix in Fig. 1D). For example, for T versus NI category pairs (top-middle 16-entry square in the classification matrix), we computed classification accuracy separately for pairs of categories with the same preceding cue and the same category type (SS, light pink entries: A-a vs B-a following Cue A, A-i vs B-i following Cue A, B-a vs A-a following Cue B, B-i vs A-i following Cue B) and then averaged across them (Fig. 2D, top, leftmost bar in left group). Similar averaging was done for SD, DS, and DD cases (Fig. 2D, top, remaining bars in left group) and for T versus NC (Fig. 2D, top, middle group) and NI versus NC (Fig. 2D, top, right group). For category pairs with behaviorally irrelevant distinctions (T vs T, NI vs NI, NC vs NC), the combination of SS was the discrimination between the category and itself. This could only be computed by splitting the data into halves, which was not undertaken because of the low number of trials per condition. Therefore, for the behaviorally irrelevant distinctions, only SD, DD, and DS pairs were computed.

Because cue and category type were orthogonal to the behavioral decision, we predicted that they should be poorly coded in the MD network. In other words, we predicted that there would be no difference in classification accuracy between SS, SD, DS, and DD pairs within behaviorally relevant category distinctions and between SD, DS, and DD pairs within behaviorally irrelevant category distinctions. Consistent with this prediction, we found that discrimination was similar regardless of whether the same or different cue or category type was used (for behaviorally relevant category distinctions, one-way repeated-measures ANOVA with levels SS, SD, DS, DD: T vs NI:  $F_{(3,51)} = 1.83$ ,  $p = 0.15$ ; T vs NC:  $F_{(3,51)} = 0.58$ ,  $p = 0.63$ ; NI vs NC:  $F_{(3,51)} = 1.36$ ,  $p = 0.27$ ; for behaviorally irrelevant category distinctions, one-way repeated-measures ANOVA with levels SD, DS, DD: T vs T:  $F_{(2,34)} = 0.37$ ,  $p = 0.7$ ; NI vs NI:  $F_{(2,34)} = 0.67$ ,  $p = 0.52$ ; NC vs NC:  $F_{(2,34)} = 0.94$ ,  $p = 0.4$ ). These data suggest that category discriminability in the MD network is not modulated by aspects of the task that are orthogonal to the behavioral decision.

### Category discrimination based on behavioral relevance in LOC

The MD network is thought to impose top-down modulations on occipital visual regions to control attention. We therefore sought to test whether category discrimination is modulated by behavioral relevance in the visual object region LOC and pursued an analysis similar to the one conducted for the MD network (Fig. 3A). Across all pairs of categories (all entries in the classification matrix), discrimination was above the chance level (50%), but this trend was not significant (one-tailed  $t$  test:  $t_{(17)} = 1.15$ ,  $p = 0.13$ , rightmost bar in Fig. 3A). Category discrimination was not significantly above chance across behaviorally relevant (Fig. 1D, red entries in the classification matrix) or behaviorally irrelevant (blue entries in the classification matrix) distinctions (one-tailed  $t$  test: T vs NI:  $t_{(17)} = 0.88$ ,  $p = 0.19$ ; T vs NC:  $t_{(17)} = 1.36$ ,  $p = 0.1$ ; NI vs NC:  $t_{(17)} = 1.4$ ,  $p = 0.09$ ; T vs T:  $t_{(17)} = 0.9$ ,  $p = 0.19$ ; NI vs NI:  $t_{(17)} = 0.57$ ,  $p = 0.29$ ; NC vs NC:  $t_{(17)} = 0$ ,  $p = 0.5$ ). Discrim-

ination was not significantly larger for behaviorally relevant compared with behaviorally irrelevant distinctions (one-tailed paired  $t$  test:  $t_{(17)} = 1.52, p = 0.073$ ).

