RESEARCH ARTICLE | *Higher Neural Functions and Behavior*

Human frontoparietal cortex represents behaviorally relevant target status based on abstract object features

 Margaret Henderson[1] and John T. Serences[1,2,3]

[1]*Neurosciences Graduate Program, University of California, San Diego, La Jolla, California;* [2]*Department of Psychology, University of California, San Diego, La Jolla, California; and* [3]*Kavli Foundation for the Brain and Mind, University of California, San Diego, La Jolla, California*

Henderson M, Serences JT. Human frontoparietal cortex represents behaviorally relevant target status based on abstract object features. *J Neurophysiol* 121: 1410–1427, 2019. First published February 13, 2019; doi:10.1152/jn.00015.2019.—Searching for items that are useful given current goals, or "target" recognition, requires observers to flexibly attend to certain object properties at the expense of others. This could involve focusing on the identity of an object while ignoring identity-preserving transformations such as changes in viewpoint or focusing on its current viewpoint while ignoring its identity. To effectively filter out variation due to the irrelevant dimension, performing either type of task is likely to require high-level, abstract search templates. Past work has found target recognition signals in areas of ventral visual cortex and in subregions of parietal and frontal cortex. However, target status in these tasks is typically associated with the identity of an object, rather than identity-orthogonal properties such as object viewpoint. In this study, we used a task that required subjects to identify novel object stimuli as targets according to either identity or viewpoint, each of which was not predictable from low-level properties such as shape. We performed functional MRI in human subjects of both sexes and measured the strength of target-match signals in areas of visual, parietal, and frontal cortex. Our multivariate analyses suggest that the multiple-demand (MD) network, including subregions of parietal and frontal cortex, encodes information about an object's status as a target in the relevant dimension only, across changes in the irrelevant dimension. Furthermore, there was more target-related information in MD regions on correct compared with incorrect trials, suggesting a strong link between MD target signals and behavior.

NEW & NOTEWORTHY Real-world target detection tasks, such as searching for a car in a crowded parking lot, require both flexibility and abstraction. We investigated the neural basis of these abilities using a task that required invariant representations of either object identity or viewpoint. Multivariate decoding analyses of our whole brain functional MRI data reveal that invariant target representations are most pronounced in frontal and parietal regions, and the strength of these representations is associated with behavioral performance.

fMRI; frontoparietal; invariance; object; target recognition

INTRODUCTION

To flexibly guide behavior, humans can choose to hold in mind information about the identity of sought-after items or about the current state of those items. For example, when searching the parking lot at the end of a long day, you might search for the presence of your blue sedan, but when crossing the street, you might search for any car moving quickly in the rightward direction. The former task is challenging because the retinal projection of your car can have considerable variability due to changes in pose, position, and environmental conditions, whereas the latter task is challenging because relevant cars may be a variety of makes, models, sizes, and colors (DiCarlo and Cox 2007; Ito et al. 1995; Lueschow et al. 1994; Marr and Nishihara 1978; Tanaka 1993). To overcome both types of challenges, recognizing relevant targets under realistic viewing conditions is likely to require high-level, abstract search templates (Biederman 2000; Freiwald and Tsao 2010; Riesenhuber and Poggio 2000; Tarr et al. 1998).

Such abstract search templates have been found in multiple brain regions. In inferotemporal (IT) cortex, neurons encode object identity across identity-preserving transformations, even during passive viewing (Anzellotti et al. 2014; Erez et al. 2016; Freiwald and Tsao 2010; Tanaka 1996). Neurons in IT and entorhinal cortex (ERC) also signal the target status of objects, both when targets can be identified on the basis of an exact match of retinal input and when targets have to be identified across changes in size and position (Lueschow et al. 1994; Miller and Desimone 1994; Pagan et al. 2013; Roth and Rust 2018; Woloszyn and Sheinberg 2009). In addition to these ventral regions, neurons in prefrontal cortex (PFC) have been shown to signal the target status of objects, both for superordinate and subordinate identification tasks (Freedman et al. 2003; Kadohisa et al. 2013; McKee et al. 2014; Miller et al. 1996).

In agreement with the single-unit modulations in prefrontal cortex, recent studies in humans suggest that a set of frontal and parietal regions, collectively referred to as the multiple-demand (MD) network, may play a role in target selection by representing objects according to their task-relevant properties (Bracci et al. 2017; Duncan 2010; Fedorenko et al. 2013; Jackson et al. 2017; Vaziri-Pashkam and Xu 2017). Accordingly, MD representations have also been found to differentiate images on the basis of their status as a target object or category, and these representations exhibit invariance across changes in low-level image properties (Erez and Duncan 2015; Guo et al. 2012). Target representations in frontoparietal regions are also

Address for reprint requests and other correspondence: M. Henderson, Neurosciences Graduate Program, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA, 92093 (e-mail: mmhender@ucsd.edu).

associated with decision confidence and task difficulty, suggesting that they play a role in shaping decisions (Guo et al. 2012). However, in all past studies, sought targets were defined according to the identity or category of the object, dimensions whose representation in the visual system has been extensively characterized (Conway 2018; DiCarlo and Cox 2007; Grill-Spector 2003; Tanaka 1996). Less well studied is how the visual system computes matches in dimensions orthogonal to object identity, such as object pose or viewpoint. Past work has demonstrated sensitivity to object viewpoint in multiple regions of the human and primate brain, including IT cortex and the intraparietal sulcus (IPS), but it is not yet clear how viewpoint representations are involved in identification of relevant targets (Andresen et al. 2009; Grill-Spector et al. 1999; Hong et al. 2016; Tanaka 1993; Valyear et al. 2006; Ward et al. 2018). Thus, although we predict that representations of target status based on object viewpoint will be found in the same frontoparietal regions known to encode target status based on identity and category of objects, this has yet to be shown.

In the present study we tested the hypothesis that regions of the MD network support performance during a task where target status is defined on the basis of either object identity or object viewpoint. We generated a novel object stimulus set (Fig. 1), in which three-dimensional (3D) objects of multiple identities were rendered at multiple viewpoints. Subjects determined the target status of objects according to either their identity (identity task) or viewpoint (viewpoint task) while ignoring the other dimension. Critically, this paradigm required subjects to form viewpoint-invariant representations of identity and identity-invariant representations of viewpoint, both of which were defined so as not to be predictable from the retinotopic shape of an object. We used multivariate pattern analyses (MVPA) on single-trial voxel activation patterns to decode the status of each image as a target in both the task-relevant and the task-irrelevant dimensions. Our findings suggest that whereas ventral visual cortex exhibits some sensitivity to an object's status as a target, regions of the MD network encode robust, abstract target representations that are sensitive to changes in task demands and that are selectively linked with behavioral performance.

MATERIALS AND METHODS

Participants. Ten subjects (3 men) between the ages of 20 and 34 yr were recruited from the University of California, San Diego (UCSD) community (mean age 24.7 ± 4.7 yr), having normal or
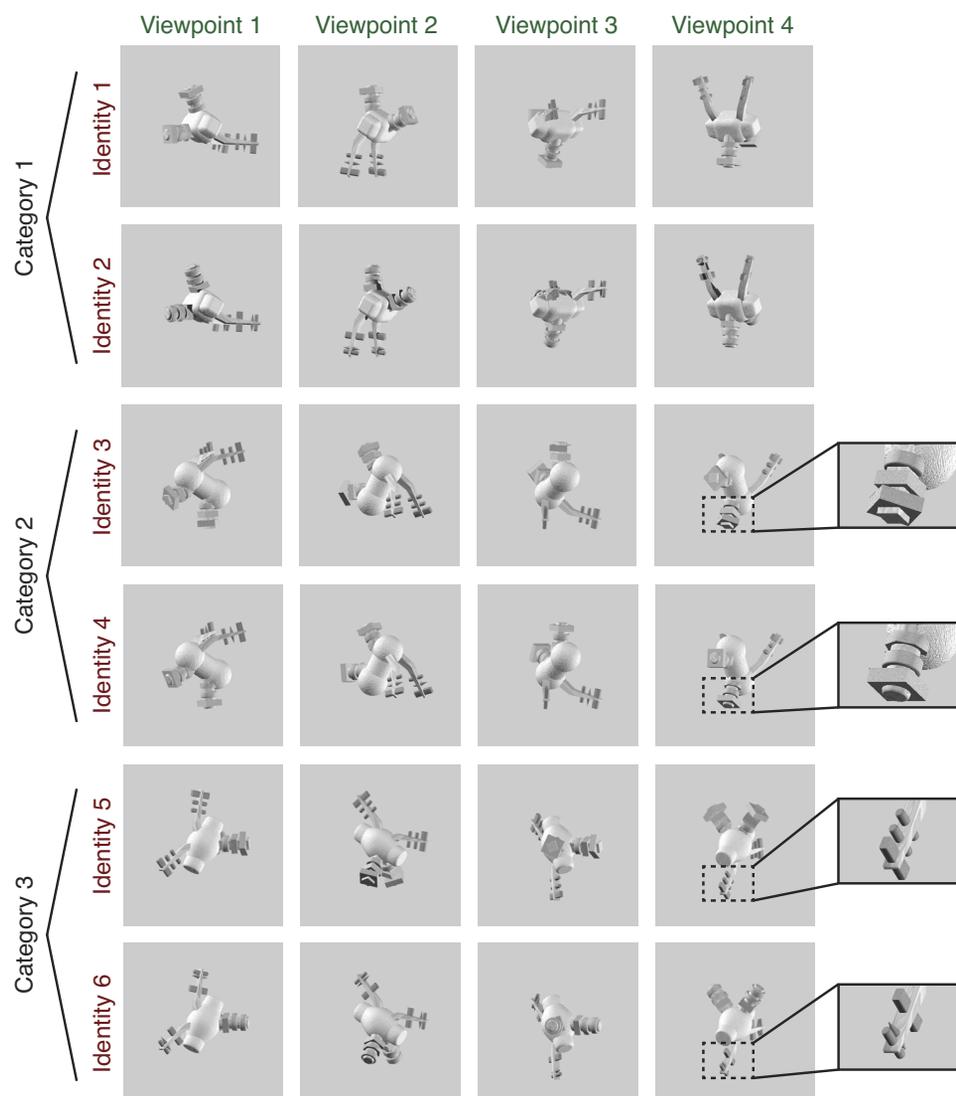

Fig. 1. Example set of images shown to a subject during scanning, consisting of 6 unique object identities, each rendered at 4 viewpoints. Subjects were instructed either to match the exact identity of the object irrespective of viewpoint (shown in rows of the matrix) or to match the viewpoint of the object irrespective of identity (columns of the matrix). The 6 identities comprised 2 exemplars in each of 3 categories, with categories defined by overall body shape and exemplars defined by details of the peripheral features (see *insets* for examples of differentiating features). Object viewpoint was generated in an arbitrarily defined coordinate system so that low-level visual features had a minimal contribution to the viewpoint matching task (see METHODS for details). Two complete sets of novel objects were generated, with half the subjects (5/10) viewing *set A* and half viewing *set B*. The images shown are from object *set B*; see Fig. 2 for examples of object *set A*.

corrected-to-normal vision. The study protocol was submitted to and approved by the Institutional Review Board at UCSD, and all participants provided written informed consent. Each subject performed a behavioral training session lasting ~1 h, followed by one or two scan sessions, each lasting ~2 h. Participants were compensated at a rate of $10/h for behavioral training and $20/h for the scanning sessions.

Novel object sets and one-back tasks. All objects were generated and rendered using Strata 3D CX software (version 7.6; Santa Clara, UT). To ensure a variety of stimuli, we generated two unique sets of objects and assigned half of our subjects (5 of 10) to object *set A* and half to object *set B* (because we did not observe a difference in performance between the 2 stimulus sets for either task, we combined our analyses across all subjects). Each stimulus set comprised 3 categories of objects, each with 36 total exemplars. Objects within a category shared a common body plan, including the shape of the main body and the configuration of peripheral features around the body (see Figs. 1 and 2). Exemplars in each category were differentiated by small variations in the details of peripheral features, such as the size or shape of a spike. These peripheral features always appeared in pairs that were attached symmetrically to the body, making the overall objects bilaterally symmetric. Feature details were always matched within each peripheral feature pair, ensuring that even when one feature in a pair was occluded at a particular viewpoint, the details could always be discerned from the other feature in the pair. During scanning, each subject viewed two exemplars in each category (se-

lected on an individual subject basis from the full set of 36 exemplars; see below for details), giving a total of six object "identities" (2 exemplars in each of 3 categories). Each of the six object identities was rendered from four different viewpoints, for a total of 24 unique images shown to each subject.

