

Angular Gyrus Responses Show Joint Statistical Dependence with Brain Regions Selective for Different Categories

Mengting Fang,^{1,2} Aidas Aglinskas,¹ Yichen Li,³ and  Stefano Anzellotti¹

¹Department of Psychology and Neuroscience, Boston College, Boston, Massachusetts 02467, ²Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, and ³Department of Psychology, Harvard University, Cambridge, Massachusetts 02138

Category selectivity is a fundamental principle of organization of perceptual brain regions. Human occipitotemporal cortex is subdivided into areas that respond preferentially to faces, bodies, artifacts, and scenes. However, observers need to combine information about objects from different categories to form a coherent understanding of the world. How is this multicategory information encoded in the brain? Studying the multivariate interactions between brain regions of male and female human subjects with fMRI and artificial neural networks, we found that the angular gyrus shows joint statistical dependence with multiple category-selective regions. Adjacent regions show effects for the combination of scenes and each other category, suggesting that scenes provide a context to combine information about the world. Additional analyses revealed a cortical map of areas that encode information across different subsets of categories, indicating that multicategory information is not encoded in a single centralized location, but in multiple distinct brain regions.

Key words: angular gyrus; fMRI; multicategory information; multivariate statistical dependence; neural networks

Significance Statement

Many cognitive tasks require combining information about entities from different categories. However, visual information about different categorical objects is processed by separate, specialized brain regions. How is the joint representation from multiple category-selective regions implemented in the brain? Using fMRI movie data and state-of-the-art multivariate statistical dependence based on artificial neural networks, we identified the angular gyrus encoding responses across face-, body-, artifact-, and scene-selective regions. Further, we showed a cortical map of areas that encode information across different subsets of categories. These findings suggest that multicategory information is not encoded in a single centralized location, but at multiple cortical sites which might contribute to distinct cognitive functions, offering insights to understand integration in a variety of domains.

Introduction

A variety of cognitive tasks require integrating information about entities from different categories. Mechanisms that integrate information about people and artifacts are already at work in early stages of development: infants are more willing to help individuals who intended to give them a desirable toy over those who did not (Dunfield and Kuhlmeier, 2010), and children observing repeated interactions between a person and an inanimate object

infer that the person has feelings of ownership toward that object (Cleroux and Friedman, 2020). In apparent contrast with these observations, visual information about animate and inanimate entities is processed by separate, specialized brain regions (Hecaen and Angelergues, 1962; Sergent et al., 1992; Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Downing et al., 2001; Martin and Chao, 2001); using fMRI, researchers have identified regions responding selectively to faces (Puce et al., 1996; Kanwisher et al., 1997; Gauthier et al., 2000); bodies (Downing et al., 2001; Beauchamp et al., 2003; Schwarzlose et al., 2005); artifacts (Chao et al., 1999; Mahon et al., 2007); and scenes (Epstein and Kanwisher, 1998; Epstein and Baker, 2019).

Investigating the computations through which representations from different category-selective regions are integrated is challenging. Key aspects of these computations might occur at temporal and spatial scales that are beyond the resolution of noninvasive neuroimaging. However, if a brain region integrates information across multiple categories, its responses should be better predicted by the responses across multiple brain regions selective for distinct object categories, than by responses in

Received June 28, 2022; revised Feb. 13, 2023; accepted Feb. 20, 2023.

Author contributions: M.F. and S.A. designed research; M.F. performed research; M.F., A.A., and Y.L. analyzed data; M.F. wrote the first draft of the paper; M.F., A.A., and S.A. edited the paper; M.F. and S.A. wrote the paper.

This work was supported by Boston College startup grant; and National Science Foundation Grant 1943862 to S.A. We thank the researchers who contributed to the *studyforrest* project (Hanke et al., 2016; Sengupta et al., 2016) for sharing their data; and the developers of fmriprep (Esteban et al., 2019) for assistance with the fmriprep preprocessing pipeline.

The authors declare no competing financial interests.

Correspondence should be addressed to Stefano Anzellotti at stefano.anzellotti@bc.edu.

<https://doi.org/10.1523/JNEUROSCI.1283-22.2023>

Copyright © 2023 the authors

regions that are all selective for a same category. We refer to this difference in prediction accuracy as “multicategory dependence” (MCD). Identifying regions that show MCD could serve as a stepping stone to understand integration.

Different hypotheses make distinct predictions about which brain regions might show MCD. According to the “Hub and Spokes” hypothesis (Patterson et al., 2007, 2016; Lambon Ralph et al., 2017), MCD should be observed in the anterior temporal lobe (ATL): a putative semantic hub integrating information in all modalities, for all semantic categories (Patterson et al., 2007). In support of this view, patients with semantic dementia, associated with neurodegeneration that affects the ATL, can present with deficits affecting multiple object categories (Hodges et al., 1992). Furthermore, transcranial magnetic stimulation to the ATL has been reported to delay the naming of both living and nonliving objects (Pobric et al., 2010).

Other studies suggest additional regions that might show MCD. Price et al. (2016) found that transcranial direct current stimulation to the angular gyrus (AG) leads to faster comprehension of semantically meaningful word combinations (“tiny radish”), but not of meaningless combinations (“fast blueberry”), and proposed that this region might be involved in semantic integration. If indeed AG is broadly involved in semantic integration, we might expect its responses to show MCD as well. More recently, a study found stronger responses in the precuneus to sentences including words from multiple categories than to sentences including only words from a single category (Rabini et al., 2021).

These results are not necessarily in conflict. Several processes need to combine representations across different object categories, such as the acquisition of semantic knowledge, the retrieval of episodic memory, social cognition, and decision-making. Therefore, MCD might be observed in multiple regions, each specialized for a different task. In addition, distinct areas might show MCD for lexical stimuli and for visual stimuli.

While most previous studies focused on lexical stimuli, in this study, we investigated MCD while participants watched quasi-naturalistic videos (Hanke et al., 2016). We used multivariate pattern dependence (MVPD) (Anzellotti et al., 2017b; Anzellotti and Coutanche, 2018) based on artificial neural networks (Fang et al., 2022) to identify brain areas in which responses are better predicted by the multivariate response patterns across multiple regions that respond preferentially to different categories, than by the response patterns in regions that are all selective for a same category. In convergence with prior work using lexical stimuli (Price et al., 2016), we found evidence for MCD in the AG. Additional tests revealed a cortical map of areas showing MCD for different subsets of categories: MCD does not occur at a single centralized location, but at multiple distinct sites.

Materials and Methods

Data. The BOLD fMRI responses ($3 \times 3 \times 3$ mm) to the movie *Forrest Gump* were obtained from the publicly available *studyforrest* dataset (<http://studyforrest.org>). Fifteen right-handed participants took part in the study (6 females; age range 21–39 years, mean 29.4 years). The data (acquired with a T2*-weighted EPI sequence) were collected on a whole-body 3 Tesla Philips Achieva dStream MRI scanner equipped with a 32 channel head coil. In addition to the fMRI responses to the movie, the dataset includes an independent functional localizer that was used to identify higher visual areas, such as the fusiform face area (FFA), the extrastriate body area (EBA), and the parahippocampal place area (PPA) (for more details, see Hanke et al., 2016).

During the category localizer, participants were shown 24 unique grayscale images from each of six stimulus categories: human faces, human bodies without heads, small artifacts, houses, outdoor scenes, and phase scrambled images. They were presented with four block-design runs and a one-back matching task. Then, to collect fMRI responses to the movie, the movie stimulus *Forrest Gump* was cut into eight segments, ~15 min long each. All eight movie segments were presented individually to participants in chronological order in 8 separate functional runs.

Preprocessing. Data were first preprocessed using fMRIPrep (<https://fmripred.readthedocs.io/en/latest/index.html>), which is a robust and convenient pipeline for preprocessing of diverse fMRI data. Anatomical images were skull-stripped with ANTs (<http://stnava.github.io/ANTs/>), and FSL FAST was used for tissue segmentation. Functional images were corrected for head movement with FSL MCFLIRT (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MCFLIRT>), and were subsequently coregistered to their anatomic scan with FSL FLIRT. The raw data of 1 participant could not pass the fMRIPrep processing pipeline. For the remaining 14 subjects, we denoised the data with CompCor (Behzadi et al., 2007) using 5 principal components extracted from the union of CSF and white matter.