Because it has been suggested that the two subregions within the LOC, the pFs and LO, may be involved in different aspects of object recognition (MacEvoy and Epstein, 2011; Harel et al., 2014), we further tested whether behavioral relevance modulates categorical representation in these subregions (Fig. 3B). Interestingly, in the pFs, classification was above chance for both behaviorally relevant and irrelevant discriminations (one-tailed  $t$  test:  $t_{(17)} = 2.6, p = 0.009$  for relevant discriminations;  $t_{(17)} = 1.7, p = 0.05$  for irrelevant discriminations), with no difference between them (one-tailed paired  $t$  test:  $t_{(17)} = 1.05, p = 0.15$ ). In contrast, in LO, classification was above chance for behaviorally relevant discriminations only (one-tailed  $t$  test:  $t_{(17)} = 2.21, p = 0.02$  for relevant discriminations;  $t_{(17)} = 0.3, p = 0.4$  for irrelevant discriminations), with larger classification for relevant than irrelevant discriminations (one-tailed paired  $t$  test:  $t_{(17)} = 2.82, p = 0.006$ ). A  $2 \times 2$  repeated-measures ANOVA with ROI (pFs, LO) and behavioral relevance (relevant, irrelevant) as factors revealed a main effect of behavioral relevance ( $F_{(1,17)} = 5.2, p = 0.036$ ) but no interaction ( $F_{(1,17)} = 1.39, p = 0.26$ ). Overall, our results show category discrimination in the pFs regardless of whether it is behaviorally relevant or not, whereas category discrimination in the LO is modulated by behavioral relevance and is above chance only when it is relevant.

### Whole-brain searchlight pattern analysis

As a complementary approach to the main ROI-based classification analysis, we searched for additional regions across the brain in which categorical information is modulated by behavioral relevance. We therefore conducted a whole-brain searchlight analysis (Kriegeskorte et al., 2006) for the main contrast of interest, behaviorally relevant discriminations versus behaviorally irrelevant discriminations. Searchlight analysis was conducted at the single-participant level and the resulting individual maps were entered into a second-level group analysis. None of the voxels survived false discovery rate correction for multiple comparisons ( $p < 0.05$ ). Using a lower threshold ( $p < 0.01$ , uncorrected), salient regions of discriminations were as follows: middle and posterior parts of the MFG, IPS and adjacent supramarginal gyrus, as well as the lateral occipital cortex on the left hemisphere; the AI, FEF, and the lateral occipital cortex on the right hemisphere. Overall, this analysis does not suggest additional brain regions beyond the MD network and higher visual cortex that show modulation of categorical information by behavioral relevance.

### Category discrimination, behavioral relevance, and task difficulty in the MD network

Activity in the MD network is highly driven by task difficulty. Therefore, we tested for the possibility that modulation of category discrimination by behavioral relevance in the MD network was driven by differences in difficulty level of the task conditions using two separate control analyses. In the first analysis, we correlated for each participant the accuracy scores for pairs of categories with the absolute difference in RT between them. Because our results showed discrimination above chance for behaviorally relevant distinctions but not for behaviorally irrelevant distinctions, we computed correlations separately for these two groups of pairs of categories (Fig. 1D, red and blue entries in the classification matrix). Overall, correlation coefficients of accuracy scores and absolute difference in RT were negligible and not sig-

nificantly different from zero (behaviorally relevant distinctions: mean  $\pm$  SEM  $0.04 \pm 0.05, t_{(17)} = 0.8, p = 0.22$ ; behaviorally irrelevant distinctions: mean  $\pm$  SEM  $0.01 \pm 0.06, t_{(17)} = 0.17, p = 0.43$ ), suggesting that accuracy levels were unrelated to differences in difficulty levels between conditions. We further tested for prediction of discriminability by difficulty level using regression of accuracy scores over absolute difference in RT. This was done at the single-subject level, and the regression was computed separately for behaviorally relevant and irrelevant distinctions. In the regression, the intercept is the expected accuracy score for a difference in RT of zero. In particular, for behaviorally relevant category distinctions, in which the overall accuracy score was above chance (50%), an intercept larger than 50% means that, even if RTs were the same for a given pair of categories, the accuracy score would still be above chance, implying that accuracy score is not driven by task difficulty. The average intercept across participants for behaviorally relevant category distinctions was marginally significantly above the chance level (mean  $\pm$  SEM:  $51.4 \pm 0.9$ , one-tailed  $t$  test against 50%:  $t_{(17)} = 1.45, p = 0.08$ ). For behaviorally irrelevant category distinctions, in which the overall accuracy score was not above chance, the average intercept was not above the chance level either (mean  $\pm$  SEM:  $48.7 \pm 1.4$ ). The average intercept was marginally significantly larger for behaviorally relevant than irrelevant distinctions (one-tailed paired  $t$  test:  $t_{(17)} = 1.57, p = 0.068$ ). The overall negligible correlation coefficients, together with the accuracy score above chance for behaviorally relevant distinctions even when there was no RT difference, suggest that task difficulty is not a serious confound in our findings.