While in the scanner, each subject performed two different one-back tasks (identity task and viewpoint task). In the identity task, subjects responded to each image on the basis of whether it matched the identity of the immediately preceding image. Identity matches had to be the same exemplar from the same category but did not have to match in viewpoint. In the viewpoint task, subjects responded on the basis of whether the current image matched the viewpoint of the immediately preceding image, whereas both category and exemplar status were irrelevant. In order for subjects to identify matches in viewpoint between objects with different identities, they were trained to recognize an arbitrary viewpoint of each object as its "frontal" viewpoint, or a rotation of [0°, 0°] about the *Y* (vertical)- and *Z* (front-back)-axes. All other viewpoints were then defined relative to this reference point. The four viewpoints used, in coordinates of [*Z* rotation, *Y* rotation], were [30°, 300°], [120°, 240°], [210°, 30°], and [300°, 150°]. We defined viewpoint in this way to ensure that subjects formed a representation of viewpoint that was largely invariant to 2D shape. Importantly, because the frontal viewpoint was chosen arbitrarily, it was not systematically predictable from the overall axis of elongation of the main body. As a result, images of the objects in
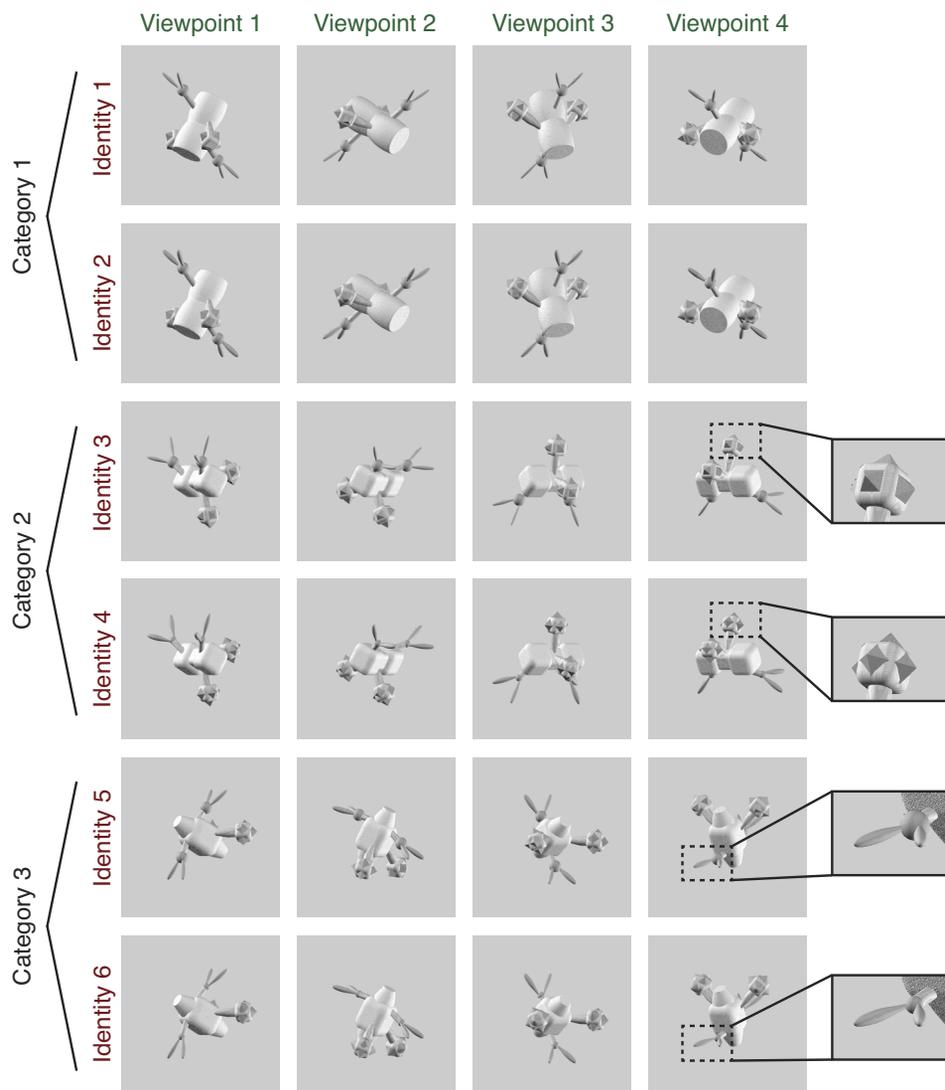

Fig. 2. Example stimulus set from object *set A* (see Fig. 1 and METHODS for details).

different categories at the same viewpoint differed dramatically in the shape they produced when projected onto a 2D plane.

Prescan training. Subjects were familiarized with the novel object viewpoints during a self-guided session performed before scanning. During the first half of the training session, subjects viewed a 3D model of each of the three novel object categories and were able to rotate the model around two axes using the arrow keys on a keyboard (each key press gave a rotation of 30° about either the *Y*- or *Z*-axis). During this entire exercise, the angular position of the object, expressed using the format "*X* rotation = *m* degrees, *Y* rotation = *n* degrees," was displayed at the top of the screen. Subjects were encouraged to use the angular coordinates to learn how each view of the object was defined relative to the arbitrarily chosen frontal ([0°, 0°]) position. During the second half of this training, subjects were presented with images of all three categories simultaneously, at matching viewpoints, and encouraged to study how the same viewpoint was defined across categories. Subjects performed both parts of this training at least once and were allowed to return to it as many times as they wished. On average, subjects spent ~20 min on the self-guided training. In addition to the self-guided training, subjects performed several practice runs of the viewpoint task. During these runs, the four object viewpoints that were used during scanning were never used so that even though subjects were familiarized with different viewpoints of each object category, they were not overexposed to the target viewpoints. After each viewpoint task practice run, subjects could return to the self-guided viewpoint training, and they repeated as many iterations of self-guided training and practice runs as were necessary to reach 70% performance (between 4 and 10 runs across subjects).

The six object identities viewed by each subject were selected on the basis of a behavioral thresholding experiment performed before scanning. This allowed us to control the difficulty of the identity task by manipulating the similarity between the exemplars in each category. The task used for thresholding was identical to the one-back identity task used during scanning, but it used only objects from a single category, presented at four random viewpoints. Subjects performed six runs of this task, with two runs for each object category. On the basis of their performance, we selected two exemplars in each object category that were confusable ~70% of the time. Following this thresholding procedure, subjects performed two practice runs of the identity task, using the final set of exemplars that they would view during scanning.

Immediately before each scanning session, subjects performed another short self-guided training exercise (~5 min), in which they were shown examples of the exact images they would see during scanning. First, they were presented with the two exemplars in each object category, side by side, and allowed to freely rotate the objects using the arrow keys to compare the appearance of the two exemplars from many viewpoints. Next, they were presented with images from each category, side by side, at each of the four viewpoints they would see during the task and encouraged to use this information to prepare for the viewpoint task.

Behavioral task in the scanner. During scanning, subjects performed between 8 and 11 runs each of the identity and viewpoint tasks. Task runs always occurred in pairs with the identity task followed by the viewpoint task. An identical sequence of visual stimuli was presented on both runs in each pair so that visual stimulation was perfectly matched between conditions. Each 6-min run consisted of 48 trials, and each trial consisted of a single image presentation for 1,500 ms, followed by a jittered intertrial interval ranging from 2,000 to 6,000 ms. Each of the 24 images was shown twice per run. The sequence of image presentations was pseudorandomly generated, with the constraint that there was a 0.50 probability that the current stimulus was from the same category as the previous stimulus (within-category trials). This constraint was adopted to more closely equate the difficulty of the two tasks, because the viewpoint task was inherently more difficult to solve on across-category trials,

and the identity task was more difficult to solve on within-category trials. This resulted in a probability of 0.23 of any trial being a match in either viewpoint or identity, and a probability of 0.04 of a match in both dimensions.

In both tasks, subjects responded to every image by pressing a button using either their index finger ("1") or their middle finger ("2"), depending on the current response mapping rule. Response mapping rules were counterbalanced within each subject so that on half of the runs the subject responded with 1 for "match" and 2 for "non-match," and on the other half of runs they responded with 1 for "non-match" and 2 for "match." The purpose of these different response mapping rules was to ensure that match-related information was not confounded with motor responses.

Magnetic resonance imaging. All MRI scanning was performed on a General Electric (GE) Discovery MR750 3.0-T research-dedicated scanner at the UC San Diego Keck Center for Functional Magnetic Resonance Imaging. Functional echo-planar imaging (EPI) data were acquired using a Nova Medical 32-channel head coil (NMSC075-32-3GE-MR750) and the Stanford Simultaneous Multi-Slice (SMS) EPI sequence (MUX EPI), with a multiband factor of 8 and 9 axial slices per band (total slices = 72; 2-mm^3 isotropic; 0-mm gap; matrix = 104 × 104; field of view = 20.8 cm; TR/TE = 800/35 ms; flip angle = 52°; in-plane acceleration = 1). Image reconstruction procedures and unaliasing procedures were performed on local servers using reconstruction code from CNI (Center for Neural Imaging at Stanford). The initial 16 repetition times (TRs) collected at sequence onset served as reference images required for the transformation from k-space to the image space. Two short (17 s) "topup" data sets were collected during each session, using forward and reverse phase-encoding directions. These images were used to estimate susceptibility-induced off-resonance fields (Andersson et al. 2003) and to correct signal distortion in EPI sequences using FSL (FMRIB Software Library) topup functionality (Jenkinson et al. 2012).

During each functional session, we also acquired an accelerated anatomical scan using parallel imaging [GE ASSET on a fast spoiled gradient-echo (FSPGR) T1-weighted sequence; 1 × 1 × 1-mm^3 voxel size; TR = 8,136 ms; TE = 3,172 ms; flip angle = 8°; 172 slices; 1-mm slice gap; 256 × 192-cm matrix size] using the same 32-channel head coil. We also acquired one additional high-resolution anatomical scan for each subject (1 × 1 × 1-mm^3 voxel size; TR = 8,136 ms; TE = 3,172 ms; flip angle = 8°; 172 slices; 1-mm slice gap; 256 × 192-cm matrix size) during a separate retinotopic mapping session using an Invivo eight-channel head coil. This scan produced higher quality contrast between gray and white matter and was used for segmentation, flattening, and visualizing retinotopic mapping data.

In addition to the multiband scan protocol described above, five subjects participated in retinotopic mapping experiments using a different scan protocol, previously reported (Sprague and Serences 2013). The remaining five subjects participated in retinotopic mapping runs using the multiband protocol described above. Where possible, the data used to generate retinotopic maps (see *Retinotopic mapping stimulus protocol*) were combined across these sessions.

Preprocessing. First, the structural scan from each session was processed in BrainVoyager 2.6.1 to align the anatomical and the functional data sets. Automatic algorithms were used to adjust the structural image intensity to correct for inhomogeneities, as well as to remove the head and skull tissue. Structural scans were then aligned to the anterior commissure-posterior commissure (AC-PC) plane using manual landmark identification. Finally, an automatic registration algorithm was used to align the structural scan to the high-definition structural scan collected during each subject's retinotopic mapping session. This high-definition structural scan was transformed into Talairach space, and the parameters of this transformation were used to transform all other scans for this subject into Talairach space.

Next, each functional run was aligned to the same-session structural scan. We then used BrainVoyager 2.6.1 to perform slice-time correction, affine motion correction, and temporal high-pass filtering

to remove first-, second-, and third-order signal drifts over the course of each functional run. These data were spatially transformed into Talairach space to align with the anatomical images. Finally, the blood oxygen level-dependent (BOLD) signal in each voxel was z-transformed within each run.

General linear model to estimate trial-by-trial responses. After preprocessing, single-trial activation estimates (beta weights), which were used for subsequent MVPA, were obtained using a general linear model (GLM) with a design matrix created by convolving the trial sequence with the canonical two-gamma hemodynamic response function (HRF) as implemented in BrainVoyager (peak at 5 s, undershoot peak at 15 s, response undershoot ratio 6, response dispersion 1, undershoot dispersion 1). Throughout this study, the same HRF parameters were used for all GLM analyses.

Retinotopic mapping stimulus protocol. We followed previously published retinotopic mapping protocols to define the visual areas V1, V2, V3, V3AB, V4, IPS0–1, and IPS2–3 (Engel et al. 1997; Jerde and Curtis 2013; Sereno et al. 1995; Swisher et al. 2007; Wandell et al. 2007; Winawer and Witthoft 2015). Subjects performed mapping runs in which they viewed a contrast-reversing checkerboard stimulus (4 Hz) configured as a rotating wedge (10 cycles, 36 s/cycle), an expanding ring (10 cycles, 36 s/cycle), or a bowtie (8 cycles, 40 s/cycle). To increase the quality of data from parietal regions, subjects performed a covert attention task on the rotating wedge stimulus, which required them to detect contrast-dimming events that occurred occasionally (on average, 1 event occurred every 7.5 s) in a row of the checkerboard (mean accuracy = $61.8 \pm 13.9\%$). This stimulus was limited to a $22° \times 22°$ field of view.

Multiple-demand localizer. To define regions of interest (ROIs) in the MD network, we used an independent functional localizer to identify voxels whose BOLD response was significantly modulated by the load of a spatial working memory task (Duncan 2010; Fedorenko et al. 2013). Subjects performed one or two runs of this task during each functional scanning session. During each trial of this task, subjects were first presented with an empty rectangular grid comprising either 8 or 16 squares. Half of the squares in the grid were then highlighted one at a time, and subjects were required to remember the locations of the highlighted squares. Subjects were then shown a probe grid and asked to report whether the highlighted squares matched the remembered locations. Runs were divided into blocks with either high or low load. Performance was significantly poorer on high-load blocks (mean d' for low load = 2.48 ± 0.28, mean d' for high load = 1.04 ± 0.21, $P < 0.001$; paired 2-tailed t-test).

We used the data from these runs to generate a statistical parametric map for each subject, which expressed the degree to which each voxel showed elevated BOLD signal for high-load vs. low-load working memory blocks. We defined a GLM with a regressor for each block type and solved for the β-coefficients corresponding to each load condition. Coefficients were then entered into a one-tailed, repeated-measures t-test against a distribution with a mean of 0 [false discovery rate (FDR)-corrected $q = 0.05$]. This resulted in a single mask of load-selective voxels for each subject.

To subdivide this mask into the typical MD ROIs, we used a group-level parcellation from a previously published data set (Fedorenko et al. 2013). We used this parcellation to generate masks for five ROIs of interest: the intraparietal sulcus (IPS), the superior precentral sulcus (sPCS), the inferior precentral sulcus (iPCS), the anterior insula/frontal operculum (AI/FO), and the inferior frontal sulcus (IFS). Because we had already defined two posterior subregions of the IPS, IPS0–1 and IPS2–3 (see *Retinotopic mapping stimulus protocol*), we removed all voxels belonging to these retinotopic regions from the larger IPS mask and used the remaining voxels to define a region that we refer to as the superior IPS (sIPS). This was done within each subject separately. We intersected each subject's mask of load-selective voxels with the mask for each ROI to generate the final MD ROI definitions.