ROI definition. Four sets of category-selective brain regions were identified using the first block-design run in the category localizer session (Fig. 1a): face-selective regions (occipital face area [OFA], FFA, and face-selective posterior superior temporal sulcus [face STS]), body-selective regions (EBA, fusiform body area [FBA], and body-selective posterior STS [body STS]), artifact-selective regions (medial fusiform gyrus [mFus] and middle temporal gyrus [MTG]), and scene-selective regions (transverse occipital sulcus [TOS], PPA, and retrosplenial cortex [RSC]). Data were modeled with a standard GLM using FSL FEAT (Woolrich et al., 2001), and each seed ROI was defined as a 9 mm radius sphere centered in the peak for its corresponding contrast (i.e., face-selective contrast: faces > bodies, artifacts, scenes, and scrambled images; body-selective contrast: bodies > faces, artifacts, scenes, and scrambled images; artifact-selective contrast: artifacts > faces, bodies, scenes, and scrambled images; scene-selective contrast: scenes > faces, bodies, artifacts, and scrambled images). The house condition from the category localizer was not used for consistency with the other analyses because it was not one of the categories tested in our main analyses. We combined data from both left and right hemisphere for each ROI and then selected the 80 voxels which showed the highest Z value for the contrast between the preferred category and other categories. Finally, the face-, body-, scene-selective regions were identified with 240 voxels each, and the artifact-selective regions consist of 160 voxels.

Additionally, we created a group-average gray matter mask using the gray matter probability maps generated during preprocessing, with a total of 53,539 voxels. This mask was used as the target of prediction ROI in the MVPD analyses (see below).

MVPD network (MVPN). Most research on the interactions between brain regions has focused on the mean responses across voxels in different regions. However, fine-grained patterns of response encode important information that could be lost by spatial-averaging. Over the past two decades, multivariate pattern analysis (Haxby et al., 2001; Norman et al., 2006) of fMRI data has led to progress in the investigation of neural coding at a level of specificity that could not be achieved with univariate analyses (Kriegeskorte et al., 2007; Soon et al., 2008; Nestor et al., 2011; Koster-Hale et al., 2013; Anzellotti et al., 2014). Despite this, relatively few attempts have been made to leverage the potential of multivariate analyses to study brain connectivity (for a recent review, see Anzellotti and Coutanche, 2018). A recent study (Anzellotti et al., 2017b) has developed a technique that investigates the interactions between brain regions in terms of the multivariate relationship between their response patterns (MVPD). MVPD has been shown to offer greater sensitivity than univariate connectivity methods (Anzellotti et al., 2017b), and uses independent training and testing data, thus offering improved robustness to noise.

The original MVPD formulation (Anzellotti et al., 2017b) used principal component analysis to reduce the dimensionality of fMRI response patterns, and subsequently used linear regression as a model of the

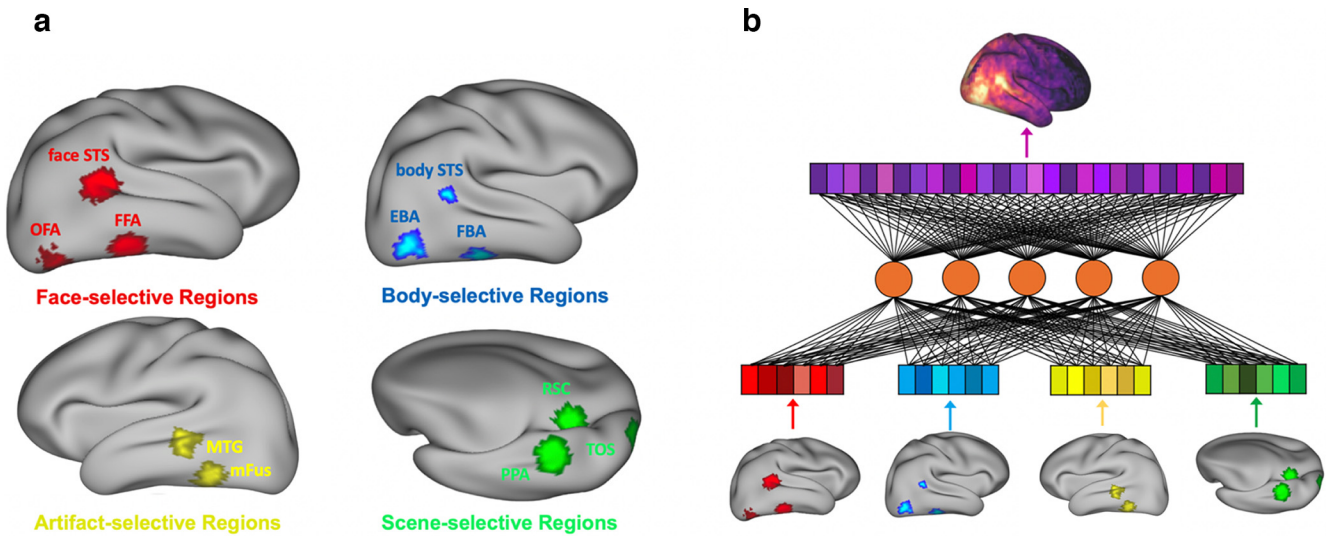


Figure 1. *a*, Category-selective ROIs of one example participant shown on an inflated cortical surface. Red represents face-selective regions (OFA, FFA, and face STS). Blue represents body-selective regions (EBA, FBA, and body STS). Yellow represents artifact-selective regions (mFus and MTG). Green represents scene-selective regions (TOS, PPA, and RSC). *b*, Illustration of an artificial neural network using the response patterns from all category-selective regions combined as input to predict neural activities across the whole brain. The hidden layer is displayed with five nodes for visualization purposes, whereas the networks used for the analyses in this article have hidden layers with 100 nodes.

statistical dependence between brain regions. A more recent version of MVPD used simple artificial neural networks (Anzellotti et al., 2017a) but was limited to a small number of nodes in the hidden layer, and still relied on principal component analysis for dimensionality reduction. Artificial neural networks can themselves perform dimensionality reduction if needed (Fyfe, 1997). In addition, using state-of-the-art software packages for artificial neural networks paves the way for the training of more complex network architectures thanks to the use of general purpose graphic processing units.

To take advantage of these benefits, in this work we extended MVPD to larger artificial neural networks (MVPN), and we implemented it in PyTorch to train the networks on four Tesla V100 graphic processing units. The networks received as inputs multivariate patterns of response in one or more sets of category-selective regions, and were trained to predict the patterns of response in the whole brain.

More formally, let us consider an fMRI scan with m experimental runs. We denote the multivariate time courses in the predictor region by $X_1, \dots, X_m$. Each matrix X_i is of size $n_X \times T_i$, where n_X is the number of voxels in the predictor region, and T_i is the number of time points in the experimental run i . Analogously, $Y_1, \dots, Y_m$ denote the multivariate time courses in the region that is the target of prediction, where each matrix Y_i is of size $n_Y \times T_i$, and n_Y is the number of voxels in the target region.

MVPN was trained with a leave-one-run-out procedure to learn a function f such that

$$Y_{train} = f(X_{train}) + E_{train},$$

where X_{train} and Y_{train} denote data in the predictor region and data in the target region, respectively, during the training runs. E_{train} is the error term. Specifically, for each choice of an experimental run i , data in the remaining runs were concatenated as the training set as follows

$$D_i = \{(X_1, Y_1), \dots, (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), \dots, (X_m, Y_m)\},$$

while data $D_i = \{(X_i, Y_i)\}$ in the left-out run i were used as a testing set.

As a measurement of the multivariate statistical dependence, we calculated the proportion of variance explained between the predictor region and every other voxel in a group-level gray matter mask created from the gray matter probability maps generated during preprocessing. For each voxel j in the target region, variance explained in run i was calculated as follows:

$$\text{varExpl}_i(j) = \max\left\{0, 1 - \frac{\text{var}(Y_i(j) - f_j(X_i))}{\text{var}(Y_i(j))}\right\},$$

where X_i is the time course in the predictor region for run i , and $f_j(X_i)$ is the prediction by MVPN for voxel j . The values $\text{varExpl}_i(j)$ obtained for the different runs $i = 1, \dots, m$ were averaged, yielding $\text{varExpl}(j)$.

Exploring MCD sites. To identify brain regions with MCD, we used the 8 experimental runs during which participants watched the movie *Forrest Gump*. The runs were used as separate folds for cross validation. In a first analysis, we used MVPN to calculate the variance explained in each gray matter voxel using each of the four category-selective regions (face, body, artifact, and scene) individually. In a second analysis, we combined all category-selective regions jointly as inputs of MVPN to predict the fMRI responses of each voxel in the gray matter mask.