In a second analysis to control for task difficulty differences between conditions, we regressed out RT at the GLM level (Todd et al., 2013; Woolgar et al., 2014) and repeated the classification analysis. This analysis yielded accuracy scores and results similar to the ones that were obtained in our main analysis. In the MD network, the average classification accuracy for behaviorally relevant discriminations was significantly above chance (one-tailed  $t$  test:  $t_{(17)} = 2.65, p = 0.008$ ) and was larger than the classification accuracy for the behaviorally irrelevant discriminations (one-tailed paired  $t$  test:  $t_{(17)} = 5.1, p < 0.001$ ). The average classification for the irrelevant discriminations was not significantly above chance (classification accuracy  $< 50\%$ ). The similar results obtained using this control analysis further reassured us that our findings were not driven by differences in difficulty level.

## Discussion

In this study, we found that the behavioral relevance of a categorical distinction modulates categorical discrimination across the MD network, as represented in the distributed pattern of activity. Using MVPA, classification accuracies for pairs of visual categories for which the categorical distinction was behaviorally relevant were above chance, implying that information about this distinction is coded across the MD network. In contrast, information about categorical distinctions that were behaviorally irrelevant was not coded.

Our results support the involvement of the MD network in adaptive control of cognitive function and goal-directed behavior (Norman and Shallice, 1980; Desimone and Duncan, 1995; Miller and Cohen, 2001; Duncan, 2013). Several previous neuroimaging studies have demonstrated the involvement of frontal and parietal cortex in the representation of task-relevant information (Jiang et al., 2007; Li et al., 2007; Liu et al., 2011; Woolgar et al., 2011a; Woolgar et al., 2011b; Chen et al., 2012; Lee et al., 2013; Harel et al., 2014; Woolgar et al., 2015). Our findings pro-



vide another tier of evidence for MD network function: using a categorization-based task, we showed that task-critical information is represented across the network, whereas task-irrelevant information is represented less strongly or not at all. In our paradigm, distinctions between category pairs that had different behavioral status (e.g., targets and nontargets) were consistently represented across the MD network, in contrast to categorical distinctions (such as two target categories) that were behaviorally irrelevant. Moreover, two other dimensions of the task, namely cue (Cue A, Cue B) and category type (Animate, Inanimate), were orthogonal to the decision required and were not significantly discriminated. Notably, animate and inanimate category types were chosen to maximize the potential discrimination, as animate and inanimate objects have been shown to be highly discriminable in the visual system (Kriegeskorte et al., 2008). Nevertheless, no discrimination was seen in the current task across MD regions. Lack of modulation of neural representation by irrelevant features of a task is a critical property of a system that is involved in attentional control: not only is it important to focus attention on behaviorally relevant information, but at the same time, it is essential to filter out information that is currently irrelevant. Our findings provide a broad demonstration of this principle across the MD network.

Our results complement and extend previous findings from the NHP reporting task-dependent coding of categorical distinctions in single cells of both LPFC (Freedman et al., 2001) and parietal cortex (Fitzgerald et al., 2012; Swaminathan and Freedman, 2012). Although these studies provide detailed information at the level of single-cell activity, they are limited in the breadth of brain coverage and scale of brain network under study. Here, we go beyond this by looking at the entire MD network. Similarly to the evidence from the NHP, we find modulation of information coding by behavioral relevance. Of particular relevance to this study are the findings from the NHP by Kadohisa et al. (2013) because the two studies share a very similar design. In this NHP study, a cue at trial onset indicated which of two visual stimuli (drawings of real-world objects) was the current target. The results showed many cells in the LPFC with selective responses as early as 100 ms after stimulus onset based on the behavioral relevance of the stimuli (T, NI, NC, defined analogously to the current study), rather than on physical stimulus identity. In partial agreement with the current findings, coding of cue was less common, although present in some individual cells. The accumulating evidence across both humans and NHPs further demonstrate the ability of the MD network to focus on currently relevant information whether it is category based (visually or verbally related), cue associated, or more abstract.