For one subject, this procedure failed to yield any voxels in the sPCS ROI. Therefore, when performing group-level ANOVA of decoding performance, we performed linear interpolation (Roth 1994) based on sPCS responses in the remaining nine subjects to generate an estimate of the missing value. We did this by calculating a t-score comparing the missing subject's d' score with those of the other nine subjects in each ROI and condition where it was defined and using the mean of these t-scores to estimate d' in sPCS for each condition. As an alternative to this interpolation method, we also ran the repeated-measures ANOVA with all values for the missing subject removed: we observed similar results (see Table 2). For all ANOVA results reported in this paper, Mauchly's test revealed that the data did not violate the assumption of sphericity, so we report uncorrected P values.

Lateral occipital complex localizer. We identified two subregions of the lateral occipital complex, LO and pFus, using a functional localizer developed by the Stanford Vision and Perception Laboratory (Stigliani et al. 2015) to identify voxels that showed enhanced responses to intact objects (cars and guitars) vs. phase-scrambled versions of the same images. Between two and four runs of this task were performed during functional scanning sessions, with each run lasting 5 min and 16 s. During each run, subjects viewed blocks of sequentially presented images in a particular category (cars, guitars, faces, houses, body parts, scrambled objects) and performed a one-back repeat detection task (mean $d' = 3.09 \pm 0.49$). We used a GLM to define voxels that showed significantly higher BOLD responses during car/guitar blocks vs. scrambled blocks (FDR-corrected $q = 0.05$). We then projected this mask onto a computationally inflated mesh of the gray matter-white matter boundary in each subject and defined LO and pFus on this mesh, based on the mask in conjunction with anatomical landmarks (Vinberg and Grill-Spector 2008).

Object localizer task. After the ROIs described above were defined, voxels within each ROI were thresholded on the basis of their visual responsiveness during performance on an independent novel object matching task. Subjects performed two to three runs of this task during each scanning session. This task was identical to the identity task described above, except that the alternate object set was used (e.g., if the subject viewed *set A* during the main one-back task runs, they viewed *set B* during the localizer). The object exemplars shown during this task were randomly selected for each run. Performance on this task was consistently lower than performance on the main one-back tasks, due to the fact that subjects had not been trained on this stimulus set (mean $d' = -0.23 \pm 0.14$).

For each subject, we combined data from all object localizer runs to generate a statistical parametric map of voxel responsiveness, based on a GLM in which all image presentations were modeled as a single predictor. We then selected only the voxels whose BOLD signal was significantly modulated by image presentation events (FDR-corrected $q = 0.05$). This limited the voxels selected in each ROI to those that were responsive to object stimuli that were visually similar (but not identical) to those presented during the main task. For one ROI in the MD network (AI/FO), this thresholding procedure yielded fewer than 10 voxels for several subjects, so for this ROI we chose to analyze all the voxels that were defined by the MD localizer. The final definitions of each ROI (centers and sizes) following this thresholding procedure are summarized in Table 1.

MVPA decoding. The goal of our MVPA analysis was to estimate the amount of linearly decodable information about object "match" status in each task dimension (identity and viewpoint) that was represented in each ROI during each task. Because match status depended on the relation of each object to the previous one in the sequence, each object had an equal probability of appearing as a match or a nonmatch in each dimension, so this decoding was orthogonal to the visual properties of the objects. To evaluate the behavioral relevance of information about match/nonmatch status in each ROI, we also assessed how match decoding was affected by the

Table 1. *Centers and sizes of the final ROIs defined for each subject, following functional localization and additional thresholding with a novel object localizer*

Subject	Center LH	Center RH	No. of Voxels LH	No. of Voxels RH
	V1			
1	[−16, −94, −12]	[12, −92, −17]	331	368
2	[−9, −82, 8]	[6, −82, 6]	797	444
3	[−15, −97, 7]	[11, −97, 5]	84	192
4	[−12, −97, −8]	[14, −94, 4]	249	186
5	[−6, −95, −4]	[16, −95, 1]	273	164
6	[−8, −90, −12]	[13, −89, −1]	269	434
7	[−9, −88, −5]	[11, −92, −5]	406	393
8	[−11, −92, −8]	[10, −91, −4]	348	254
9	[−12, −93, −8]	[15, −96, −5]	477	324
10	[−16, −89, −18]	[12, −92, −9]	299	239
	V2			
1	[−23, −92, −11]	[14, −90, −15]	325	311
2	[−14, −86, 7]	[19, −82, 5]	495	602
3	[−17, −92, 4]	[15, −93, 3]	271	226
4	[−13, −93, −7]	[18, −92, 4]	354	212
5	[−10, −94, −8]	[19, −93, 5]	179	167
6	[−13, −85, −13]	[18, −89, −1]	272	388
7	[−14, −93, −6]	[15, −93, −5]	363	419
8	[−14, −87, −12]	[16, −89, −8]	416	308
9	[−13, −92, −6]	[14, −90, −6]	277	300
10	[−20, −88, −20]	[17, −90, −7]	370	364
	V3			
1	[−25, −85, −8]	[22, −88, −9]	331	360
2	[−22, −83, 1]	[12, −88, 6]	500	515
3	[−19, −87, 6]	[21, −89, 3]	247	320
4	[−20, −90, −3]	[22, −85, 6]	359	216
5	[−20, −88, −7]	[29, −85, 1]	382	292
6	[−23, −85, −9]	[27, −84, −0]	259	474
7	[−24, −90, −3]	[23, −88, −5]	646	703
8	[−22, −83, −16]	[25, −88, −7]	432	348
9	[−18, −89, −3]	[22, −86, −6]	442	318
10	[−23, −87, −16]	[24, −84, −8]	366	630
	V3AB			
1	[−26, −84, 7]	[28, −83, 7]	136	200
2	[−28, −85, 19]	[23, −80, 30]	425	253
3	[−25, −84, 21]	[28, −85, 22]	269	214
4	[−26, −83, 11]	[26, −77, 19]	141	167
5	[−26, −87, 17]	[34, −79, 19]	281	396
6	[−31, −86, 12]	[31, −78, 20]	120	137
7	[−29, −83, 12]	[24, −78, 15]	357	306
8	[−36, −84, 5]	[36, −83, 1]	350	474
9	[−24, −89, 18]	[30, −83, 10]	341	329
10	[−25, −91, −2]	[28, −78, 18]	450	311
	V4			
1	[−27, −75, −16]	[26, −75, −14]	116	245
2	[−29, −76, −8]	[25, −76, −8]	336	299
3	[−29, −78, −7]	[22, −78, −9]	303	154
4	[−22, −80, −13]	[25, −75, −9]	75	158
5	[−27, −79, −16]	[32, −75, −12]	243	139
6	[−29, −75, −17]	[32, −72, −14]	115	272
7	[−29, −76, −14]	[29, −76, −14]	406	380
8	[−31, −75, −23]	[30, −83, −18]	261	164
9	[−24, −76, −18]	[26, −77, −16]	304	362
10	[−32, −76, −22]	[24, −75, −17]	243	329
	LO			
1	[−43, −76, −13]	[37, −78, −10]	617	798
2	[−40, −77, 3]	[37, −78, 5]	911	1113
3	[−38, −81, 5]	[35, −81, 1]	253	314
4	[−33, −87, 1]	[34, −80, 0]	163	177
5	[−35, −84, 4]	[38, −81, 5]	207	132

Continued

Table 1. *Continued*

Subject	Center LH	Center RH	No. of Voxels LH	No. of Voxels RH
6	[−40, −78, −7]	[40, −73, −3]	933	724
7	[−35, −88, −4]	[32, −82, −1]	372	449
8	[−41, −73, −13]	[38, −76, −14]	812	347
9	[−34, −80, −5]	[37, −78, −7]	792	634
10	[−47, −76, −14]	[35, −81, 2]	119	77
	pFus			
1	[−34, −75, −16]	[34, −51, −15]	215	54
2	[−37, −63, −10]	[36, −62, −9]	354	415
3	[−35, −69, −8]	[34, −70, −9]	233	98
4	[−37, −80, −10]	[36, −67, −8]	252	341
5	[−43, −66, −8]	[40, −64, −10]	201	127
6	[−37, −63, −16]	[33, −55, −15]	234	268
7	[−38, −68, −15]	[34, −61, −17]	87	184
8	[−37, −54, −21]	[32, −62, −20]	391	656
9	[−33, −63, −12]	[35, −61, −15]	96	447
10	[−45, −67, −12]	[39, −61, −15]	207	117
	IPS0−1			
1	[−26, −80, 15]	[25, −79, 21]	162	236
2	[−31, −78, 27]	[22, −61, 40]	369	960
3	[−22, −70, 32]	[24, −75, 33]	331	280
4	[−22, −70, 28]	[25, −67, 33]	185	175
5	[−25, −70, 26]	[30, −67, 30]	524	397
6	[−30, −67, 26]	[22, −64, 40]	30	269
7	[−25, −76, 25]	[24, −73, 28]	478	586
8	[−34, −74, 15]	[32, −71, 13]	401	375
9	[−24, −75, 24]	[30, −74, 16]	675	482
10	[−27, −88, 9]	[26, −67, 33]	473	499
	IPS2−3			
1	[−29, −66, 32]	[25, −72, 39]	140	126
2	[−25, −63, 35]	[24, −46, 51]	488	491
3	[−25, −54, 46]	[22, −61, 44]	283	203
4	[−24, −62, 45]	[24, −60, 39]	135	47
5	[−27, −63, 43]	[22, −63, 43]	368	270
6	[−26, −62, 35]	[24, −54, 45]	221	175
7	[−22, −71, 39]	[24, −61, 37]	395	352
8	[−26, −68, 27]	[27, −63, 23]	214	161
9	[−29, −62, 31]	[25, −63, 29]	283	388
10	[−27, −71, 23]	[25, −55, 45]	390	510
	sIPS			
1	[−34, −55, 36]	[34, −54, 37]	65	204
2	[−34, −40, 42]	[31, −47, 50]	155	283
3	[−31, −50, 43]	[26, −60, 45]	203	145
4	[−24, −62, 43]	[29, −55, 44]	190	559
5	[−32, −53, 47]	[34, −50, 45]	472	708
6	[−35, −48, 43]	[27, −51, 48]	527	434
7	[−35, −54, 41]	[33, −52, 44]	335	919
8	[−28, −60, 45]	[28, −59, 41]	540	654
9	[−34, −53, 42]	[32, −54, 40]	486	805
10	[−28, −55, 41]	[31, −51, 44]	1057	665
	sPCS			
1				
2	[−22, 4, 60]	[30, 2, 54]	4	28
3	[−25, −4, 54]	[29, 0, 53]	35	19
4	[−31, −4, 50]	[36, −3, 49]	5	52
5	[−29, −4, 57]	[29, −4, 52]	56	142
6	[−33, 2, 54]	[36, 2, 55]	152	172
7	[−28, −5, 54]	[29, −1, 55]	19	348
8	[−37, −1, 56]	[29, 1, 53]	19	349
9	[−27, −3, 52]	[27, −3, 50]	189	209
10	[−25, −3, 56]	[26, 2, 55]	95	109

Continued

Table 1.　*Continued*

Subject	Center		No. of Voxels	
	LH	RH	LH	RH
		iPCS		
1	[−47, 10,34]	[47, 6, 35]	169	222
2	[−45, 7,35]	[46, 8, 37]	25	44
3	[−48, 8,23]	[47, 8, 29]	103	121
4	[−34,−2,32]	[42, 6, 35]	30	217
5	[−46, 2,32]	[44, 4, 28]	66	135
6	[−47, 6,29]	[46, 7, 28]	216	150
7	[−42,−1,33]	[41, 6, 32]	150	422
8	[−46, 8,30]	[50, 5, 32]	124	154
9	[−43, 5,33]	[42, 5, 34]	472	483
10	[−45, 1,30]	[45, 6, 33]	130	326
		AI/FO		
1	[−35, 19, 4]	[37,17, 5]	244	532
2	[−37, 17, 3]	[38,18, 1]	267	340
3	[−35, 15, 2]	[32,18, 5]	120	115
4	[−37, 18, 4]	[37,18, 3]	354	510
5	[−34, 20, 1]	[33,19,−1]	47	34
6	[−38, 18, 2]	[37,20, 2]	289	238
7	[−35, 17, 2]	[35,18, 2]	145	388
8	[−36, 16, 4]	[35,17, 3]	218	485
9	[−37, 19, 4]	[37,17, 5]	486	551
10	[−37, 19, 3]	[33,20, 5]	53	150
		IFS		
1	[−37, 35,28]	[40,30, 28]	9	46
2	[−29, 47,17]	[34,37, 20]	47	235
3	[−44, 33,22]	[42,31, 22]	44	55
4	[−42, 29,23]	[38,30, 23]	5	96
5	[−28, 49,21]	[34,41, 18]	53	105
6	[−35, 42,26]	[41,39, 26]	105	85
7	[−40, 36,19]	[40,40, 21]	152	460
8	[−40, 33,18]	[41,33, 22]	71	308
9	[−41, 34,26]	[44,30, 27]	242	189
10	[−39, 37,21]	[38,43, 24]	33	102

Data are centers and sizes of final regions of interest (ROIs) in early visual cortex (V1, V2, V3, V3AB, V4), lateral occipital complex (LO, pFus), intraparietal sulcus [IPS0–1, IPS2–3, superior IPS (sIPS)], and the multiple-demand network [superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), anterior insula/frontal operculum (AI/FO), and inferior frontal sulcus (IFS)] for each subject (see METHODS for details). Coordinates of each ROI center are described in Talairach space, where X = left-right axis (negative is left), Y = anterior-posterior axis (negative is posterior), and Z = inferior-superior axis (negative is inferior).

task relevance of each match dimension, as well as how it differed on correct and incorrect trials.

