ARTIFICIAL NEURAL NETWORKS REVEAL MULTIVARIATE INTEGRATION OF INFORMATION FROM MULTIPLE CATEGORY-SELECTIVE REGIONS

A PREPRINT

Mengting Fang
Department of Psychology
Boston College
Boston, MA 02467
fangmg@bc.edu

Aidas Aglinskas
Department of Psychology
Boston College
Boston, MA 02467
aglinska@bc.edu

Yichen Li
Courant Institute of Mathematical Sciences
Department of Computer Science
New York University
New York, NY 10012
yl3506@nyu.edu

Stefano Anzellotti
Department of Psychology
Boston College
Boston, MA 02467
stefano.anzellotti@bc.edu

March 20, 2020

ABSTRACT

Human visual cortex is organized into regions that respond preferentially to different categories of objects (i.e. faces, bodies, artifacts, scenes). However, often people need to integrate information about objects from different categories to make inferences about the world. How does the brain integrate information represented in different category-selective regions? In this work, we investigated this question taking advantage of a new analysis approach. We modeled the multivariate statistical dependence between fMRI responses in different brain regions using artificial neural networks. Regions whose responses were predicted significantly better by a combination of multiple category-selective regions than by the best-predicting category-selective region taken individually were identified as candidate hubs for multi-category integration. We used this approach to analyze fMRI responses to complex dynamic stimuli (the movie Forrest Gump), and identified five candidate hubs for multi-category integration: 1) angular gyrus, 2) posterior cingulate, 3) middle cingulate, 4) thalamus, and 5) cerebellum. Hubs were identified robustly across different artificial neural network architectures. Furthermore, representational similarity analysis revealed different representational geometry in integration hubs as compared to category-selective regions. These results indicate that a small set of localized regions integrates visual information about different object categories, and suggest that multi-category integration leads to a transformation of the similarity structure of neural representations.

1 Introduction

Visual information about different object categories is processed by distinct, specialized brain regions [1, 2, 3, 4, 5, 6]. Using functional magnetic resonance imaging (fMRI), researchers have observed preferential responses to human faces in the occipital face area (OFA), fusiform face area (FFA) and face-selective posterior superior temporal sulcus (face STS) [7, 3, 8]. The viewing of human bodies selectively activates the extrastriate body area (EBA), fusiform body area (FBA), and body-selective posterior superior temporal sulcus (body STS) [5, 9, 10], whereas artifacts lead to greater activity in the medial fusiform gyrus (mFus) and middle temporal gyrus (MTG) [11, 12]. Furthermore, the transverse

occipital sulcus (TOS), parahippocampal place area (PPA), and retrosplenial cortex (RSC) respond preferentially to scenes and buildings [4, 13].

In contrast with the category-selective organization of visual cortex, solving problems people encounter in everyday life often requires to integrate perceptual information about objects that belong to different categories. Imagine observing a person taking a box of cereal from a shelf in a store - you might infer that s/he is going to purchase it. However, if the person is taking the cereal from a cabinet in her/his own kitchen, you might conclude that s/he plans to eat it. While this type of inference is seemingly effortless, it requires recognizing objects (the cereal box), actions, and scenes, and integrating this information across multiple categories to infer the person's intention.

Two main hypotheses have been proposed to account for how the brain integrates information from regions selective for different categories. According to the 'Hub Hypothesis', information from different categories is integrated within dedicated brain regions ('hubs'). Damasio and colleagues [14] hypothesized the existence of 'convergence zones' that integrate perceptual information into entities and later into complex events. More recently, in light of evidence from patients with semantic dementia [15] and transcranial magnetic stimulation (TMS) [16], Patterson and colleagues [17, 18, 19] proposed that the anterior temporal lobes (ATL) serve as an integration hub for information from different modalities and categories.