In our study, MD coding of behaviorally relevant distinctions (T, NI, or NC) could not simply be interpreted as coding of responses. Critically, the S-R mapping linking the decision (target present or absent) to the response (left or right keypress) was reversed every block, unconfounding decision and response. Also relevant is the significant discrimination between the NI and NC categories. Arguably, NI and NC stimuli were classified on a different basis, NI based on the current cue but NC based on long-term learning. The data suggest that these different classification procedures led to different patterns of MD activity despite the same final response.

In our results, all regions within the MD network exhibited a similar pattern of response, discriminating only those categories that were critical to the task, which is consistent with previous findings (Duncan, 2010; Woolgar et al., 2011b). Various suggestions have been made about division of the MD system into sep-

arate components, for example, into cingulo-opercular and frontoparietal subnetworks (Dosenbach et al., 2007; Sadaghiani and D'Esposito, 2014), but separate MD regions also show many similar properties. Similar properties may reflect, at least in part, the coarse time scale of the BOLD signal. Frontal and parietal regions in the NHP, for example, exchange information at a time scale of tens of milliseconds (Salazar et al., 2012). Similarly, neurons within a single region code task events with rapidly changing population vectors of activity (Meyers et al., 2008; Kadohisa et al., 2013; Stokes et al., 2013). Rapid evolution of signals within each region and rapid exchange from one region to another may make functional differences between regions hard to observe at the time scale of BOLD signal.

In our study, classification accuracies were low but statistically significant. Similar classification rates obtained using MVPA of fMRI data are not uncommon in decoding task events across the MD network (Woolgar et al., 2011a; Woolgar et al., 2011b) and in the auditory and visual cortex (MacEvoy and Epstein, 2011; Alink et al., 2012). The expected effect size of classification accuracy is generally unknown and may be limited, for example, by the underlying spatial arrangement of critical neural populations (Dubois et al., 2015). Regardless of absolute classification accuracy, our results show the robust difference in MD coding of task-critical and task-irrelevant categorizations.

Using whole-brain imaging, we could investigate categorical discrimination not only across the MD network, but also in the visual cortex, in particular the high-level visual region, the LOC. Several other studies have looked at attentional modulations in the visual cortex using low-level stimuli such as gratings (Li et al., 2007; Serences et al., 2009; Jehee et al., 2011) or high-level objects (Chen et al., 2012; Lee et al., 2013; Harel et al., 2014) and addressing issues including training (Jiang et al., 2007), visual clutter (Reddy and Kanwisher, 2007; Reddy et al., 2009; Erez and Yovel, 2014), and attentional effects on functional connectivity (Chadick and Gazzaley, 2011). Contextual modulations of object representation have been found in LOC, with some studies reporting such modulations in pFs (Lee et al., 2013; Harel et al., 2014), but not in LO (Harel et al., 2014), and others reporting flexible categorization in LO (Jiang et al., 2007; Li et al., 2007). More broadly in the visual system, attentional modulations were reported in category-selective regions such as the fusiform face area (FFA) and parahippocampal scene-selective area (PPA) (Reddy and Kanwisher, 2007; Reddy et al., 2009; Chadick and Gazzaley, 2011; Chen et al., 2012) and early visual cortex (Serences et al., 2009; Jehee et al., 2011). Interestingly, in our study, a division of LOC into its two subregions revealed representation of both behaviorally relevant and irrelevant category distinctions in the pFs, whereas the representation in the LO was modulated by behavioral relevance and limited to relevant category distinctions only. Nonetheless, the overall category discrimination in LOC was weak. Possibly, a combination of several factors in the complex design of the paradigm that we used contributed to suboptimal classification power. These factors include the peripheral stimulus display while fixating, very brief presentation (150 ms), fast event-related design, and frequent changes of target categories. The accumulating mixed evidence suggests that the pFs and LO may have different roles in object processing and its contextual modulation. Further studies are required to better understand the contribution of LOC and its subregions, as well as other regions within the visual system, to context-dependent object processing.