Several ROIs did show a significant difference in mean signal between the identity and viewpoint tasks (data not shown). Therefore, before performing MVPA, we first mean-centered the voxel activation pattern on each trial by calculating the mean across voxels on each trial and subtracting this value from the voxel activation pattern. This ensured that classification was based on information encoded in the relative pattern of activity across voxels in each ROI, rather than information about mean signal changes across conditions.

We performed all decoding analyses using a binary classifier based on the normalized Euclidean distance. To avoid overfitting, we used a leave-one-run-out cross-validation scheme so that each run served as the test set once. Before starting this analysis, we removed all trials that were the first in a block, because they could not be labeled as a match or nonmatch. Next, we divided data in the training set into two groups based on status as a match in the dimension of interest. For each of these two groups, we then calculated a mean voxel activation pattern (e.g., averaging the response of each voxel over all trials in the group). We also calculated the pooled variance of each voxel's

response across the two groups. Next, for each trial in the test set, we calculated the normalized Euclidean distance to each of the mean patterns of the training set groups, weighting each voxel's contribution on the basis of its pooled variance. We then assigned each test set trial to the group with the minimum normalized Euclidean distance. Specifically, for a training set including n total trials, with n_a trials in *condition A* and n_b trials in *condition B*, and v voxels in each activation pattern, we can define $\bar{a}$, σ_a^2, $\bar{b}$, and σ_b^2 as vectors of size $[1xv]$ describing the mean and variance of each voxel's response within *conditions A* and *B*, respectively. If x is a $[1xv]$ vector describing a voxel activation pattern from a single trial in the test set, the normalized Euclidean distance from x to each of the two training set conditions is

$$d_{x \to a} = \sqrt{\sum_{i=1}^{v} \left(\frac{x_i - \bar{a}_i}{\sigma_{p_i}^2} \right)^2}$$

$$d_{x \to b} = \sqrt{\sum_{i=1}^{v} \left(\frac{x_i - \bar{b}_i}{\sigma_{p_i}^2} \right)^2}$$

where σ_p^2 is a $[1xv]$ vector describing the pooled variance of each voxel over *conditions A* and *B*:

$$\sigma_p^2 = \frac{n_a \sigma_a^2 + n_b \sigma_b^2}{n_a + n_b}$$

The final label assigned to each test set trial by the classifier was obtained by finding the minimum value between $d_{x \to a}$ and $d_{x \to b}$. Finally, we computed a single value for classifier performance across the entire data set by calculating d' with the formula

$$d' = Z(hit\ rate) - Z(false\ positive\ rate),$$

where the hit rate is defined as the proportion of test samples in *condition A* accurately classified as belonging to *condition A*, and the false positive rate is the proportion of test samples in *condition B* inaccurately classified as belonging to *condition A*. The function $Z(p), p \in [0,1]$ is the inverse of the cumulative distribution of the Gaussian distribution.

Because the frequency of matches in our task was less than 50%, the training set for the classifier was initially unbalanced. To correct for this, we performed downsampling on the larger training set group (nonmatch trials) by randomly sampling N trials without replacement from the larger set, where N is the number of samples in the smaller set. We performed 1,000 iterations of this random downsampling and averaged the results for d' over all iterations.

We assessed the significance of classifier decoding performance in each ROI using a permutation test in which we shuffled the labels of all trials in the training set and computed decoding performance on this shuffled data set. We repeated this procedure over 1,000 iterations to compute a null distribution of d' for each subject and each ROI. For each shuffling iteration, we performed downsampling to balance the training set as described above, but to reduce the computational time, we used only 100 iterations. To compute significance at the group level, we averaged the null distributions over all subjects to obtain a single distribution of 1,000 d' values, and averaged the d' values for the real data set over all subjects to obtain a subject-average d' value. We obtained a P value by calculating the proportion of shuffling iterations on which the shuffled d' value exceeded the real d' value, and the proportion on which the real d' value exceeded the shuffled d' value, and taking the minimum value multiplied by 2. We then performed FDR correction across ROIs within each condition and match type, at the 0.01 and 0.05 significance levels (Benjamini and Yekutieli 2001).

The above analysis was carried out separately within each ROI, task, and match dimension separately to estimate information about viewpoint and identity match status when each dimension was rele-

vant and irrelevant. Next, we entered all d' values into a three-way repeated-measures ANOVA with factors of task, ROI, and relevance. Following this, to more closely investigate interactions between ROI and relevance, we used nonparametric paired t-tests to compare the d' distributions for the relevant match dimension vs. the irrelevant match dimension, within each task and ROI separately. This test consisted of performing 10,000 iterations in which we randomly permuted the relevance labels corresponding to the d' values, maintaining the subject labels. After randomly permuting the labels, we calculated the difference in d' between the two conditions for each subject and used these 10 difference values to calculate a t-statistic. We then compared the distribution of these null t-statistics with the value of the t-statistic found with the real relevance labels and used this to generate a two-tailed P value. These P values were FDR corrected across ROIs at both the 0.05 and 0.01 levels.

In the first set of analyses (see Figs. 5 and 6), which focused on the overall performance of the classifier, we performed the above steps after removing all trials where the subject was incorrect or did not respond (on average, 7% of trials were no-response trials). In the next set of analyses (see Figs. 8 and 9), we were interested in whether information in each ROI about the task-relevant match dimension was associated with task performance. To evaluate this, we included incorrect and no-response trials in both the training and testing sets. We considered all incorrect and no-response trials as a single group, which we refer to as "incorrect." For each trial in the test set, we then used the normalized Euclidean distance (calculations described above) as a metric of classifier evidence in favor of the actual trial label, where

$$d_{x \to incorrect} - d_{x \to correct} = evidence.$$

We then compared the distributions of evidence between correct and incorrect test set trials. Because there were many more correct than incorrect test set trials, we performed an additional step of downsampling to balance the training set. We divided training set trials into four groups, based on their status as a match and the correctness of the subject's response, and downsampled the number of trials in all four groups to match the number of trials in the smallest set. This was performed over 1,000 iterations, and the resulting values for classifier evidence were averaged. We assessed significance of the difference between correct and incorrect trials using a permutation test as described above.

In the main analyses described above, we performed MVPA using all voxels from each localized ROI. Additionally, to control for differences in the number of voxels between ROIs, we repeated all analyses after restricting the number of voxels to 50 in each area. Overall, reducing the number of voxels did not lead to a dramatic change in the patterns of decoding performance across ROIs. These analyses are all reported in the figures.

Experimental design and statistical analyses. The sample size for this experiment was 10, with subjects run in 1 or 2 scanning sessions to collect at least 16–22 runs of experimental data, as well as between 2 and 6 runs of each functional localizer described above. This sample size was determined before data collection was started, based on sample sizes used by past experiments with similar methodology in our laboratory. For details of MVPA analyses and related statistics, see *MVPA decoding*. Briefly, all statistical tests, including repeated-measures ANOVAs, were performed using MATLAB R2017a (The MathWorks, Natick, MA) and were based on within-subject factors. The significance of MVPA results was assessed using permutation testing, with the final test for significance performed across all subjects. Pairwise comparisons of classifier output (e.g., classifier evidence on correct vs. incorrect trials, decoding d' for task-relevant vs. task-irrelevant match status) were performed using a nonparametric permutation-based t-test. Multiple comparisons correction was performed using FDR correction as described in Benjamini and Yekutieli (2001). We chose two thresholds because the 0.05 value

provides slightly more power to detect weaker effects, whereas the 0.01 value provides a more conservative threshold.

Image similarity analysis. The viewpoint and identity tasks were designed so that the low-level shape similarity of the images would not be explicitly informative about the status of each image as a match. Thus, when we performed classification on the status of each image as a target in each dimension, we intended to capture information that was related to perception of the abstract dimensions of each object, rather than low-level properties such as its shape in a 2D projection. However, because of factors such as the small number of objects in our stimulus set, it is possible that there was some coincidental, systematic structure in the similarity between pairs of objects such that low-level image similarity was partially informative about the status of each image as a match in either identity or viewpoint.

To evaluate this possibility, for each trial in the sequence of images shown to each subject, we determined the image similarity between the current and previous images by unwrapping each image (1,000 × 1,000 pixels) into a single vector and calculating the Pearson correlation coefficient between each pair of vectors. For this analysis, we removed trials that were a match in both category and viewpoint (identical images by design) and sorted the remaining trials according to whether the current and previous images were actually a match in the dimension of interest. Mean image similarity between match and nonmatch trials was compared using a one-tailed t-test (see Fig. 6D). This resulted in a P value for each subject in each condition, which was used to evaluate the extent to which low-level image similarity may have been informative about match status.

In addition to using the Pearson correlation to measure similarity, we also assessed similarity by passing each image through a Gabor wavelet model meant to simulate the responses of V1 neurons to the spatial frequency and orientation content of the image (Pinto et al. 2008). We then compared the effectiveness of this V1 model and the simpler pixel model at capturing the responses in V1 from our fMRI data, by calculating a similarity matrix for each pairwise comparison of images (24 × 24), based on *1*) the V1 model, *2*) the pixel model, and *3*) the voxel activation patterns recorded in V1 for each subject. We found that across all subjects, the pixel model was more correlated with the V1 voxel responses than was the V1 model (data not shown). Therefore, in the interest of parsimony, we used the simpler pixel model for all image similarity analyses.

RESULTS

Behavioral performance. Subjects ($n = 10$) performed alternating runs of the identity task and the viewpoint task while undergoing fMRI. Runs were always presented in matched pairs so that the object sequence and visual stimulation were identical between runs of the identity task and viewpoint task. We found no significant difference in performance (d' for identity: 1.47 ± 0.24, d' for viewpoint: 1.81 ± 0.34; paired 2-tailed t-test, $P = 0.2054$; Fig. 3A;) and no significant difference in response times (RT for identity: 1.15 ± 0.06 s, RT for viewpoint: 1.17 ± 0.05 s; paired 2-tailed t-test, $P = 0.6124$; Fig. 3B) on the two tasks across subjects. Performance and response time for each task also did not differ as a function of the object set subjects had been assigned to (d' for identity, *set A*: 1.50 ± 0.22, d' for identity, *set B*: 1.43 ± 0.27, $P = 0.8930$; d' for viewpoint, *set A*: 1.76 ± 0.30, d' for viewpoint, *set B*: 1.86 ± 0.42, $P = 0.9024$; RT for identity, *set A*: 1.11 ± 0.04 s, RT for identity, *set B*: 1.19 ± 0.07 s, $P = 0.5343$; RT for viewpoint, *set A*: 1.18 ± 0.02, RT for viewpoint, *set B*: 1.16 ± 0.08, $P = 0.8843$; all are 2-tailed t-tests).

Univariate BOLD signal does not show match suppression. First, we examined whether the status of each image as a match in the task-relevant dimension (e.g., identity match status
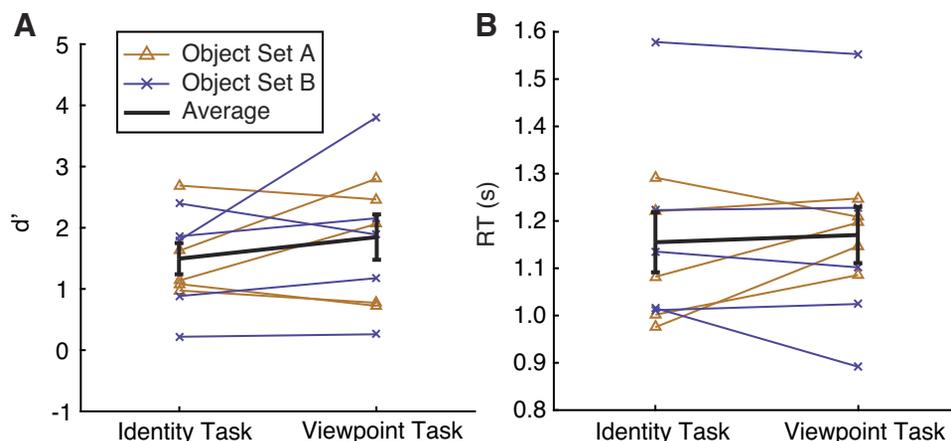
Fig. 3. Behavioral performance (d'; A) and response time (RT; B) were similar across tasks and stimulus sets. Each line represents performance of a single subject, averaged over runs of each task. Error bars are ±SE.

during the identity task; viewpoint match status during the viewpoint task) was reflected in a change in the mean amplitude of BOLD response in any visual area. On the basis of previous fMRI studies (Grill-Spector et al. 1999; Henson 2003) and electrophysiology studies (Meyer and Rust 2018; Miller et al. 1991), we predicted that the repetition of object identity or viewpoint might result in response suppression (often referred to as repetition suppression). However, in all but 1 of the 14 ROIs we examined (Fig. 4), we found no significant difference in the mean signal amplitude between matches and nonmatches. In the one ROI showing a significant effect (AI/FO), mean signal was actually higher on match trials than on nonmatch trials. We suspect that the absence of repetition suppression is due to differences in task demands between our experiment and previous work, particularly the fact that identity and viewpoint matches were task relevant in our paradigm and thus may have evoked larger attention-modulated responses that counteracted any repetition suppression effects (see DISCUSSION for details). Note that in all subsequent multivariate analyses, we first de-meaned responses across all voxels in each ROI so that single-trial voxel activation patterns were centered at zero and any small univariate effects could not contribute to decoding performance (see METHODS).

Multivariate activation patterns reflect task-relevant match status. Next, we examined how voxel activation patterns in each ROI reflected the status of a stimulus as a match in viewpoint and identity, and how representations of viewpoint and identity match status were influenced by the task relevance of each dimension. During each task, a correct behavioral response depended on the object being a match to the previous object in the relevant dimension (identity or viewpoint), whereas match status in the other dimension was irrelevant. Therefore, we expected that information about the match status of an object in each dimension would be more strongly represented when that dimension was task relevant.