Another variant of the Hub Hypothesis proposes that multiple 'semantic' brain regions operate as hubs [20]. A meta-analysis of fMRI studies on semantic knowledge [21] compiled and analyzed results from experiments in which participants accessed the meaning of words (as opposed to reading pseudowords or listening to distorted speech). Brain regions reliably engaged by such semantic contrasts include angular gyrus and posterior cingulate, as well as medial prefrontal cortex and inferior frontal gyrus [21]. A recent study investigated the contribution of angular gyrus to combinatorial semantics - the ability to integrate multiple concepts into a single meaning (i.e. representing 'plaid jacket' as a combination of 'plaid' and 'jacket') [22]. Meaningful word combinations were found to activate the angular gyrus more than non-meaningful combinations [22]. In addition, in patients with neurodegenerative disease, the degree of atrophy in the angular gyrus correlated with the amount of impairment at detecting meaningful word combinations [22]. TMS applied to the angular gyrus disrupts judgments about the thematic association between objects (i.e. trombone-orchestra) [23].

Like angular gyrus, posterior cingulate also responds more to meaningful word combinations (i.e. 'lake house') than to non-meaningful combinations (i.e. 'house lake') [24]. Furthermore, posterior cingulate encodes representations of people and places that can be accessed from multiple stimulus modalities (pictures and words) [25, 26]. Portions of posterior cingulate and angular gyrus respond to both semantic and episodic knowledge about people [27]. A recent study found that despite activity in posterior cingulate decreased for more difficult semantic tasks, correlation ('functional connectivity') between responses in posterior cingulate and dorsolateral prefrontal cortex increased as a function of task difficulty [28]. Evidence supporting the role of posterior cingulate and angular gyrus as 'convergence zones' has been leveraged to challenge the view that the anterior temporal lobe acts as the single highest-level integration hub, leading to the proposal that convergence across modalities and categories might occur in multiple regions [20].

An alternative to the Hub Hypothesis - the 'Synchrony Hypothesis' - states that the information encoded in different brain regions is integrated via synchrony in their responses over multiple frequency bands [29, 30]. Growing evidence suggests that neuronal synchronization between responses in distinct cortical areas plays a crucial role in effective inter-regional brain communication [31, 32]. In addition, recent studies using functional connectivity computed over sliding temporal windows found that the pattern of correlations between brain regions alternates between segregated states, in which responses in different regions are less correlated with each other, and integrated states, in which responses are more correlated [33, 34]. If integration of information across multiple categories relies exclusively on synchronization, spatially localized integration hubs might not be necessary.

However, the Hub Hypothesis and the Synchrony Hypothesis are not mutually exclusive. Research on the alternation between segregated and integrated states showed that the emergence of integrated states is most evident in frontoparietal cortex, in default mode regions (including angular gyrus and posterior cingulte), in the striatum and in the thalamus [33]. This suggests that large scale information integration might be both temporally and spatially localized.

Several studies on the integration of information across multiple brain regions relied on structural data or resting state data, without focusing on category-selective regions [35]. Other studies focused on integration across modalities (i.e. vision, audition, olfaction, language [25, 36, 37]). It remains unknown whether regions implicated in information integration in these studies also play a role for the integration of information encoded in regions selective for different categories.

In this study, we investigated the integration of information across multiple category-selective brain regions using Multivariate Pattern Dependence (MVPD, [38, 39]). We extended MVPD to model the multivariate statistical dependence between the response patterns in different brain regions using a variety of artificial neural network architectures.

Integration hubs were defined as regions whose responses are better predicted by the response patterns across regions selective for multiple categories, than by the response patterns in the regions selective for the best-predicting category in isolation. Applying this method to fMRI data collected during the viewing of complex dynamic stimuli [40], we identified five integration hubs robustly across different neural network architectures. Furthermore, representational similarity analysis (RSA, [41]) revealed that unlike face-, body-, and artifact-selective regions, the representational geometry in the hubs is not driven by animacy. Our findings are consistent with the Hub Hypothesis (and compatible with a hybrid Hub/Synchrony hypothesis), and indicate that other regions beyond ATL contribute to the integration of information across multiple category-selective regions.