If a target region only encodes information from one category of objects, using the responses from regions selective for multiple categories as predictors should not improve over using only the responses in the regions from the one category yielding the best predictions. Instead, if the responses in a target brain region are predicted significantly better by a model including all category-selective regions combined, than by the best of the category-selective regions in isolation, we can conclude that multiple category-selective regions have unique contributions to the overall statistical dependence with that target region. We refer to this approach as the “combined-minus-max” approach.

To make things more precise, for each voxel j , we can denote with $\text{varExpl}_{all}(j)$ the variance explained by MVPN using as input the responses in all category-selective regions (Fig. 1b), and with $\text{varExpl}_{max}(j)$ the variance explained by MVPN using as input the responses in regions corresponding to the single best-predicting category. We then calculated for each voxel the difference as follows:

$$\Delta\text{varExpl}(j) = \text{varExpl}_{all}(j) - \text{varExpl}_{max}(j),$$

and used it as a metric to evaluate whether a voxel exhibits MCD by jointly encoding information across multiple regions selective for different categories. For each subject, we calculated the difference $\Delta\text{varExpl}$ for each voxel in the gray matter mask, and we computed statistical significance across participants with statistical nonparametric mapping using the SnPM extension for SPM (<http://warwick.ac.uk/snpm>), obtaining pseudo- t statistics. Regions consisting of voxels with statistically significant pseudo- t values were defined as candidate MCD sites.

Using the combined-minus-max approach alone, however, we cannot rule out the possibility that the better predictive accuracy of the combined MVPN model in candidate MCD sites is simply because of the increased number of voxels: control analyses are needed. To test whether combined-minus-max effects are driven by differences in the number of voxels, we further conducted two control analyses. In the first control analysis (Control 1), we used voxels extracted from the primary visual cortex (V1) as predictors. Specifically, we first identified the V1 region using the calcarine cortex mask across both left and right hemispheres from WFU_PickAtlas (https://www.nitrc.org/projects/wfu_pickatlas/). Next, we randomly selected 880 nonoverlapping voxels from V1 and randomly divided them into four nonoverlapping groups, such that these groups have the same number of voxels as our four sets of category-selective regions. We then used responses from these groups of voxels in V1 to run a control analysis matched in the number of voxels to the original combined-minus-max analysis. Specifically, we ran MVPN analyses using as inputs the responses from each of the four V1 groups in isolation, and then using responses in all groups combined. Next, we performed a combined-minus-max analysis and computed the statistical significance of $\Delta \text{varExpl}$ across participants. If a region we previously identified as an MCD site also shows statistical significance in the control analysis, the observed results might be driven by the number of voxels rather than by MCD, and therefore the region is removed from the set of candidate MCD sites.

In the second control analysis (Control 2), we followed an analogous procedure, except that voxels used as predictors were extracted from face-selective regions instead of V1. Specifically, we defined each face ROI (FFA, face STS, OFA) within an 18-mm-radius sphere centered in the peak for the face contrast using the first localizer run. For each subject, we combined data from the two hemispheres and took the union of all three face ROIs. Next, to include only voxels with preferential responses to faces, we excluded voxels from the union where the Z value for the face contrast is not the highest compared with the ones for all other three contrasts (i.e., body, artifact, scene). We then selected the 880 voxels that yielded the highest Z value, and randomly divided them into four nonoverlapping groups in which the number of voxels matches that in the original four sets of category-selective regions. We then used responses from these groups of voxels as predictors in the second control analysis, following the same procedure used for the original data and for the first control analysis.

It is important to note that there is some overlap between adjacent category-selective ROIs (e.g., face-selective FFA and body-selective FBA); indeed, this overlap has been reported in several previous studies (Schwarzlose et al., 2005; Peelen et al., 2006). However, the overlap actually makes our analysis more stringent when seeking for MCD sites. Suppose that one voxel in the face-selective ROI overlaps with the body-selective ROI and encodes information about both face and body. When face is selected as the best-predicting category, the “max model” using face-selective ROIs as single predictors is more powerful than it is supposed to be since it also carries some body information. As a consequence, the difference $\Delta \text{varExpl}$ would be smaller, making regions jointly encoding information about face and body fail to survive our combined-minus-max analysis. The more voxels with overlapping category selectivity, the harder it would be to find MCD sites. In contrast, the nonoverlapping selection of the control V1 voxels makes it easier to find regions showing combined-minus-max effects in the control analysis, thus excluding voxels showing significant results in the control analysis further increases the rigor of our findings.

Neural network architectures. To make sure that the results do not depend on choosing a very specific neural network architecture, we trained MVPN using three different neural network architectures, and identified regions that showed significant effects in the combined-minus-max analysis in all architectures. All network architectures were linear, because previous studies did not find an advantage for using nonlinear neural networks in MVPD (Poskanzer et al., 2021). All architectures used 100 hidden nodes in each hidden layer. The first architecture was a one-layer feedforward network. Since previous studies have shown that deeper networks can approximate the same classes of functions as

shallower networks using fewer parameters (Mhaskar et al., 2017), we then tested a second, deeper architecture: a five-layer feedforward network. Finally, a challenge encountered in training deep neural networks is the vanishing-gradient problem (Hochreiter, 1998): as the gradient of the loss function is back-propagated across multiple layers, the weight updates can become progressively smaller, affecting learning in early layers of the network. For this reason, we also tested a 5-Layer DenseNet (Huang et al., 2017). The DenseNet architecture includes connections that bypass multiple layers, enabling more direct backpropagation of the loss function to early layers.

All architectures were trained over 5000 epochs using stochastic gradient descent on a mean squared error loss, with a learning rate of 0.001 and a momentum of 0.9. We used a batch size of 32, and batch normalization was applied to the inputs of each layer. The original code implemented in PyTorch is available at <https://github.com/scnlab/PyMVPD>. More details are provided in the PyMVPD toolbox (Fang et al., 2022).

Pairwise and three-way MCD analysis. Multicategory dependence might not only occur across all four object categories. Instead, it is possible to find MCD effects in other brain areas when considering a pair or a triplet of category-selective regions. To investigate MCD in more depth, we implemented the combined-minus-max approach on different pairs and triplets of categories, and conducted the following pairwise and three-way MCD analysis.

In the pairwise MCD analysis, we used responses in each pair of category-selective regions (face and body, face and artifact, face and scene, body and artifact, body and scene, artifact and scene) as MVPN inputs to calculate the variance explained in each gray matter voxel. We then calculated $\Delta \text{varExpl}_{\text{pair}}$ for each voxel by subtracting the variance explained predicted by responses in one pair of category-selective regions from the variance explained predicted by responses in regions selective for the better-predicting category within that pair. We used the average $\Delta \text{varExpl}_{\text{pair}}$ across different runs as a metric to identify candidate brain areas that jointly encode information from two category-selective regions in each pair. A control analysis was then performed following the same approach described previously (in the section, Exploring MCD sites): we repeated the combined-minus-max analysis with randomly selected subsets of V1 voxels to exclude regions that were simply better predicted by the combined model because of the greater number of voxels.

In the three-way MCD analysis, responses in each triplet of category-selective regions (face and body and artifact, face and body and scene, face and artifact and scene, body and artifact and scene) were used as inputs to our MVPN model. The $\Delta \text{varExpl}_{\text{triplet}}$ for each voxel was calculated by subtracting the variance explained predicted by responses in one triplet of category-selective regions from the variance explained predicted by responses in regions selective for the best-predicting category within that triplet. We took the average $\Delta \text{varExpl}_{\text{triplet}}$ across runs as a metric to identify candidate areas that jointly encode information from three category-selective regions in each triplet. Finally, a control analysis was performed (following the approach described above) to rule out voxel-driven regions that did not contribute to the three-way MCD.

Representational similarity analysis. We used representational similarity analysis (Kriegeskorte et al., 2008b; Diedrichsen and Kriegeskorte, 2017) to study the representational geometry of MCD sites and to investigate how they differ from the representational geometry in category-selective regions. Representational similarity analysis is a multivariate method that calculates the pairwise dissimilarities between multivariate activation patterns in a brain region, yielding a representational dissimilarity matrix (RDM). In this study, we used the correlation distance (one minus the Pearson correlation) as dissimilarity metric. Before calculating the correlation distance, for each subject the average response pattern across all categories was subtracted from the data (Friston et al., 2019).