Indeed, status as a match along the task-relevant dimension, measured by classifier performance (d'), was represented widely within the ROIs we examined, whereas the irrelevant match was not represented at an above-chance level in any ROI (Fig. 5). Information about the task-relevant match increased along a posterior-to-anterior axis such that match status was represented most strongly in MD and IPS ROIs but was comparatively weaker in early visual cortex and the lateral occipital complex (LOC). Relevant match decoding performance was above chance for all MD ROIs for both the identity

and viewpoint tasks, and was also above chance for LO, V3AB, and V2 for both tasks. Decoding performance was above chance in V1, V4, and pFus for the identity task only.

The general pattern of decoding performance was similar across tasks, although there was a trend toward higher relevant match decoding performance in IPS for the viewpoint task than for the identity task. There was also an opposite trend in V4 and pFus for higher relevant match decoding during the identity task than during the viewpoint task. A three-way repeated-measures ANOVA with factors of task, ROI, and relevance revealed a main effect of relevance [$F(1,9) = 46.219$, $P = 10^{-4}$], a main effect of ROI [$F(13,117) = 9.930$, $P < 10^{-12}$], and a relevance $\times$ ROI interaction [$F(13,117) = 13.981$, $P < 10^{-17}$], but no main effect of task [$F(1,9) = 1.819$, $P = 0.2104$]. The interactions task $\times$ ROI [$F(13,117) = 1.838$, $P = 0.0450$] and task $\times$ relevance [$F(1,9) = 1.019$, $P = 0.3392$] were not significant at $\alpha = 0.01$. We further investigated the ROI $\times$ relevance interaction, using paired t-tests to compare decoding of the relevant and irrelevant dimensions, and found that the effect of relevance was significant for all MD regions in both tasks, LO in both tasks, as well as V3AB, V4, and pFus for the identity task only (Fig. 5). These results were similar when we used all voxels in each area (Fig. 5, A and B) and when we used 50 voxels in each area (Fig. 5, C and D).

Because our MD localizer failed to yield any voxels in the sPCS ROI for one subject, we used an interpolation method to fill in this value before performing the repeated-measures ANOVA (see METHODS for details). As an alternative method, we also performed the same test after removing all data from the subject that was missing sPCS, and we found similar results (Table 2). We note, however, that the task $\times$ ROI interaction term, which was not significant when the interpolation method was used, was significant when the missing subject was removed entirely [$F(13,104) = 2.530$, $P = 0.0047$].

Control analyses for visually driven match representations. To perform the identity and viewpoint tasks, subjects were required to use representations of object identity and viewpoint that were largely invariant to shape. However, a subset of trials in each task could, in principle, be solved based only on shape similarity between the previous and current objects. In the viewpoint task, because subtle differences between exemplars were not task relevant, the group of trials that could be solved based only on shape similarity included all trials that were matches in both category and viewpoint. In the identity task,
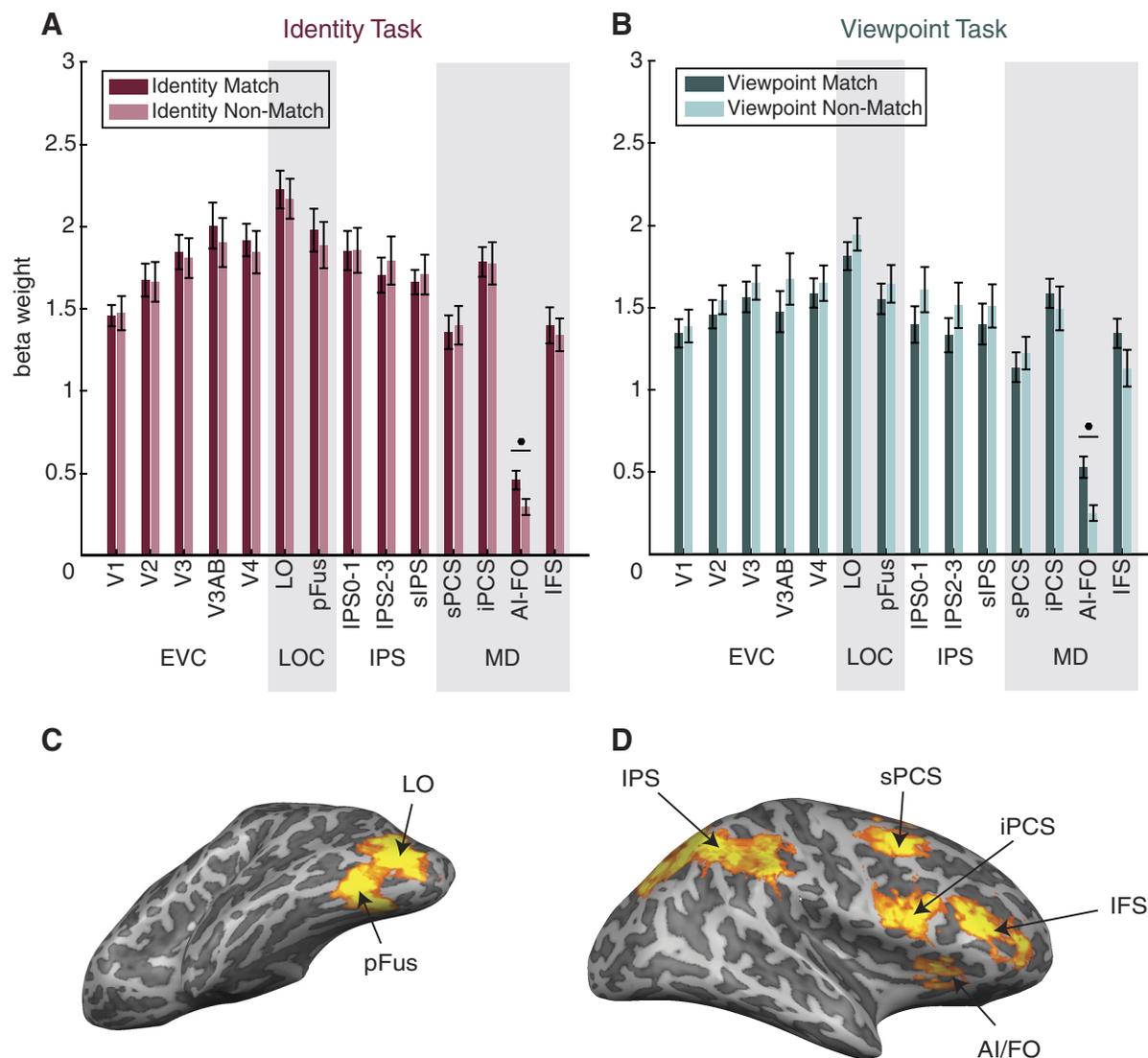
Fig. 4. Match status in the current task cannot be determined solely from mean signal change. *A* and *B*: mean beta weights are plotted on the *y*-axis, and individual regions of interest (ROIs) are plotted on the *x*-axis. In all functional MRI analyses, we used 14 ROIs: early visual cortex (EVC; comprising V1, V2, V3, V3AB, and V4), lateral occipital complex (LOC; comprising LO and pFus), intraparietal sulcus [IPS; comprising IPS0−1, IPS2−3, and superior IPS (sIPS)], and the multiple-demand (MD) network [superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), anterior insula/frontal operculum (AI/FO), and inferior frontal sulcus (IFS)], all of which were defined using functional localizers (see METHODS). ROIs are organized into 4 groups for convenience. Note that although we have visually separated the groups, we consider IPS to be a part of the MD network. In the identity task (*A*), in all but one ROI, univariate activation (mean beta weight) did not differ between trials according to their status as an identity match. Similarly, in the viewpoint task (*B*), univariate activation in most ROIs did not reflect viewpoint match status. Match and nonmatch trials were compared using paired *t*-tests. *P* values were FDR-corrected over all conditions. Circles indicate significance at *q* = 0.01. Error bars are ±SE. *C*: example definitions of the LOC ROIs shown on an inflated mesh (left hemisphere of *subject 1*). *D*: example definitions of the MD ROIs shown on an inflated mesh (right hemisphere of *subject 1*). Subdivisions of IPS are not shown. For a summary of the definitions of all ROIs in all subjects, see Table 1.

the group of trials that could be solved on the basis of shape similarity included all trials that were an exact match to the previous image in category, exemplar, and viewpoint. Therefore, it is possible that some of the match status classification we observed in Fig. 5 was driven by the detection of shape similarity. To test for this possibility, we removed all trials that were a match in both category and viewpoint, leaving a set of trials in which the shape similarity was entirely uninformative about match status. We then performed classification on this reduced data set as before.

For the viewpoint task, we now found an important difference between visual and MD regions: whereas decoding of viewpoint matches remained above chance in all MD and IPS

regions, it dropped to chance in LOC and early visual cortex (Fig. 6*B*). Thus, whereas early visual and LOC representations of viewpoint match status appeared to rely largely on low-level shape similarity, MD regions encoded viewpoint match status even when shape similarity could not be used to define a match.

During the identity task, however, even after removal of all trials that were a match in both category and viewpoint, decoding of identity match status remained above chance in all ROIs examined (Fig. 6*A*). The observation of above-chance identity match decoding in early visual areas was surprising given that these areas are not expected to encode abstract, viewpoint-independent representations of target status. There-
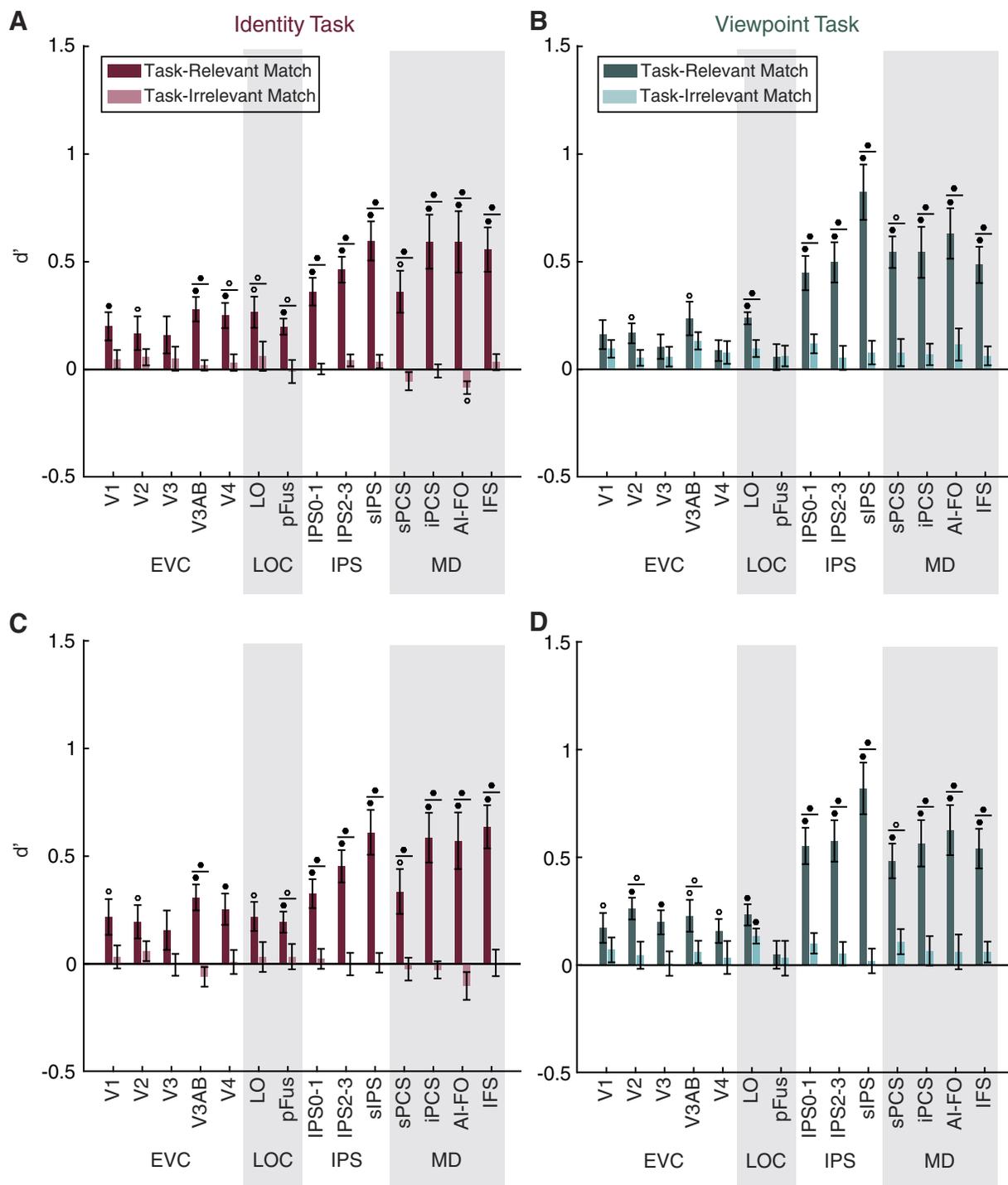
Fig. 5. Task-relevant matches are represented more strongly than task-irrelevant matches. *A–D*: a linear classifier was trained to discriminate between voxel activation patterns measured during runs of the identity task (*A* and *C*) or the viewpoint task (*B* and *D*), according to whether the viewed image was a match in the task-relevant or the task-irrelevant dimension. Classifier performance (d') is plotted on the *y*-axis for 14 regions of interest (ROIs): early visual cortex (EVC; comprising V1, V2, V3, V3AB, and V4), lateral occipital complex (LOC; comprising LO and pFus), intraparietal sulcus [IPS; comprising IPS0–1, IPS2–3, and superior IPS (sIPS)], and the multiple-demand (MD) network [superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), anterior insula/frontal operculum (AI/FO), and inferior frontal sulcus (IFS)]. Circles above individual bars indicate above-chance classification performance (test against 0); circles above pairs of bars (denoted by horizontal lines) indicate significant differences between bars (paired *t*-test). *P* values were FDR-corrected over all conditions. Open circles indicate significance at $q = 0.05$; closed circles indicate significance at $q = 0.01$. Error bars are ±SE. *A* and *B* show decoding performance using all voxels in each ROI; *C* and *D* show decoding performance using only 50 voxels in each ROI.