To conclude, our data suggest that coding of stimulus events in the MD network is determined by behavioral relevance. In a cued category-detection task, neural patterns of activity across the MD network conveyed information about the categorical distinctions that were behaviorally relevant. In contrast, discrimination was not seen for categorical distinctions that were behaviorally irrelevant and, at least in our data, discrimination was also not visible for other aspects of the task that were orthogonal to the behavioral decision. The ability to carry information about those aspects of the task that are relevant to decision making and filter out aspects currently not relevant is an essential property of a neural system adapting to the dynamic requirements of behavior. Our findings are consistent with the view of the key role of the MD network in attentional and cognitive control and provide important evidence for its ability to focus attentional resources on behaviorally relevant information.

## References

- Alink A, Euler F, Kriegeskorte N, Singer W, Kohler A (2012) Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Hum Brain Mapp* 33:969–978. [CrossRef Medline](#)
- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451–459. [Medline](#)
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436. [CrossRef Medline](#)
- Chadick JZ, Gazzaley A (2011) Differential coupling of visual cortex with default or frontal-parietal network based on goals. *Nat Neurosci* 14:830–832. [CrossRef Medline](#)
- Chen AJ, Britton M, Turner GR, Vytlačil J, Thompson TW, D'Esposito M (2012) Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. *Front Hum Neurosci* 6:187. [Medline](#)
- Cromer JA, Roy JE, Miller EK (2010) Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66:796–807. [CrossRef Medline](#)
- Cusack R, Veldsman M, Naci L, Mitchell DJ, Linke AC (2012) Seeing different objects in different ways: measuring ventral visual tuning to sensory and semantic features with dynamically adaptive imaging. *Hum Brain Mapp* 33:387–397. [CrossRef Medline](#)
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222. [CrossRef Medline](#)
- Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE (2007) Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci U S A* 104:11073–11078. [CrossRef Medline](#)
- Dubois J, de Berker AO, Tsao DY (2015) Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci* 35:2791–2802. [CrossRef Medline](#)
- Duncan J (2010) The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci* 14:172–179. [CrossRef Medline](#)
- Duncan J (2013) The structure of cognition: attentional episodes in mind and brain. *Neuron* 80:35–50. [CrossRef Medline](#)
- Erez Y, Yovel G (2014) Clutter modulates the representation of target objects in the human occipitotemporal cortex. *J Cogn Neurosci* 26:490–500. [CrossRef Medline](#)
- Everling S, Tinsley CJ, Gaffan D, Duncan J (2002) Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nat Neurosci* 5:671–676. [CrossRef Medline](#)
- Everling S, Tinsley CJ, Gaffan D, Duncan J (2006) Selective representation of task-relevant objects and locations in the monkey prefrontal cortex. *Eur J Neurosci* 23:2197–2214. [CrossRef Medline](#)
- Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci U S A* 110:16616–16621. [CrossRef Medline](#)
- Fitzgerald JK, Swaminathan SK, Freedman DJ (2012) Visual categorization and the parietal cortex. *Front Integr Neurosci* 6:18. [Medline](#)
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291:312–316. [CrossRef Medline](#)
- Harel A, Kravitz DJ, Baker CI (2014) Task context impacts visual object processing differentially across the cortex. *Proc Natl Acad Sci U S A* 111: E962–E971. [CrossRef Medline](#)
- Jehee JF, Brady DK, Tong F (2011) Attention improves encoding of task-relevant features in the human visual cortex. *J Neurosci* 31:8210–8219. [CrossRef Medline](#)
- Jiang X, Bradley E, Rini RA, Zeffiro T, Vanmeter J, Riesenhuber M (2007) Categorization training results in shape- and category-selective human neural plasticity. *Neuron* 53:891–903. [CrossRef Medline](#)
- Kadohisa M, Petrov P, Stokes M, Sigala N, Buckley M, Gaffan D, Kusunoki M, Duncan J (2013) Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron* 80:235–246. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141. [CrossRef Medline](#)
- Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2010) Target detection by opponent coding in monkey prefrontal cortex. *J Cogn Neurosci* 22: 751–760. [CrossRef Medline](#)
- Lee SH, Kravitz DJ, Baker CI (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci* 16:997–999. [CrossRef Medline](#)
- Li S, Ostwald D, Giese M, Kourtzi Z (2007) Flexible coding for categorical decisions in the human brain. *J Neurosci* 27:12321–12330. [CrossRef Medline](#)
- Liu T, Hospadaruk L, Zhu DC, Gardner JL (2011) Feature-specific attentional priority signals in human cortex. *J Neurosci* 31:4484–4495. [CrossRef Medline](#)
- Luria AR (1966) Higher cortical functions in man. London: Tavistock.
- MacEvoy SP, Epstein RA (2011) Constructing scenes from objects in human occipitotemporal cortex. *Nat Neurosci* 14:1323–1329. [CrossRef Medline](#)
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139. [CrossRef Medline](#)
- Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T (2008) Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J Neurophysiol* 100:1407–1419. [CrossRef Medline](#)
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202. [CrossRef Medline](#)
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16: 5154–5167. [Medline](#)
- Milner B (1963) Effects of different brain lesions on card sorting. *Arch Neurol* 9:90–100. [CrossRef](#)
- Norman DA, Shallice T (1980) Attention to action: willed and automatic control of behavior. San Diego: University of California Center for Human Information Processing.
- Poser BA, Koopmans PJ, Witzel T, Wald LL, Barth M (2010) Three dimensional echo-planar imaging at 7 Tesla. *Neuroimage* 51:261–266. [CrossRef Medline](#)
- Reddy L, Kanwisher N (2007) Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol* 17: 2067–2072. [CrossRef Medline](#)
- Reddy L, Kanwisher NG, VanRullen R (2009) Attention and biased competition in multi-voxel object representations. *Proc Natl Acad Sci U S A* 106:21447–21452. [CrossRef Medline](#)
- Roca M, Parr A, Thompson R, Woolgar A, Torralva T, Antoun N, Manes F, Duncan J (2010) Executive function and fluid intelligence after frontal lobe lesions. *Brain* 133:234–247. [CrossRef Medline](#)
- Sadaghiani S, D'Esposito M (2014) Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cereb Cortex*. In press.
- Sakagami M, Niki H (1994) Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: relation to relevant task conditions. *Exp Brain Res* 97:423–436. [Medline](#)
- Salazar RF, Dotson NM, Bressler SL, Gray CM (2012) Content-specific