Since we used the first run of the category localizer to identify category-selective ROIs, we used the remaining three runs of the localizer for the following analyses. For the four sets of category-selective ROIs (e.g., face ROIs) and each category-selective ROI separately (e.g., FFA), response patterns in each of the regions in the set were concatenated, yielding four RDMs for each of the 14 participants. For each of the

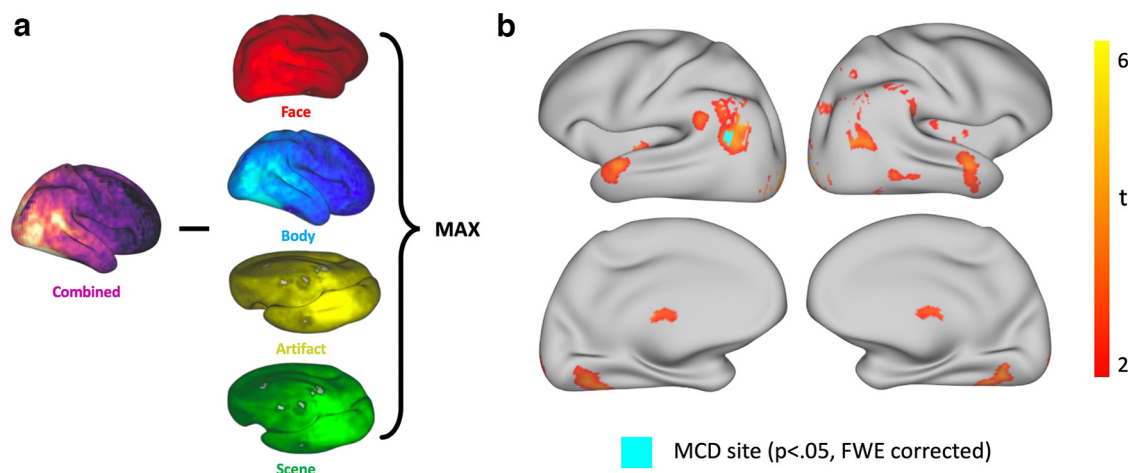


Figure 2. *a*, The MVPN combined-minus-max approach to detect MCD sites jointly encoding information from multiple category-selective regions. *b*, Statistical *t* maps of the 5-Layer Dense MVPN model computed from Δ varExpl across subjects. Δ varExpl is the voxelwise difference between the variance explained predicted by the combined model using the response patterns from all category-selective regions as input, and the variance explained predicted by the model using the response patterns from regions selective for the single best-predicting category as input. The SnPM threshold corrected at $p < 0.05$ FWE is 5.11. *t* maps are thresholded within the range [2, 6] for visualization purposes. Aqua represents the identified MCD site within left AG.

candidate MCD sites, we first defined a 9 mm radius sphere (123 voxels) centered in the peak of its SnPM *t* contrast map obtained from the combined-minus-max analysis using the best-performing MVPN model (i.e., 5-Layer Dense MVPN). Next, we selected 50 voxels with the highest *t* values for each candidate MCD site and calculated the RDMs on the new set of voxels. RDMs were then averaged across participants, and the SEM was calculated as a measure of the intersubject variability of correlation distance for each pair of stimuli. We used radar charts to visualize the within-region and between-region differences in dissimilarity. Because RDMs are symmetric about a diagonal of zeros, we only extracted the upper (or equivalently the lower) triangle of the matrices for radar chart visualization. We also performed a set of statistical analyses to quantitatively compare the RDM patterns between the MCD site and the category-selective ROIs.

Results

Identification of MCD sites

To identify candidate sites that jointly encode information from regions selective for different object categories, we calculated an MCD index for each voxel in the brain (mathematical details are reported in Materials and Methods). The index was computed as the difference between the proportion of variance explained by a model using all category-selective regions as predictors (henceforth the “combined model”), and the proportion of variance explained by a model using regions selective for the single best predicting category as predictor (henceforth, the “max model,” Fig. 2*a*). To ensure the robustness of the results across different neural network architectures, we computed this index using 1-Layer, 5-Layer, and 5-Layer Dense MVPN architectures. The results showed that the 5-Layer Dense network outperformed the 1-Layer network and the 5-Layer network without dense connections in terms of the combined variance explained in each candidate MCD site. For each neural network architecture, we used a group-level analysis to identify voxels in which the MCD index was significantly >0 ($p < 0.05$ corrected with SnPM). Then, we selected voxels where the MCD index was significantly >0 for all three network architectures as candidate sites for MCD.

The nature of this analysis is such that the number of voxels used as predictors in the combined model is greater than the number of voxels used as predictors in the max models. Therefore, we performed two control analyses using the best-predicting 5-Layer Dense MVPN model to rule out the possibility

that positive values of the MCD index might be driven by differences in the number of input voxels. In the control analyses, we repeated the same procedure we used to identify candidate sites for MCD, but we replaced the category-selective regions with control subregions consisting of voxels randomly sampled from primary visual cortex (V1) in Control 1 and from face-selective regions in Control 2. Importantly, the control subregions were matched in number of voxels to the category-selective regions. To make the control more stringent, we adopted a less conservative threshold for the MCD index. Specifically, we identified voxels whose pseudo-*t* values (computed with a nonparametric test using SnPM) in the control analyses were significantly >0 without applying multiple comparison correction (pseudo- $t_{(13)} = 1.77$, $p < 0.05$, one-sided); any MCD effects in these voxels were discarded as possible artifacts of the number of voxels.

In a region within left AG (Fig. 2*b*, peak MNI coordinates: -57 , -69 , 21), the MCD index was significantly >0 for all neural network architectures tested (lowest pseudo- $t_{(13)} = 5.20$, $p < 0.05$). At the same time, this region did not show significant effects in the control (indeed, in this region all pseudo-*t* values in Control 1 analysis were negative: highest pseudo- $t_{(13)} = -1.27$; and all pseudo-*t* values in Control 2 are below the significance threshold: highest pseudo- $t_{(13)} = 0.07$; for a complete report of the pseudo-*t* values for all voxels in two control analyses, see Table 1). The location of this region within AG was confirmed with Neurosynth (Yarkoni et al., 2011). The AG has been previously found to integrate information across multiple sensory modalities (Bonnici et al., 2016). Our results indicate that it also jointly encodes information across multiple object categories. Three other regions (two in the vermis, one in occipital gyrus) showed significant effects in the MCD analysis, but also in the Control 1 or Control 2 analysis; thus, they were removed from subsequent analyses (for details, see Table 1).

Representational geometry of the MCD site: AG

Having identified an MCD site in the left AG, we asked whether this region inherits its representational similarity structure from category-selective regions, showing more similar responses to pairs of objects that are both animate or both inanimate. To address this question, we used response patterns to different object categories during the functional localizer to calculate

Table 1. Candidate MCD sites that jointly encode information across all four category-selective ROIs^a

Candidate MCD sites	MNI coordinates	5-Layer Dense (<i>t</i>)	1-Layer (<i>t</i>)	5-Layer (<i>t</i>)	Control 1 (<i>t</i>)	Control 2 (<i>t</i>)	No. of voxels in each site
AG	(−57, −69, 21)	5.72	7.51	6.57	−1.27	−0.20	8
	(−57, −69, 24)	5.20	7.30	6.81	−1.81	0.02	
	(−57, −66, 18)	5.28	6.07	5.25	−1.69	−0.43	
	(−57, −66, 21)	5.53	7.16	6.09	−2.86	−0.59	
	(−54, −72, 21)	5.58	7.69	6.48	−1.62	0.07	
	(−54, −69, 21)	5.61	8.03	7.31	−2.54	−0.38	
	(−54, −69, 24)	5.30	8.24	8.06	−2.79	−0.04	
	(−54, −66, 24)	5.16	7.81	7.17	−3.68	0.05	
Vermis 1	(−3, −75, −9)	5.37	7.98	7.98	5.59	2.98	7
	(−3, −72, −12)	5.41	7.43	6.36	3.67	2.95	
	(−3, −72, −5)	5.85	8.07	7.75	5.97	2.32	
	(−3, −69, −5)	5.40	7.89	6.59	5.40	2.02	
	(0, −75, −12)	5.26	7.85	7.30	5.11	3.81	
	(0, −75, −9)	5.52	8.18	8.19	5.76	3.23	
	(0, −72, −12)	5.72	7.90	6.64	3.86	3.28	
	(3, −72, −5)	5.11	7.17	6.15	5.11	1.08	
Occipital gyrus	(24, −102, 8)	5.26	6.14	5.35	3.83	1.97	4
	(24, −99, 8)	5.47	7.46	6.64	3.75	1.20	
	(24, −99, 11)	5.52	6.89	6.33	3.90	1.38	
	(27, −99, 8)	5.62	6.85	6.05	3.49	1.87	

^aWe selected candidate MCD sites from brain regions with significant *t* values ($p < 0.05$, FWE-corrected) computed from the combine-minus-max analysis across all three network architectures: 1-Layer MVPN, 5-Layer MVPN, and 5-Layer Dense MVPN. The location of each site is expressed in MNI coordinates. For each voxel in the candidate MCD sites, we also computed the corresponding *t* value in two control analyses using the best-performing 5-Layer Dense MVPN. Control 1 used voxels sampled from V1 as predictors, and Control 2 used voxels sampled from face-selective regions as predictors.