fore, we wondered whether the remaining set of images may have had some shared features that supported match classification. Indeed, when we assessed the pixelwise similarity between images (see METHODS) belonging to each stimulus set,

we found that in four of the subjects assigned to object *set A*, pixelwise similarity between images was significantly predictive of identity match status (Fig. 6*D*). We thus hypothesized that the above-chance decoding of identity match status ob-

Table 2. *Results of three-way repeated-measures ANOVA on decoding results*

	SumSq	df	MeanSq	*F* Statistic	*P* Value
	RM ANOVA, using interpolation				
(Intercept)	23.3373	1	23.3373	55.1700	**<10⁻⁴**
Error	3.8071	9	0.4230		
(Intercept):ROI	4.7641	13	0.3665	9.9297	**<10⁻¹²**
Error(ROI)	4.3180	117	0.0369		
(Intercept):Task	0.1433	1	0.1433	1.8193	0.2104
Error(Task)	0.7090	9	0.0788		
(Intercept):Relevance	13.4705	1	13.4705	46.2189	**0.0001**
Error(Relevance)	2.6230	9	0.2914		
(Intercept):ROI:Task	0.6030	13	0.0464	1.8375	0.0450
Error(ROI:Task)	2.9535	117	0.0252		
(Intercept):ROI:Relevance	5.5957	13	0.4304	13.9806	**<10⁻¹⁷**
Error(ROI:Relevance)	3.6022	117	0.0308		
(Intercept):Task:Relevance	0.1618	1	0.1618	1.0185	0.3392
Error(Task:Relevance)	1.4296	9	0.1588		
(Intercept):ROI:Task:Relevance	0.3903	13	0.0300	1.0157	0.4411
Error(ROI:Task:Relevance)	3.4585	117	0.0296		
	RM ANOVA, with subject 1 removed				
(Intercept)	22.4745	1	22.4745	50.5309	**0.0001**
Error	3.5581	8	0.4448		
(Intercept):ROI	4.7895	13	0.3684	9.9486	**<10⁻¹²**
Error(ROI)	3.8514	104	0.0370		
(Intercept):Task	0.1458	1	0.1458	1.6572	0.2340
Error(Task)	0.7039	8	0.0880		
(Intercept):Relevance	13.2417	1	13.2417	44.5775	**0.0002**
Error(Relevance)	2.3764	8	0.2970		
(Intercept):ROI:Task	0.7977	13	0.0614	2.5299	**0.0047**
Error(ROI:Task)	2.5227	104	0.0243		
(Intercept):ROI:Relevance	5.0628	13	0.3894	13.4838	**<10⁻¹⁵**
Error(ROI:Relevance)	3.0038	104	0.0289		
(Intercept):Task:Relevance	0.0959	1	0.0959	0.5569	0.4769
Error(Task:Relevance)	1.3779	8	0.1722		
(Intercept):ROI:Task:Relevance	0.4434	13	0.0341	1.0929	0.3734
Error(ROI:Task:Relevance)	3.2455	104	0.0312		

Data are the results of a 3-way repeated-measures (RM) ANOVA with the factors region of interest (ROI), task, and relevance, performed on the decoding results shown in Fig. 5, *A* and *B*. We were unable to define the superior precentral sulcus (sPCS) ROI in 1 of 10 subjects and used two different methods to address this missing value before running the RM ANOVA. Results are shown for both an interpolation method (see METHODS for details) and with all data removed corresponding to the subject who was missing sPCS (using 9/10 subjects). Bold values indicate *P* values that were significant at $\alpha = 0.01$. For both tests, Mauchly's test revealed that the data did not violate the assumption of sphericity, so we report uncorrected *P* values. df, degrees of freedom; MeanSq, mean square; SumSq, sum of squares.

served in early visual areas might be driven by this group of subjects.

In line with this prediction, when we reanalyzed decoding performance using only subjects from *set B*, identity match decoding was no longer significant in V1, V2, V3, V4, and pFus, even though it was still well above chance in other parietal and MD areas. This pattern suggests that the above-chance decoding accuracy for identity match status in these earlier visual ROIs may have been related to low-level image features (Fig. 6*C*). In contrast, even after this additional source of low-level image similarity was removed, MD and IPS regions still encoded robust representations of identity match status. Furthermore, match decoding in the MD regions was individually strong in the majority of subjects (Fig. 7).

Behavioral performance is closely linked to activation patterns in the MD network. Having established that several ROIs, including all MD network ROIs, represent the status of an object as a match in the task-relevant dimension, we next sought to determine whether these representations were related to behavioral performance. To answer this question, we used the normalized Euclidean distance to calculate a continuous measure of the classifier's evidence at predicting the correct

label for each trial in the test set (see METHODS). We then calculated the mean classifier evidence for all correct and incorrect trials. Because the previous analysis (Fig. 5) indicated no significant effect of task on relevant match decoding, we combined all trials across the identity and viewpoint tasks for this analysis (Fig. 8).

In all regions of the MD network and IPS, we found that classifier evidence was significantly higher on correct than incorrect trials. In contrast, we found no significant effect of behavior on classifier evidence in early visual cortex or in the LOC. Thus representations in IPS and MD ROIs, but not any of the other regions that we evaluated, were significantly associated with task performance. We also verified that this effect was similar using data from each task separately (Fig. 9) and when voxel number was matched across ROIs (Fig. 8*B* and Fig. 9, *C* and *D*).

DISCUSSION

In this study we used fMRI and pattern classification methods to investigate the role of different brain areas in signaling task-relevant matches across identity-preserving transforma-

tions. Specifically, subjects performed a task that required them to identify matches in either the identity or the viewpoint of novel objects. Consistent with previous work, we found that areas of ventral visual cortex represent information about the status of an object as a relevant match (Fig. 5; Lueschow et al. 1994; Miller and Desimone 1994; Pagan et al. 2013; Woloszyn and Sheinberg 2009). However, our results also suggest a key role for the MD network in this match identification task, with MD match representations showing specificity to task-relevant dimensions (Fig. 5), as well as invariance to low-level visual features (Fig. 6). Importantly, the present data also establish a

significant link between task performance and the strength of MD match representations (Fig. 8). In contrast, in early visual and ventral object-selective cortex, we found comparatively weaker evidence for representations of match status and no significant associations with task performance. These results suggest that MD regions play a key role in computing flexible and abstract target representations that are likely to be important for task performance.

In contrast to our MVPA results, univariate signal amplitude in almost all ROIs was not significantly modulated by match status. This finding differs from many past fMRI
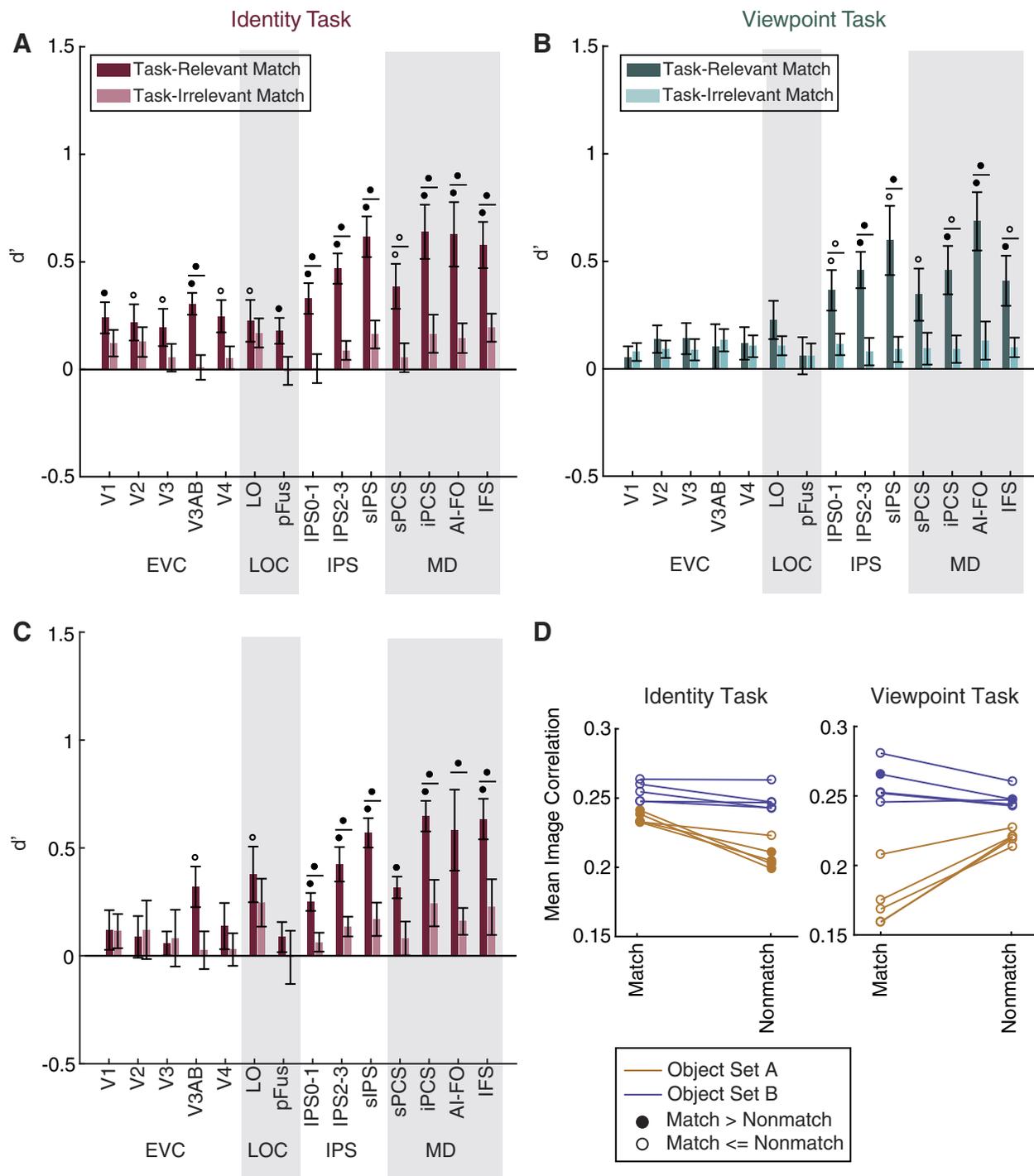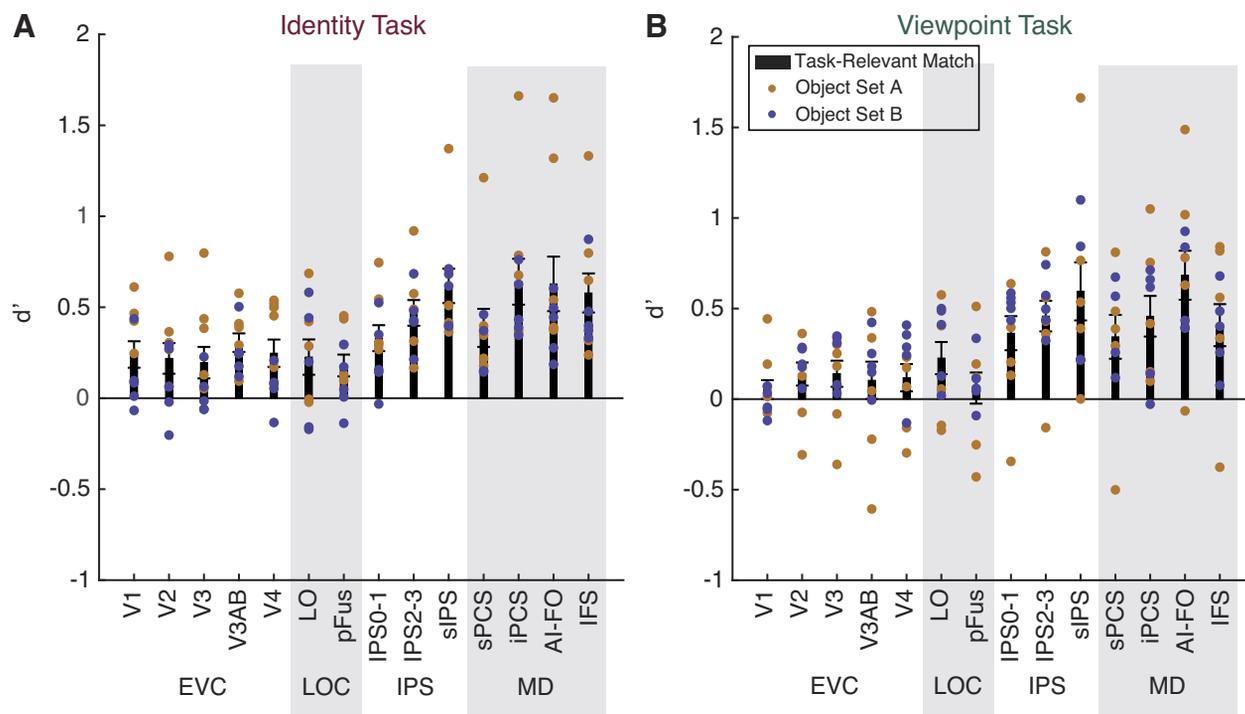
Fig. 7. After stimulus similarity confounds were addressed, most individual subjects in both stimulus sets still show above-chance match decoding in multiple-demand (MD) network regions of interest (ROIs): superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), anterior insula/frontal operculum (AI/FO), and inferior frontal sulcus (IFS). Relevant match decoding performance was calculated after removal of all trials in which the current and previous objects had a high degree of shape similarity (same analysis as Fig. 6, A and B). Black bars show the subject average $\pm$ SE and are identical to the dark colored bars in Fig. 6, A and B. Colored dots indicate individual subject decoding performance, with different stimulus sets shown in different colors. EVC, early visual cortex (comprising V1, V2, V3, V3AB, and V4); LOC, lateral occipital complex (comprising LO and pFus); IPS, intraparietal sulcus [comprising IPS0–1, IPS2–3, and superior IPS (sIPS)].

studies (Grill-Spector et al. 1999; Henson 2003; Turk-Browne et al. 2007) that have observed response suppression as a result of object repetition (or "repetition suppression"). One explanation for this divergence in findings is that in our task, the repetition of identity and viewpoint is task relevant, meaning that repetition-related signals are mixed with signals related to task performance. This account is consistent with prior electrophysiology studies reporting that when an object's match status is task relevant, neural response modulations are heterogeneous, including both enhancement and suppression (Engel and Wang 2011; Lui and Pasternak 2011; Miller and Desimone 1994; Pagan et al. 2013; Roth and Rust 2018). This type of signal would be detectable using multivariate decoding methods but in univariate analyses may be obscured by averaging across all

voxels in an ROI (Kamitani and Tong 2005; Norman et al. 2006; Serences and Saproo 2012). Therefore, our data are consistent with the interpretation that match representations are not an automatic by-product of stimulus repetition, but are linked to the task relevance of each stimulus.