2 Methods

2.1 Data

High-resolution (3×3×3 mm) BOLD fMRI responses 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, mean=29.4). The BOLD fMRI responses (acquired with a T2*-weighted echo-planar imaging sequence) were collected on a whole-body 3 Tesla Philips Achieva dStream MRI scanner equipped with a 32 channel head coil (see Hanke et al., 2016 for more details). 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). See [40] for more details.

During the category localizer session, participants were shown 24 unique gray-scale 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. During the movie localizer session, the movie stimulus 'Forrest Gump' was cut into eight segments, approximately 15 min long each. All eight movie segments were presented individually to participants in a chronological order in eight separate functional runs.

2.2 Preprocessing

Data were first preprocessed using fMRIPrep (`https://fmriprep.readthedocs.io/en/latest/index.html`), which is a robust and convenient pipline 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 anatomical scan with FSL FLIRT. The raw data of one participant could not pass the fMRIPrep processing pipeline. We used the data of the remaining 14 subjects for further analyses.

2.3 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 (FFA, OFA, and face STS), body-selective regions (EBA, FBA, and body STS), artifact-selective regions (mFus and MTG), and scene-selective regions (TOS, PPA, and RSC). Data were modeled with a standard GLM using FSL FEAT [42], and each seed ROI was defined as a 9mm radius sphere centered in the peak for its corresponding contrast (e,g. face-selective contrast: faces > bodys, artifacts, scenes, and scrambled images). We combined data from both left and right hemisphere for each ROI and then selected the 80 voxels which showed the highest t-values for the contrast between the preferred category and other categories.

Additionally, we created a group-average gray matter mask using the gray matter probability maps generated during preprocessing, with a total of 53539 voxels, that was used as the target of prediction.

2.4 MVPN: Multivariate pattern dependence network

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 (MVPA, [43, 44]) 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 [45, 46, 47, 48, 49]. Despite this, relatively few attempts have been made to leverage the potential of multivariate analyses to study brain connectivity (see [39] for a recent review). A recent study [38] has developed a technique that investigates the interactions between brain regions in terms of the multivariate relationship between their response patterns (multivariate pattern dependence - MVPD). MVPD has been shown to offer greater sensitivity than univariate connectivity methods [38], and uses independent training and testing data, thus offering improved robustness to noise.

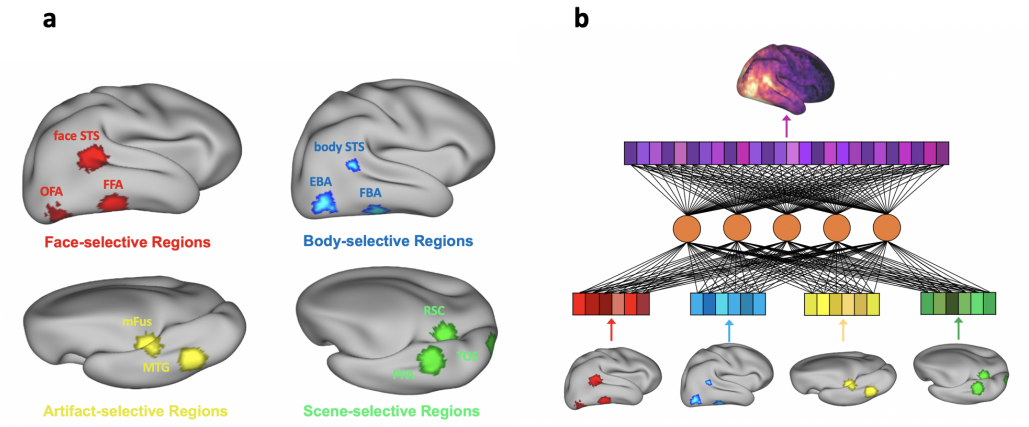

Figure 1: **a.** Category-selective ROIs of one example participant shown on an inflated cortical surface. Face-selective regions are shown in red (FFA, OFA, and face STS); Body-selective regions are shown in blue (EBA, FBA, and body STS); Artifact-selective regions are shown in yellow (mFus and MTG); Scene-selective regions are shown in green (TOS, PPA, and RSC). **b.** Illustration of an artificial neural network using the information from all category-selective regions combined as the 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.