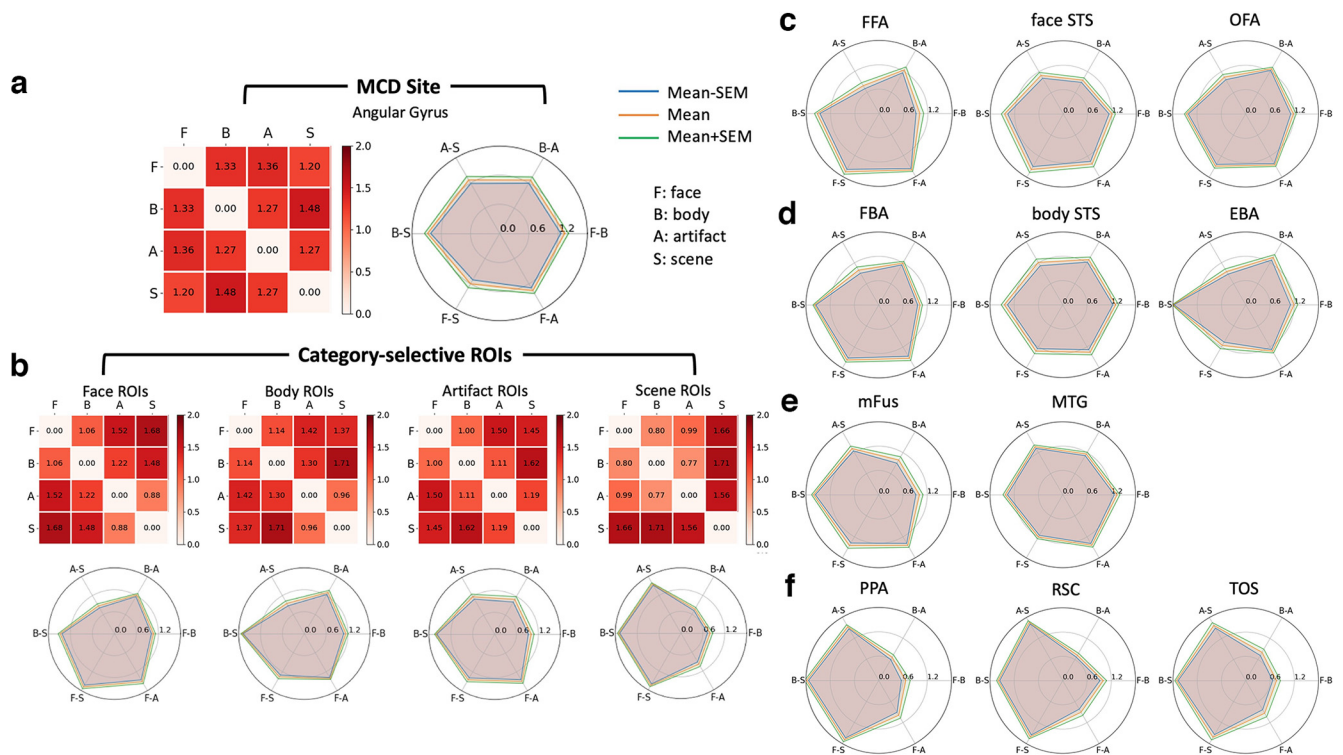


Figure 3. Using RDMs and radar charts to discriminate between MCD site and category-selective ROIs. **a**, RDMs (top) and radar charts (bottom) for average brain activities across subjects in the MCD site: the left AG. **b**, RDMs (top) and radar charts (bottom) for average brain activities across subjects in four category-selective ROIs: face-selective ROIs, body-selective ROIs, artifact-selective ROIs, and scene-selective ROIs. **c–f**, RDMs for average brain activities across subjects in three face-selective ROIs: FFA, face STS, and OFA; in three body-selective ROIs: FBA, body STS, and EBA; in two artifact-selective ROIs: mFus and MTG; and in three scene-selective ROIs: PPA, RSC, and TOS. In each radar chart, the middle boundary represents the average representational dissimilarity across subjects, the inner boundary represents the average representational dissimilarity minus the SEM, and the outer boundary represents the average representational dissimilarity plus SEM.

RDMs (Kriegeskorte et al., 2008b) for the MCD site and for each set of category-selective regions (Fig. 3).

First, we replicated the finding that representations in inferior temporal cortex are organized by animacy (Kriegeskorte et al.,

2008a; Bracci et al., 2019). We found that in face-, body-, and artifact-selective ROIs, object pairs that were both animate (faces and bodies) or both inanimate (artifacts and scenes) elicited more similar responses than object pairs with different animacy

(faces and scenes, faces and artifacts, or bodies and scenes) (Fig. 3*b*). This similarity structure leads to asymmetrical radar charts for the category-selective ROIs (Fig. 3*b*). In scene-selective ROIs, response patterns to scenes showed high dissimilarity from the responses to all other object categories, while the dissimilarity between pairs of nonscene categories was low (Fig. 3*b*). However, these effects were not observed in the left AG MCD site (Fig. 3*a*). By contrast, in the MCD site, all pairs of categories showed comparable dissimilarity (Fig. 3*a*).

Next, we performed a Region $\times$ Category-pair ANOVA to analyze how different regions and category pairs influence the RDM patterns. Specifically, we created a representational dissimilarity vector (RDV) by extracting values of the upper triangle of each RDM and performed the two-way ANOVA between RDVs in different regions. The results revealed that there was a statistically significant interaction between the effects of Region and Category-pair ($F_{(20,390)} = 12.11$, $p = 3.32 \times 10^{-30} < 0.05$). Simple main effects analysis showed that Category-pair had a statistically significant effect on RDVs ($F_{(5,390)} = 51.93$, $p = 3.24 \times 10^{-42} < 0.05$), whereas the effect of Region did not show statistical significance ($F_{(4,390)} = 1.28$, $p = 0.27$).

To further quantify the dissimilarity patterns for different category-pairs in the AG MCD site and the four category-selective regions, we conducted a one-way ANOVA on the RDVs for each region across all subjects. The results showed that there is no significant difference in the RDV values within the AG MCD site ($F_{(5,78)} = 1.47$, $p = 0.21 > 0.05$), but significant differences across RDV values in all category-selective regions (face ROIs: $F_{(5,78)} = 27.55$, $p = 5.86 \times 10^{-16} < 0.05$; body ROIs: $F_{(5,78)} = 19.78$, $p = 1.12 \times 10^{-12} < 0.05$; artifact ROIs: $F_{(5,78)} = 12.17$, $p = 1.00 \times 10^{-8} < 0.05$; scene ROIs: $F_{(5,78)} = 63.47$, $p = 4.50 \times 10^{-26} < 0.05$). The *post hoc* one-sided *t* tests with Bonferroni correction for multiple comparisons further showed that, in all four category-selective regions, body-scene pair is significantly more dissimilar than body-artifact pair (face ROIs: $p = 0.027$; body ROIs: $p = 0.000$; artifact ROIs: $p = 0.000$; scene ROIs: $p = 1.887 \times 10^{-13}$) and face-body pair (face ROIs: $p = 0.000$; body ROIs: $p = 1.692 \times 10^{-7}$; artifact ROIs: $p = 3.534 \times 10^{-6}$; scene ROIs: $p = 2.834 \times 10^{-10}$). We then performed a one-sided *t* test to compare the SD of RDV values in the AG MCD site to the average SD of RDV values in four category-selective regions. The results showed that the AG MCD site has a significantly lower SD of RDV values ($t_{(13)} = -1.89$, $p = 0.03 < 0.05$), which demonstrates that the more symmetrical radar chart we observed in Figure 3*a* for the AG MCD site is significantly different from those observed for the category-selective regions.