Representations of identity and viewpoint match status were weaker in early visual and ventral ROIs compared with ROIs in the MD network. Moreover, several control analyses suggest that the representations in some of these occipital and ventral regions were driven primarily by low-level image statistics, as opposed to an object's status as a match in the task-relevant dimension. First, after removing all trials in which the current and previous objects were matches in both category and viewpoint, we found that viewpoint match decoding during the viewpoint task dropped to chance in all ROIs except for those

Fig. 6. Control analyses related to Fig. 5: viewpoint and identity match information in MD regions is not driven by low-level image statistics. A and B: to address the possibility that match status could have been inferred from low-level visual properties, we removed all trials in which an object had a high degree of shape similarity to the previous object and repeated the analyses of Fig. 5. A: identity match information remained above chance in all regions. B: viewpoint match information dropped to chance in the early visual cortex (EVC; comprising V1, V2, V3, V3AB, and V4) and lateral occipital complex (LOC; comprising LO and pFus) but remained above chance in multiple-demand (MD) network [superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), anterior insula/frontal operculum (AI/FO), and inferior frontal sulcus (IFS)] regions of interest (ROIs). For individual subject data, see Fig. 7. C: after identifying that the pairwise similarity between images in object set A was informative about identity match status, we reanalyzed the identity task data using only the subjects shown object set B. Identity match classification in early visual and ventral visual cortex drops below significance when subjects shown object set A are removed but remains above chance in MD regions. P values were computed at the subject level over these 5 subjects and FDR-corrected across ROIs. Open circles indicate significance at $q = 0.05$; closed circles indicate significance at $q = 0.01$. Circles above individual bars indicate above-chance classification performance (test against 0); circles above pairs of bars (denoted by horizontal lines) indicate significant differences between bars (paired t-test). Error bars are $\pm$SE. D: image correlation is predictive of identity match status for several subjects in object set A. To assess the possibility that identity match classification in EVC may have been driven by low-level similarity between pairs of images, we used the Pearson correlation coefficient to calculate the similarity between all pairs of object images (excluding image pairs that were matched in both category and viewpoint and thus highly similar). For each subject, we calculated the mean correlation coefficient between pairs of images that were a match in each feature vs. images that were not a match. In the plot, closed circles indicate that this mean value was higher for matching pairs than for nonmatching pairs ($\alpha = 0.01$). This finding suggests that the above-chance classification of identity match status observed in set A subjects may have been driven by low-level image features.
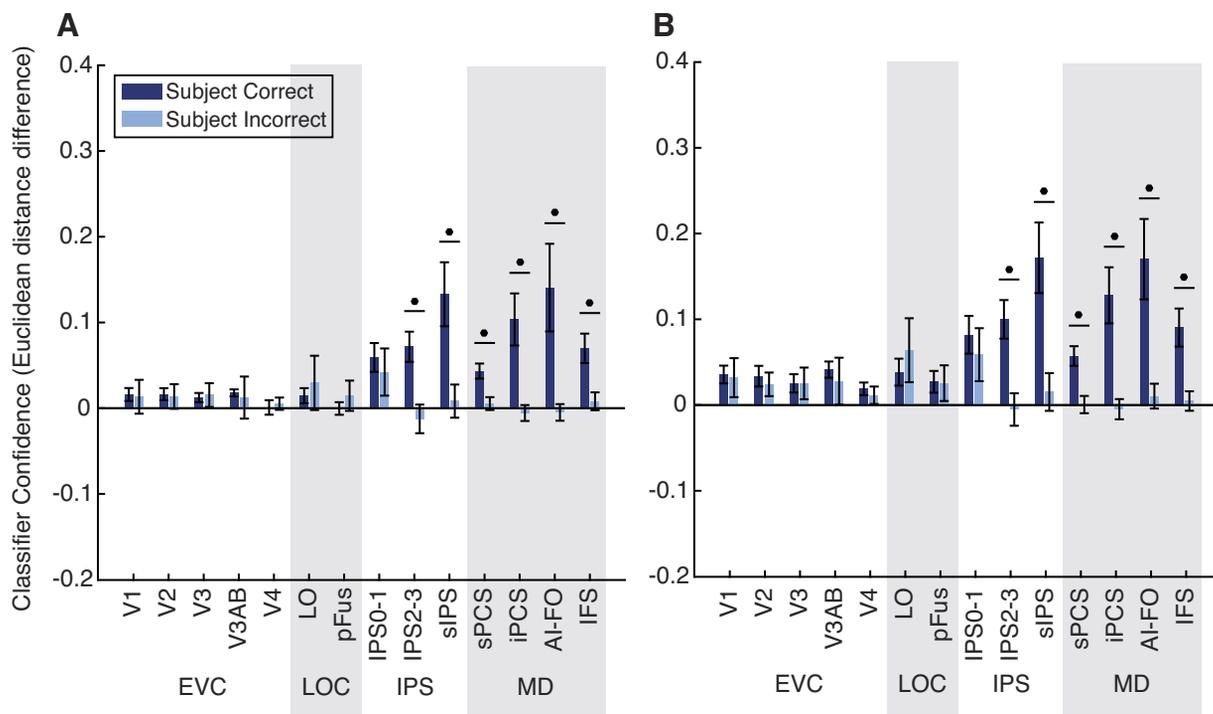
Fig. 8. Classifier evidence is associated with task performance. *A* and *B*: trials were classified according to their status as a match in the task-relevant feature, and the normalized Euclidean distance was used as a metric of evidence in favor of the correct classification (distance to incorrect label minus distance to correct label, plotted on the *y*-axis). Circles above pairs of bars indicate a significant difference between correct and incorrect trials (paired *t*-test). Closed circles indicate significance at $q = 0.01$ (FDR corrected). Error bars are ±SE. *A* shows results using all voxels; *B* shows results using 50 voxels per region of interest (ROI): early visual cortex (EVC; comprising V1, V2, V3, V3AB, and V4), lateral occipital complex (LOC; comprising LO and pFus), intraparietal sulcus [IPS; comprising IPS0–1, IPS2–3, and superior IPS (sIPS)], and the multiple-demand (MD) network [superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), anterior insula/frontal operculum (AI/FO), and inferior frontal sulcus (IFS)].

in the MD network. Additionally, a post hoc image-based analysis revealed that one of our stimulus sets had a subtle bias in pixel similarity such that the degree of image similarity between each pair of objects was informative about the status of the pair as an identity match. When we repeated our decoding analysis using only subjects who had seen the non-biased stimulus set, we found that identity match decoding dropped to chance in V1, V2, V3, V4, and pFus, although it remained above chance in V3AB and LO (and in regions of IPS and the MD network). Together, these findings suggest that low-level visual features may be partially responsible for the viewpoint and identity match representations we initially observed in early visual and ventral visual areas. We note, however, that this is not a powerful test; a more targeted experiment would be needed to rigorously assess the role of image similarity in driving match representations in early and ventral areas.

Our findings suggest that representations of object target status in MD regions show more invariance to identity preserving transformations than representations in ventral regions, which are more strongly influenced by low-level visual features. This view is consistent with several previous studies in nonhuman primates. For instance, in tasks that require identification of targets based on their membership in abstract categories whose boundaries are not predicted from visual similarity, neurons in both prefrontal cortex (PFC; Cromer et al. 2010; Freedman et al. 2001, 2003; Roy et al. 2010) and premotor cortex (Cromer et al. 2011) encode objects' target status. Studies that directly compare PFC and inferotemporal cortex (ITC) responses during these tasks have found that PFC

neurons show a higher degree of abstract category selectivity (Freedman et al. 2003), as well as a stronger influence of task-related effects on object representation (McKee et al. 2014), compared with ITC neurons. Past work in humans is also consistent with a high degree of abstraction in MD representations. A recent fMRI study found that abstract face-identity information is represented more strongly in IPS than in LO and the fusiform face area (Jeong and Xu 2016). Another study found that during the delay period of a working memory task, object information could be decoded from PFC only when the task was nonvisual and from the posterior fusiform area only when the task was visual, supporting a dissociation between these regions in representing abstract and visual object information, respectively (Lee et al. 2013). Overall, our findings provide additional support for the conclusion that target representations in frontoparietal cortex reflect abstract signals that flexibly update to guide behavior, while ventral representations are more linked to the details of currently-viewed images.

That said, our results do not rule out the possibility of abstract object information or target-related information in the ventral stream. Many past studies have demonstrated viewpoint-invariant object information in human and primate ventral visual cortex (Anzellotti et al. 2014; Erez et al. 2016; Freiwald and Tsao 2010; Tanaka 1996). Furthermore, several studies have found that match representations in ITC are capable of generalizing across changes in size and position of objects (Lueschow et al. 1994; Roth and Rust 2018). Despite this, our multivariate decoding analysis failed to detect an association with behavior in any ventral ROI. One explanation
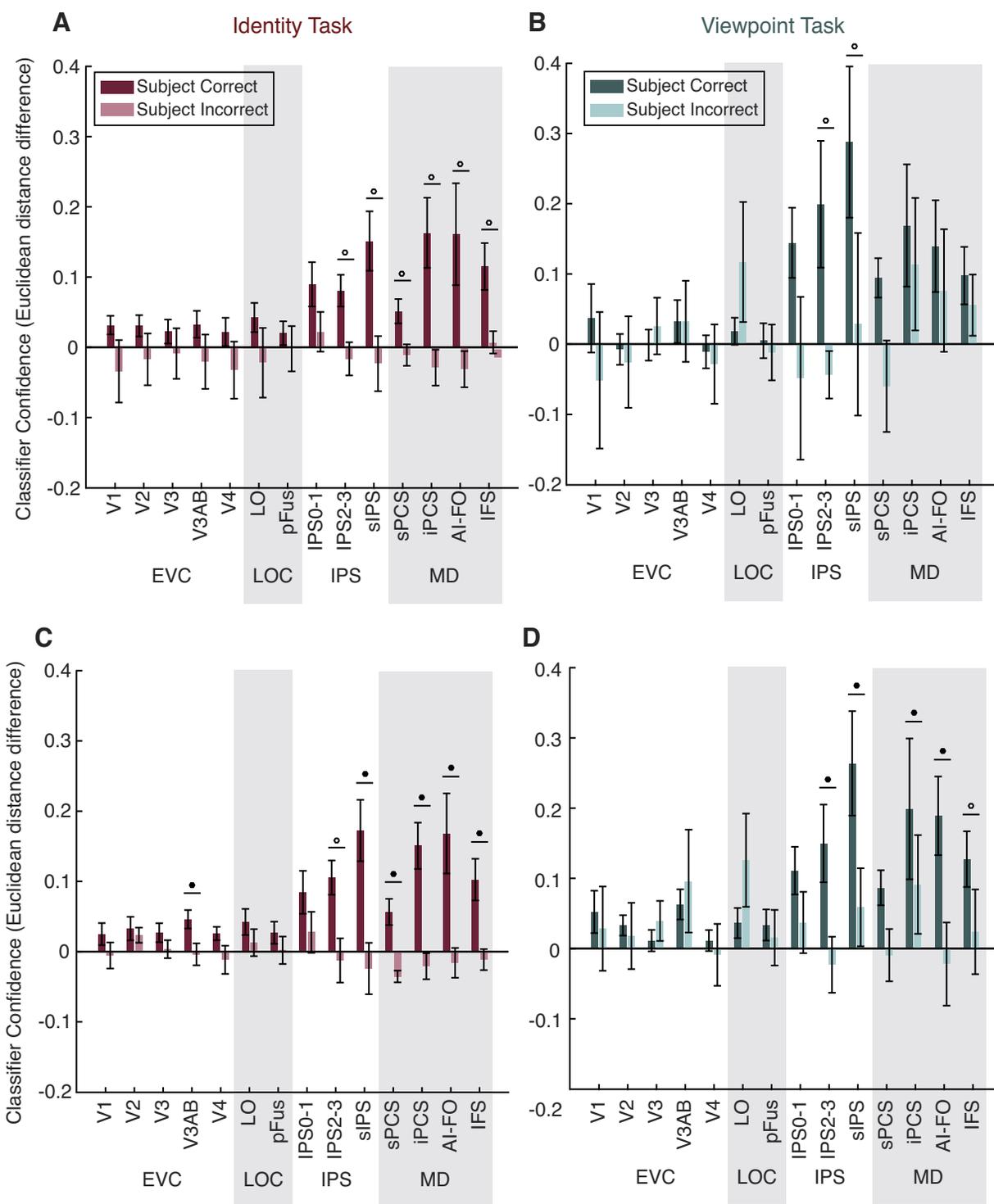
Fig. 9. Classifier evidence is associated with performance on both tasks individually. *A–D*: analysis is identical to that in Fig. 8 but was carried out using trials from the identity task only (*A* and *C*) or the viewpoint task only (*B* and *D*). *A* and *B* show results using all voxels; *C* and *D* show results using 50 voxels per region of interest (ROI): early visual cortex (EVC; comprising V1, V2, V3, V3AB, and V4), lateral occipital complex (LOC; comprising LO and pFus), intraparietal sulcus [IPS; comprising IPS0–1, IPS2–3, and superior IPS (sIPS)], and the multiple-demand (MD) network [superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), anterior insula/frontal operculum (AI/FO), and inferior frontal sulcus (IFS)].

for this finding is that target information was present at a level of granularity that is too fine to be detectable with our methods, or that the signal-to-noise regime in ventral object-selective cortex prevented us from detecting abstract object representations (Dubois et al. 2015). Furthermore, it is possible that linearly decodable information about abstract object properties

may be more readily detectable at later stages of the ventral visual stream, such as perirhinal cortex, compared with LO and pFus (Erez et al. 2016; Pagan et al. 2013). Future studies will be needed to determine the extent to which abstract matches in identity and viewpoint may be represented within ventral visual cortex.