- fronto-parietal synchronization during visual working memory. *Science* 338:1097–1100. [CrossRef Medline](#)
- Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *Neuroimage* 44:223–231. [CrossRef Medline](#)
- Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78:364–375. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Swaminathan SK, Freedman DJ (2012) Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nat Neurosci* 15:315–320. [CrossRef Medline](#)
- Todd MT, Nystrom LE, Cohen JD (2013) Confounds in multivariate pattern analysis: Theory and rule representation case study. *Neuroimage* 77:157–165. [CrossRef Medline](#)
- van der Kouwe AJ, Benner T, Salat DH, Fischl B (2008) Brain morphology with multiecho MPRAGE. *Neuroimage* 40:559–569. [CrossRef Medline](#)
- Watanabe M (1986) Prefrontal unit activity during delayed conditional Go/No-Go discrimination in the monkey. II. Relation to Go and No-Go responses. *Brain Res* 382:15–27. [CrossRef Medline](#)
- Woolgar A, Parr A, Cusack R, Thompson R, Nimmo-Smith I, Torralva T, Roca M, Antoun N, Manes F, Duncan J (2010) Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proc Natl Acad Sci U S A* 107:14899–14902. [CrossRef Medline](#)
- Woolgar A, Thompson R, Bor D, Duncan J (2011a) Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *Neuroimage* 56:744–752. [CrossRef Medline](#)
- Woolgar A, Hampshire A, Thompson R, Duncan J (2011b) Adaptive coding of task-relevant information in human frontoparietal cortex. *J Neurosci* 31:14592–14599. [CrossRef Medline](#)
- Woolgar A, Golland P, Bode S (2014) Coping with confounds in multivoxel pattern analysis: what should we do about reaction time differences? A comment on Todd, Nystrom, and Cohen, 2013. *Neuroimage* 98:506–512. [CrossRef Medline](#)
- Woolgar A, Williams MA, Rich AN (2015) Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *Neuroimage* 109:429–437. [CrossRef Medline](#)