In addition, we calculated RDMs for each category-selective region separately and plotted radar charts using the upper triangle of the matrices. Similarly, the asymmetrical radar charts still show the animate-inanimate distinction in each of the regions selective for faces, bodies, and artifacts (Fig. 3*c–e*). In each of the scene-selective regions, response patterns between nonscene

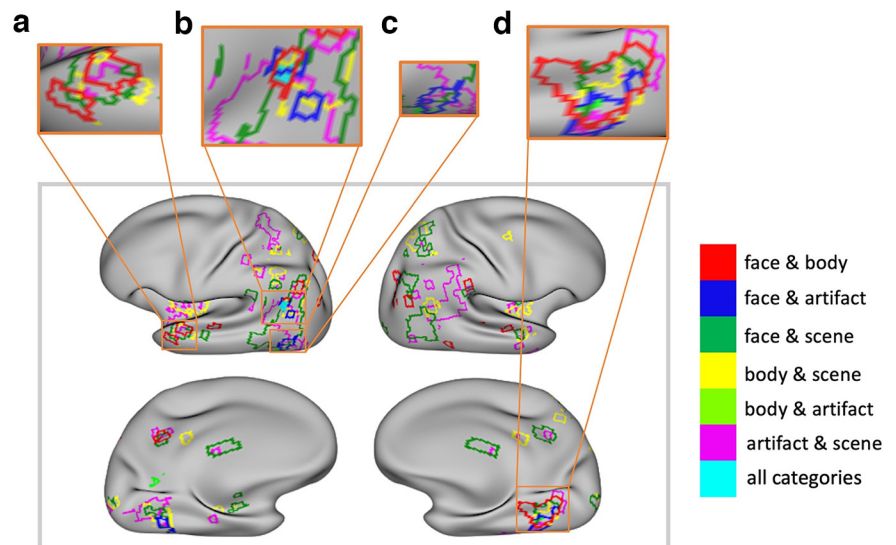


Figure 4. Cortical map of pairwise MCD. The outlines of brain areas with significant pairwise MCD effects ($p < 0.05$, FWE-corrected) but not in the control analysis ($p < 0.05$, uncorrected) are colored in red (face & body), blue (face & artifact), green (face & scene), yellow (body & scene), lime (body & artifact), and purple (artifact & scene). Aqua represents the left AG MCD site identified across all four categories. The four orange boxes on the top panel represent the enlarged patterns of pairwise MCD in the dorsal ATL, AG, left MTG, and the ventral occipitotemporal cortex. In the main figure, the AG MCD site is set as the foreground, covering other regions with pairwise MCD effects. In orange box *b*, the AG MCD site is set as the background so as to show the overlaps between the different pairwise MCD sites. *a–d*, close-up views of selected regions.

category pairs are more similar than those between pairs including the scene category (Fig. 3*f*). We then compared the RDMs from each category-selective region to the RDMs in the AG MCD site for each subject using Kendall's tau correlation between the matrices upper triangle (because they are symmetrical). The results showed that RDMs of all category-selective regions are weakly positively correlated with that of the AG MCD site, except for FBA and body STS. Among these category-selective regions, MTG and TOS have the highest Kendall's tau values with AG, indicating that their representational geometry is most similar to the representational geometry in AG.

Pairwise and three-way MCD

Combining information about faces and bodies could facilitate the recognition of individuals and their actions, while combining information about artifacts and scenes could help to search for objects in their habitual contexts. Several distinct cortical sites might encode information about multiple object categories jointly, in the service of distinct cognitive functions. Can we identify brain regions that show MCD for specific subsets of categories? To address this question systematically, we calculated the MCD index for all pairs and triplets of categories. The resulting SnPM pseudo-*t* contrast maps were thresholded at $p < 0.05$ (FWE-corrected). For each pair and triplet of categories, we also ran control analyses using subregions of V1 matched in terms of the number of voxels, following the same procedure we used in the analysis with all four categories (for details, see Materials and Methods).

We plotted the outlines of sites showing significant MCD of pairs of categories on an inflated cortical surface (Fig. 4). This analysis revealed that the left AG MCD site identified in the first portion of this study is surrounded by a cortical map of regions showing significant pairwise MCD for different pairs of categories. This cortical map includes regions showing MCD effects between scenes and each of the other three categories separately,

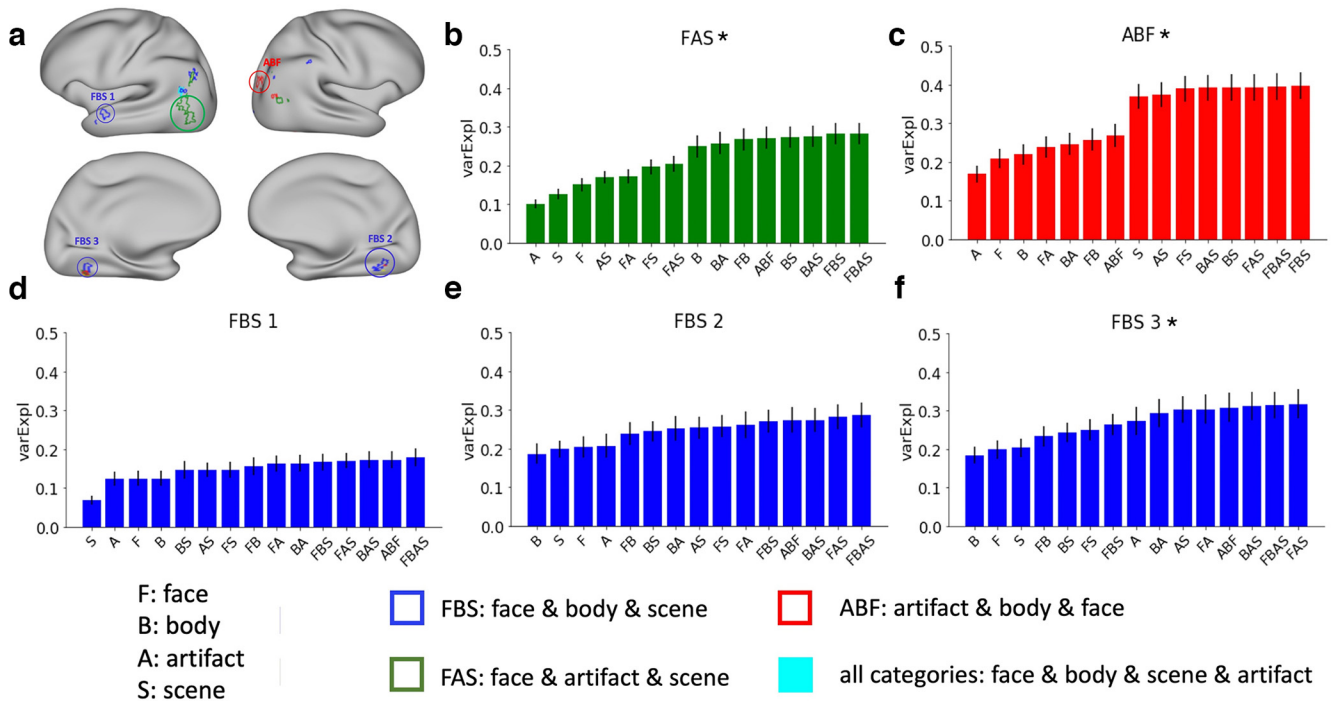


Figure 5. *a*, Cortical sites showing three-way MCD. *b–f*, Predictive accuracy across three-way MCD regions using response patterns from different sets of category-selective regions as inputs (e.g., F, face-selective regions; FB, face- and body-selective regions; ABF, artifact-, body-, and face-selective regions) with the 5-Layer Dense MVPN model. Error bars indicate SEM. *Regions showing the triplet subordinate phenomenon.

suggesting that scenes might provide a context to combine information about the world (Fig. 4; face and scene, in green; body and scene, in yellow; artifact and scene, in purple).