Additionally, although match representations in early visual cortex and the LOC were weaker than those measured in MD regions, we did observe a consistent modulation of these representations by task such that the relevant match was generally represented more strongly than the irrelevant match. This modulation was significant in LO during both tasks and in V4, V3AB, and pFus during the identity task (Fig. 5). One interpretation for the enhanced representation of relevant match status is the presence of top-down feedback that influences the content of visual representations in early visual and ventral object-selective cortex. This feedback could act to enhance representations of visual properties that are informative for computing match status, which would result in an enhancement of signals related to repetition of these properties. Alternatively, these feedback signals could contain nonspecific information about the presence of a relevant target. In either case, our data do not provide strong evidence that the match representations in early visual and ventral ROIs are involved in task performance.

Finally, one novel finding of the present study is that MD regions encode representations of matches in viewpoint across objects, in addition to matches in identity. This is especially interesting given that subjects had to be trained to recognize an arbitrary viewpoint of each object as its front. Therefore, the viewpoint dimension of an object may be regarded as a learned, semantic dimension rather than an intuitive physical dimension of the object. This may be another reason why we do not see invariant representations of viewpoint match signals in the ventral visual areas examined, because these regions may not encode such arbitrary dimensions that define potentially relevant objects. Further work will be needed to understand the extent to which neural representations of object viewpoint can be separated from visual features such as the axis of elongation.

In conclusion, our findings support a role of MD network regions in computing high-level and abstract target representations based on features such as viewpoint and identity, even across changes in retinal input patterns. The match representations in these regions are modulated by task relevance and are associated with behavioral performance on a trial-by-trial basis. In contrast, early visual and ventral object selective cortex contained weaker evidence for match representations, with some modulation by task, consistent with a role for top-down feedback in enhancing representations of relevant information at the earliest levels of processing.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

M.M.H. and J.T.S. conceived and designed research; M.M.H. performed experiments; M.M.H. analyzed data; M.M.H. and J.T.S. interpreted results of experiments; M.M.H. prepared figures; M.M.H. drafted manuscript; M.M.H. and J.T.S. edited and revised manuscript; M.M.H. and J.T.S. approved final version of manuscript.

REFERENCES

Andersson JL, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* 20: 870–888, 2003. doi:10.1016/S1053-8119(03)00336-7.

Andresen DR, Vinberg J, Grill-Spector K. The representation of object viewpoint in human visual cortex. *Neuroimage* 45: 522–536, 2009. doi:10.1016/j.neuroimage.2008.11.009.

Anzellotti S, Fairhall SL, Caramazza A. Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex* 24: 1988–1995, 2014. doi:10.1093/cercor/bht046.

Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165–1188, 2001. doi:10.1214/aos/1013699998.

Biederman I. Recognizing depth-rotated objects: a review of recent research and theory. *Spat Vis* 13: 241–253, 2000. doi:10.1163/156856800741063.

Bracci S, Daniels N, Op de Beeck H. Task context overrules object- and category-related representational content in the human parietal cortex. *Cereb Cortex* 27: 310–321, 2017. doi:10.1093/cercor/bhw419.

Conway BR. The organization and operation of inferior temporal cortex. *Annu Rev Vis Sci* 4: 381–402, 2018. doi:10.1146/annurev-vision-091517-034202.

Cromer JA, Roy JE, Buschman TJ, Miller EK. Comparison of primate prefrontal and premotor cortex neuronal activity during visual categorization. *J Cogn Neurosci* 23: 3355–3365, 2011. doi:10.1162/jocn_a_00032.

Cromer JA, Roy JE, Miller EK. Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66: 796–807, 2010. doi:10.1016/j.neuron.2010.05.005.

DiCarlo JJ, Cox DD. Untangling invariant object recognition. *Trends Cogn Sci* 11: 333–341, 2007. doi:10.1016/j.tics.2007.06.010.

Dubois J, de Berker AO, Tsao DY. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci* 35: 2791–2802, 2015. doi:10.1523/JNEUROSCI.4037-14.2015.

Duncan J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci* 14: 172–179, 2010. doi:10.1016/j.tics.2010.01.004.

Engel SA, Glover GH, Wandell BA. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 7: 181–192, 1997. doi:10.1093/cercor/7.2.181.

Engel TA, Wang X-J. Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *J Neurosci* 31: 6982–6996, 2011. doi:10.1523/JNEUROSCI.6150-10.2011.

Erez J, Cusack R, Kendall W, Barense MD. Conjunctive coding of complex object features. *Cereb Cortex* 26: 2271–2282, 2016. doi:10.1093/cercor/bhv081.

Erez Y, Duncan J. Discrimination of visual categories based on behavioral relevance in widespread regions of frontoparietal cortex. *J Neurosci* 35: 12383–12393, 2015. doi:10.1523/JNEUROSCI.1134-15.2015.

Fedorenko E, Duncan J, Kanwisher N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci USA* 110: 16616–16621, 2013. doi:10.1073/pnas.1315235110.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291: 312–316, 2001. doi:10.1126/science.291.5502.312.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23: 5235–5246, 2003. doi:10.1523/JNEUROSCI.23-12-05235.2003.

Freiwald WA, Tsao DY. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330: 845–851, 2010. doi:10.1126/science.1194908.

Grill-Spector K. The neural basis of object perception. *Curr Opin Neurobiol* 13: 159–166, 2003. doi:10.1016/S0959-4388(03)00040-0.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24: 187–203, 1999. doi:10.1016/S0896-6273(00)80832-6.

Guo F, Preston TJ, Das K, Giesbrecht B, Eckstein MP. Feature-independent neural coding of target detection during search of natural scenes. *J Neurosci* 32: 9499–9510, 2012. doi:10.1523/JNEUROSCI.5876-11.2012.

Henson RN. Neuroimaging studies of priming. *Prog Neurobiol* 70: 53–81, 2003. doi:10.1016/S0301-0082(03)00086-8.

Hong H, Yamins DL, Majaj NJ, DiCarlo JJ. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat Neurosci* 19: 613–622, 2016. doi:10.1038/nn.4247.

Ito M, Tamura H, Fujita I, Tanaka K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73: 218–226, 1995. doi:10.1152/jn.1995.73.1.218.

Jackson J, Rich AN, Williams MA, Woolgar A. Feature-selective attention in frontoparietal cortex: multivoxel codes adjust to prioritize task-relevant information. *J Cogn Neurosci* 29: 310–321, 2017. doi:10.1162/jocn_a_01039.

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. *Neuroimage* 62: 782–790, 2012. doi:10.1016/j.neuroimage.2011.09.015.

Jeong SK, Xu Y. Behaviorally relevant abstract object identity representation in the human parietal cortex. *J Neurosci* 36: 1607–1619, 2016. doi:10.1523/JNEUROSCI.1016-15.2016.

Jerde TA, Curtis CE. Maps of space in human frontoparietal cortex. *J Physiol Paris* 107: 510–516, 2013. doi:10.1016/j.jphysparis.2013.04.002.

Kadohisa M, Petrov P, Stokes M, Sigala N, Buckley M, Gaffan D, Kusunoki M, Duncan J. Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron* 80: 235–246, 2013. doi:10.1016/j.neuron.2013.07.041.

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8: 679–685, 2005. doi:10.1038/nn1444.

Lee SH, Kravitz DJ, Baker CI. Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci* 16: 997–999, 2013. doi:10.1038/nn.3452.

Lueschow A, Miller EK, Desimone R. Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex* 4: 523–531, 1994. doi:10.1093/cercor/4.5.523.

Lui LL, Pasternak T. Representation of comparison signals in cortical area MT during a delayed direction discrimination task. *J Neurophysiol* 106: 1260–1273, 2011. doi:10.1152/jn.00016.2011.

Marr D, Nishihara HK. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc Lond B Biol Sci* 200: 269–294, 1978. doi:10.1098/rspb.1978.0020.

McKee JL, Riesenhuber M, Miller EK, Freedman DJ. Task dependence of visual and category representations in prefrontal and inferior temporal cortices. *J Neurosci* 34: 16065–16075, 2014. doi:10.1523/JNEUROSCI.1660-14.2014.

Meyer T, Rust NC. Single-exposure visual memory judgments are reflected in inferotemporal cortex. *eLife* 7: e32259, 2018. doi:10.7554/eLife.32259.

Miller E, Erickson C, Desimone R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16: 5154–5167, 1996. doi:10.1523/JNEUROSCI.16-16-05154.1996.

Miller EK, Desimone R. Parallel neuronal mechanisms for short-term memory. *Science* 263: 520–522, 1994. doi:10.1126/science.8290960.

Miller EK, Li L, Desimone R. A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254: 1377–1379, 1991. doi:10.1126/science.1962197.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10: 424–430, 2006. doi:10.1016/j.tics.2006.07.005.

Pagan M, Urban LS, Wohl MP, Rust NC. Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. *Nat Neurosci* 16: 1132–1139, 2013. doi:10.1038/nn.3433.

Pinto N, Cox DD, DiCarlo JJ. Why is real-world visual object recognition hard? *PLoS Comput Biol* 4: e27, 2008. doi:10.1371/journal.pcbi.0040027.

Riesenhuber M, Poggio T. Models of object recognition. *Nat Neurosci* 3, *Suppl*: 1199–1204, 2000. doi:10.1038/81479.

Roth N, Rust NC. Inferotemporal cortex reflects behaviorally-relevant target match information during invariant object search (Preprint). *bioRxiv* 152181, 2018. doi:10.1101/152181.

Roth P. Missing data: a conceptual review for applied psychologists. *Pers Psychol* 47: 537–560, 1994. doi:10.1111/j.1744-6570.1994.tb01736.x.

Roy JE, Riesenhuber M, Poggio T, Miller EK. Prefrontal cortex activity during flexible categorization. *J Neurosci* 30: 8519–8528, 2010. doi:10.1523/JNEUROSCI.4837-09.2010.

Serences JT, Saproo S. Computational advances towards linking BOLD and behavior. *Neuropsychologia* 50: 435–446, 2012. doi:10.1016/j.neuropsychologia.2011.07.013.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268: 889–893, 1995. doi:10.1126/science.7754376.

Sprague TC, Serences JT. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat Neurosci* 16: 1879–1887, 2013. doi:10.1038/nn.3574.

Stigliani A, Weiner KS, Grill-Spector K. Temporal processing capacity in high-level visual cortex is domain specific. *J Neurosci* 35: 12412–12424, 2015. doi:10.1523/JNEUROSCI.4822-14.2015.

Swisher JD, Halko MA, Merabet LB, McMains SA, Somers DC. Visual topography of human intraparietal sulcus. *J Neurosci* 27: 5326–5337, 2007. doi:10.1523/JNEUROSCI.0991-07.2007.

Tanaka K. Neuronal mechanisms of object recognition. *Science* 262: 685–688, 1993. doi:10.1126/science.8235589.

Tanaka K. Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139, 1996. doi:10.1146/annurev.ne.19.030196.000545.

Tarr MJ, Williams P, Hayward WG, Gauthier I. Three-dimensional object recognition is viewpoint dependent. *Nat Neurosci* 1: 275–277, 1998. doi:10.1038/1089.

Turk-Browne NB, Yi DJ, Leber AB, Chun MM. Visual quality determines the direction of neural repetition effects. *Cereb Cortex* 17: 425–433, 2007. doi:10.1093/cercor/bhj159.

Valyear KF, Culham JC, Sharif N, Westwood D, Goodale MA. A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: a human fMRI study. *Neuropsychologia* 44: 218–228, 2006. doi:10.1016/j.neuropsychologia.2005.05.004.

Vaziri-Pashkam M, Xu Y. Goal-directed visual processing differentially impacts human ventral and dorsal visual representations. *J Neurosci* 37: 8767–8782, 2017. doi:10.1523/JNEUROSCI.3392-16.2017.

Vinberg J, Grill-Spector K. Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *J Neurophysiol* 99: 1380–1393, 2008. doi:10.1152/jn.01223.2007.

Wandell BA, Dumoulin SO, Brewer AA. Visual field maps in human cortex. *Neuron* 56: 366–383, 2007. doi:10.1016/j.neuron.2007.10.012.

Ward EJ, Isik L, Chun MM. General transformations of object representations in human visual cortex. *J Neurosci* 38: 8526–8537, 2018. doi:10.1523/JNEUROSCI.2800-17.2018.

Winawer J, Witthoft N. Human V4 and ventral occipital retinotopic maps. *Vis Neurosci* 32: E020, 2015. doi:10.1017/S0952523815000176.

Woloszyn L, Sheinberg DL. Neural dynamics in inferior temporal cortex during a visual working memory task. *J Neurosci* 29: 5494–5507, 2009. doi:10.1523/JNEUROSCI.5785-08.2009.