In addition to this cortical map of MCD in the left AG, the analysis revealed other cortical areas displaying evidence for MCD between pairs of categories (a magnified view of these areas is shown in Fig. 4, top). First, a region in dorsal ATL showed evidence of overlapping face-body, face-scene, and body-scene MCD (shown in Fig. 4, leftmost magnification box). Second, pairwise MCD effects were observed in the left MTG (shown in Fig. 4, third magnification box from the left). Last, pairwise MCD effects of responses selective for faces and bodies, faces and artifacts, faces and scenes, and bodies and scenes were observed in a portion of the ventral occipitotemporal cortex (shown in Fig. 4, fourth magnification box from the left). Overlap between face-body and face-scene MCD effects was observed in the precuneus (Fig. 4). Overall, MCD effects showed a substantial degree of bilateral symmetry.

Consistent with the overlap in pairwise MCD results, we also found significant three-way MCD effects in ATL, MTG, and ventral occipitotemporal cortex (Fig. 5*a*). To quantify the contribution of different categories to predicting responses in these regions, we computed the variance explained by the best-predicting 5-Layer Dense MVPN model using all combinations of individual, pairwise, and three-way sets of categories (Fig. 5*b–f*). While in three-way MCD sites the combination of three categories better predicts fMRI responses than the best single category among those three, it is still possible that the other category outside the triplet can predict the responses better than the three categories in the triplet combined. We marked regions showing such “triplet subordinate” phenomenon with asterisks, including region FAS (Fig. 5*b*), ABF (Fig. 5*c*), and FBS 3 (Fig. 5*f*). Compared with other “triplet dominant” regions (FBS 1 and FBS 2 in Fig. 5*d,e*), these triplet subordinate regions

achieved higher variance explained on average across subjects by response patterns from the fourth left-out category (the “dominant category”) than by any individual category within the triplet or their combinations. One possible interpretation for this unintuitive finding is that such regions might predominantly represent one specific category, but might also represent associated information from the subordinate categories. To illustrate this idea more clearly, imagine seeing young students and teachers (faces and bodies), blackboards, and textbooks (artifacts): you could infer that this is probably a classroom. The ABF “triplet subordinate” region might predominantly represent scenes, but might also encode information about objects from multiple nonscene categories that are typically associated with those scenes. In sum, the pairwise and three-way analyses show that, in addition to the MCD site in AG, partial MCD effects can be observed at multiple other sites, including ATL and MTG.

Discussion

This study aimed to identify cortical sites that jointly encode responses across multiple category-selective brain regions. Using multivariate analyses of the interactions between brain regions, we identified an MCD site in the left AG, surrounded by a cortical map of regions showing pairwise and three-way MCD effects. Combining information from multiple object categories is needed for a variety of cognitive functions: what might be the functional role of the multicategory representations we identified in AG? The AG is a multimodal integration area (Bonnici et al., 2016) that has been implicated in a variety of cognitive processes. Specifically, the AG has been implicated in semantic memory (Geschwind, 1972; Binder et al., 2009), episodic memory (Wagner et al., 2005; Berryhill et al., 2007), and bottom-up attention (Corbetta and Shulman, 2002). More

recently, it has been proposed that AG might play a key role for the representation of event semantics (Binder and Desai, 2011) and schemas (Wagner et al., 2015), and that it might serve as a temporary buffer that integrates spatiotemporal information (Humphreys et al., 2021).

Previous studies reported evidence of a role of AG in combinatorial semantics, showing that AG responds more to meaningful word combinations than to nonmeaningful ones, and that the degree of atrophy in AG is inversely related to combinatorial performance in patients with semantic dementia (Price et al., 2015). In addition, transcranial direct current stimulation of the left AG selectively affects the speed of comprehension of semantically meaningful word combinations (Price et al., 2016). These observations are consistent with the sensitivity of AG to combinatorial spatial and nonspatial patterns (Wagner et al., 2015). AG was also found to respond differentially during the integration of multiple cues when making judgments about ambiguous words (Lanzoni et al., 2020), even when the cues were from different categories (facial expressions and locations). Our results directly demonstrate the statistical dependence of AG responses on responses across regions selective for multiple different categories during the observation of complex, quasi-naturalistic videos (Fig. 2*b*).

In addition, the results reveal a cortical map of regions in and around AG whose responses depend on different pairs of category-selective networks (Fig. 4). This cortical map is characterized by regions jointly encoding scene-selective responses with responses selective for one other category (scenes and artifacts, scenes and bodies, scenes and faces), suggesting that scenes might provide a context to combine representations of objects from multiple different categories into complex situations and events. This proposal is consistent with the hypothesis that AG might be involved in event representations (Binder and Desai, 2011); and that it might serve as a temporary buffer (Humphreys et al., 2021), as the locations and interactions of multiple objects and their functional significance within an event (i.e., whether they play a causal role) can vary from situation to situation.

In addition to the AG, the ATL has been implicated in the integration of information across multiple modalities and categories (Patterson et al., 2007). In addition, multivariate analyses of fMRI data suggest that ATL represents conceptual knowledge about objects (Peelen and Caramazza, 2012), and integrates multiple features of an object (Coutanche and Thompson-Schill, 2015). In the present study, we identified a region in dorsal ATL showing pairwise MCD effects of faces and bodies, faces and scenes, and bodies and scenes. However, we did not find significant MCD effects in ATL in the analysis with all four categories.

In addition to AG and ATL, we observed MCD effects in the left MTG between artifacts and scenes, artifacts and bodies, and artifacts and faces. Unlike AG, this region did not show MCD effects in the analysis with all four categories. Left MTG has been previously implicated in the representation of artifacts (Beauchamp et al., 2003; Beauchamp and Martin, 2007; Mahon et al., 2007). In addition, overlap between responses to hands and tools has been reported in the vicinity of this area (Bracci et al., 2012). The findings in the present study suggest that left MTG might encode representations not only of artifacts in isolation, but also of their interactions with the human body, including the face (i.e., flatware, glasses, hats). MCD between some pairs of categories was also observed in a portion of ventral occipitotemporal cortex (Fig. 4). This observation converges with recent reports of a heteromodal semantic hub in the fusiform gyrus

(Forseth et al., 2018; Qin et al., 2021) to suggest that some integration might already occur in posterior temporal cortex, but additional studies are needed to clarify the functional implications of this finding.

In addition to AG and ATL, another brain region often implicated in semantic integration is the precuneus (Binder et al., 2009). Indeed, a recent study used linguistic stimuli to study integration between words for objects from different categories, and found that precuneus, but not AG, shows stronger responses to sentences, including the names of entities from multiple different categories than to sentences, including only entities from a same category (Rabini et al., 2021). While in our study participants watched complex, quasi-naturalistic stimuli (as opposed to processing sentences), our results are consistent with and complementary to the findings of Rabini et al. (2021). The hypothesis that AG encodes complex spatiotemporal relations between entities predicts that it would integrate information across multiple categories when entities from different categories are present within a situation, but also when only entities from one category are present. In the study by Rabini et al. (2021), both types of sentences described events featuring spatiotemporal interactions between multiple entities. Therefore, the hypothesized role of AG would not predict a difference in the overall amount of response between the two sentence types.

It is possible that precuneus might also contribute to multicategory information representation. Indeed, we did find MCD effects in the precuneus when using two of the three artificial neural network architectures (i.e., the 1-Layer and the 5-Layer network). However, the MCD effects were not observed when using the 5-Layer Densely connected architecture. Additional studies will be needed to test the robustness of MCD effects in the precuneus during the perception of complex dynamic videos, and to rule out that these effects are specific to language processing.

More work remains to be done to fully understand how information from multiple category-selective brain regions is encoded. In this study, we investigated MCD by searching for MCD sites whose responses are better predicted by the multivariate response patterns across multiple regions that respond preferentially to different categories, than by the response patterns in regions that are all selective for a same category. We identified a cortical map of brain regions whose responses are better predicted by the joint responses across multiple category-selective regions. However, several questions remain open. Additional research will be needed to investigate what kind of computations occurs within these sites, and to determine whether they use information about objects from different categories to compute representations of their relationships and interactions. Furthermore, different MCD sites might support distinct cognitive functions. Finally, the methods used in this study are correlational: future investigations could evaluate the causal relationship between responses in category-selective regions and MCD sites using techniques, such as combined transcranial magnetic stimulation-fMRI.

In conclusion, we identified a region in the left AG whose responses are better predicted by the joint patterns of activity across multiple category-selective networks. This region is surrounded by a cortical map of areas showing MCD effects for specific pairs of object categories, including in particular pairwise effects for scenes and every other category tested. In addition, we found evidence for MCD between some subsets of categories in the dorsal ATL and in left MTG. Together, these results show

that multicategory information is not encoded in a single centralized site. Instead, MCD occurs in multiple cortical areas, which might in turn support distinct cognitive functions.

References

- Anzellotti S, Coutanche MN (2018) Beyond functional connectivity: investigating networks of multivariate representations. *Trends Cogn Sci* 22:258–269.
- Anzellotti S, Fairhall SL, Caramazza A (2014) Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex* 24:1988–1995.
- Anzellotti S, Fedorenko E, Kell AJ, Caramazza A, Saxe R (2017a) Measuring and modeling nonlinear interactions between brain regions with fMRI. *bioRxiv* 074856. <https://doi.org/10.1101/074856>.
- Anzellotti S, Caramazza A, Saxe R (2017b) Multivariate pattern dependence. *PLoS Comput Biol* 13:e1005799.
- Beauchamp MS, Martin A (2007) Grounding object concepts in perception and action: evidence from fMRI studies of tools. *Cortex* 43:461–468.
- Beauchamp MS, Lee KE, Haxby JV, Martin A (2003) fMRI responses to video and point-light displays of moving humans and manipulable objects. *J Cogn Neurosci* 15:991–1001.
- Behzadi Y, Restom K, Liu J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37:90–101.
- Berryhill ME, Phuong L, Picasso L, Cabeza R, Olson IR (2007) Parietal lobe and episodic memory: bilateral damage causes impaired free recall of autobiographical memory. *J Neurosci* 27:14415–14423.
- Binder JR, Desai RH (2011) The neurobiology of semantic memory. *Trends Cogn Sci* 15:527–536.
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796.
- Bonnici HM, Richter FR, Yazar Y, Simons JS (2016) Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. *J Neurosci* 36:5462–5471.
- Bracci S, Cavina-Pratesi C, Ietswaart M, Caramazza A, Peelen MV (2012) Closely overlapping responses to tools and hands in left lateral occipito-temporal cortex. *J Neurophysiol* 107:1443–1456.
- Bracci S, Ritchie JB, Kalfas I, Op de Beeck HP (2019) The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *J Neurosci* 39:6513–6525.
- Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2:913–919.
- Cléroux A, Friedman O (2020) Young children infer feelings of ownership from habitual use. *Emotion* 21:607–616.
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.
- Coutanche MN, Thompson-Schill SL (2015) Creating concepts from converging features in human cortex. *Cereb Cortex* 25:2584–2593.
- Diedrichsen J, Kriegeskorte N (2017) Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Comput Biol* 13:e1005508.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470–2473.
- Dunfield KA, Kuhlmeier VA (2010) Intention-mediated selective helping in infancy. *Psychol Sci* 21:523–527.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601.
- Epstein RA, Baker CI (2019) Scene perception in the human brain. *Annu Rev Vis Sci* 5:373–397.
- Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ (2019) fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature methods* 16:111–116.
- Fang M, Poskanzer C, Anzellotti S (2022) PyMVPD: a toolbox for multivariate pattern dependence. *Front Neuroinform* 16:835772.
- Forseth KJ, Kadipasaoglu CM, Conner CR, Hickok G, Knight RT, Tandon N (2018) A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain* 141:2112–2126.
- Friston KJ, Diedrichsen J, Holmes E, Zeidman P (2019) Variational representational similarity analysis. *Neuroimage* 201:115986.
- Fyfe C (1997) A neural network for PCA and beyond. *Neural Proc Lett* 6:33–41.
- Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform face area is part of a network that processes faces at the individual level. *J Cogn Neurosci* 12:495–504.
- Geschwind N (1972) Language and the brain. *Sci Am* 226:76–83.
- Hanke M, Adelhöfer N, Kottke D, Iacovella V, Sengupta A, Kaule FR, Nigbur R, Waite AQ, Baumgartner F, Stadler J (2016) A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Sci Data* 3:160092.
- Haxby JV, Gobbini MI, Furey ML, Isha A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Hecaen H, Angelergues R (1962) Agnosia for faces (prosopagnosia). *Arch Neurol* 7:92–100.
- Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int J Unc Fuzz Knowl Based Syst* 6:107–116.
- Hodges JR, Patterson K, Oxbury S, Funnell E (1992) Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. *Brain* 115:1783–1806.
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. Hawaii Convention Center, Honolulu, Hawaii.
- Humphreys GF, Ralph MA, Simons JS (2021) A unifying account of angular gyrus contributions to episodic and semantic cognition. *Trends Neurosci* 44:452–463.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Koster-Hale J, Saxe R, Dungan J, Young LL (2013) Decoding moral judgments from neural representations of intentions. *Proc Natl Acad Sci USA* 110:5648–5653.
- Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc Natl Acad Sci USA* 104:20600–20605.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008a) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141.
- Kriegeskorte N, Mur M, Bandettini PA (2008b) Representational similarity analysis: connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. *Nat Rev Neurosci* 18:42–55.
- Lanzoni L, Ravasio D, Thompson H, Vatansever D, Margulies D, Smallwood J, Jefferies E (2020) The role of default mode network in semantic cue integration. *Neuroimage* 219:117019.
- Mahon BZ, Milleville SC, Negri Gioia AL, Raffaella R, Caramazza I, Martin A (2007) Action-related properties shape object representations in the ventral stream. *Neuron* 55:507–520.
- Martin A, Chao LL (2001) Semantic memory and the brain: structure and processes. *Curr Opin Neurobiol* 11:194–201.
- Mhaskar H, Liao Q, Poggio T (2017) When and why are deep networks better than shallow ones? Thirty-First AAAI Conference on Artificial Intelligence. San Francisco, California.
- Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc Natl Acad Sci USA* 108:9998–10003.
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.
- Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8:976–987.
- Patterson K, Lambon Ralph MA (2016) The hub-and-spoke hypothesis of semantic memory. In: *Neurobiology of language*, pp 765–775. Amsterdam: Elsevier.
- Peelen MV, Caramazza A (2012) Conceptual object representations in human anterior temporal cortex. *J Neurosci* 32:15728–15736.
- Peelen MV, Wiggett AJ, Downing PE (2006) Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron* 49:815–822.

- Pobric G, Jefferies E, Lambon Ralph MA (2010) Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Curr Biol* 20:964–968.
- Poskanzer C, Fang M, Aglinskis A, Anzellotti S (2021) Controlling for spurious nonlinear dependence in connectivity analyses. *Neuroinformatics* 20:599–611.
- Price AR, Bonner MF, Peelle JE, Grossman M (2015) Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *J Neurosci* 35:3276–3284.
- Price AR, Peelle JE, Bonner MF, Grossman M, Hamilton RH (2016) Causal evidence for a mechanism of semantic integration in the angular gyrus as revealed by high-definition transcranial direct current stimulation. *J Neurosci* 36:3829–3838.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci* 16:5205–5215.
- Qin L, Lyu B, Shu S, Yin Y, Wang X, Ge J, Siok WT, Gao JH (2021) A heteromodal word-meaning binding site in the visual word form area under top-down frontoparietal control. *J Neurosci* 41:3854–3869.
- Rabini G, Ubaldi S, Fairhall SL (2021) Combining concepts across categorical domains: a linking role of the precuneus. *Neurobiol Lang* 2:354–371.
- Schwarzlose RF, Baker CI, Kanwisher N (2005) Separate face and body selectivity on the fusiform gyrus. *J Neurosci* 25:11055–11059.
- Sengupta A, Kaule FR, Guntupalli JS, Hoffmann MB, Häusler C, Stadler J, Hanke M (2016) A study for extension, retinotopic mapping and localization of higher visual areas. *Scientific data* 3:1–14.
- Sergent J, Ohta S, Brennan M (1992) Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain* 115:15–36.
- Soon C, Siong B, Marcel H, Hans J, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11:543–545.
- Wagner AD, Shannon BJ, Kahn I, Buckner RL (2005) Parietal lobe contributions to episodic memory retrieval. *Trends Cogn Sci* 9:445–453.
- Wagner IC, van Buuren M, Kroes MC, Gutteling TP, van der Linden M, Morris RG, Fernández G (2015) Schematic memory components converge within angular gyrus during retrieval. *Elife* 4:e09668.
- Woolrich MW, Ripley BD, Brady M, Smith SM (2001) Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage* 14:1370–1386.
- Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) NeuroSynth: a new platform for large-scale automated synthesis of human functional neuroimaging data. *Front Neuroinformatics Conf Abstr 4th INCF Congr Neuroinformatics*. Boston, Massachusetts.