The original MVPD formulation [38] used principal component analysis (PCA) to reduce the dimensionality of fMRI response patterns, and subsequently used linear regression as a model of the statistical dependence between brain regions. A more recent version of MVPD used simple artificial neural networks [50], but was limited to a small number of nodes in the hidden layer, and still relied on PCA for dimensionality reduction.

However, artificial neural networks can themselves perform dimensionality reduction [51]. In addition, using state-of-the-art software packages for the training and testing of artificial neural network paves the way for the training of more complex network architectures thanks to the use of general purpose graphic processing units (GPGPU).

To take advantage of these benefits, in this work we extended MVPD to larger artificial neural networks (Multivariate Pattern Dependence Network, or MVPN), and we implemented it in Pytorch to train the networks on four Tesla V100 GPUs. 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's consider an fMRI scan with m experimental runs. We can denote with $X_1, ..., X_m$ the multivariate timecourses in the predictor regions. Each matrix X_i is of size $n_X \times T_i$, where n_X is the number of voxels in the input regions, and T_i is the number of timepoints in the experimental run i. Analogously, we can denote with $Y_1, ..., Y_m$ the multivariate timecourses 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

$$Y_i(t) = f(X_i(t)) + \epsilon(t).$$

Specifically, for each choice of an experimental run i, data in the remaining runs were concatenated as the training dataset

$$D_{\setminus i} = \{(X_1, Y_1), ..., (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), ..., (X_m, Y_m)\},$$

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

As a measurement of the multivariate statistical dependence, we calculated the proportion of variance explained between the predictor regions 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, variance explained in run i was calculated as

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

where $f(X_i(j))$ is the voxel-wise prediction by MVPN. The values $\text{varExpl}_i(j)$ obtained for the different runs $i = 1, ..., m$ were averaged, yielding $\overline{\text{varExpl}}(j)$.

2.5 Exploring integration hubs

To identify hubs that integrate information from multiple category-selective regions, 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 the predictor from each of the four category-selective regions (face, body, artifact, and scene) individually. In a second analysis, we combined all the category-selective regions jointly as inputs of MVPN to predict the brain responses of each voxel in the gray matter mask.

If the neural responses in a 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 the region is receiving information from multiple category-selective regions. Otherwise, if a region only encodes information from one category of objects, using as predictors the responses from regions selective for multiple categories should not improve over using only the responses in the regions from the one category yielding the best predictions. To make things more precise, for each voxel j we can denote with $\mathrm{varExpl}_{all}(j)$ the variance explained by MVPN using as input the responses in all category-selective regions (Fig. 1b), and with $\mathrm{varExpl}_{max}(j)$ 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

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

and used it as a metric to evaluate whether a voxel integrates information across multiple regions selective for different categories (Fig. 2a). For each subject, we calculated the difference $\Delta\,\mathrm{varExpl}$ for each voxel in the gray matter mask, and we computed statistical significance across participants with statistical non-parametric mapping using the SnPM extension for SPM (`http://warwick.ac.uk/snpm`).

2.6 Neural network architectures

We trained MVPN using three different neural network architectures (Fig. 3a) to compare the predictive power of different models and to assess the robustness of our findings. All network architectures were fully connected and linear, and used 100 hidden nodes in each hidden layer. The first architecture was a one-layer feedforward network (Fig. 3a, left). Since previous studies have shown that deeper networks can approximate the same classes of functions as shallower networks using fewer parameters [52], we then tested a second, deeper architecture: a five-layer feedforward network (Fig. 3a, middle). Finally, a challenge encountered in training deep neural networks is the vanishing-gradient problem [53]: 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 five-layer DenseNet [54]. The DenseNet architecture includes connections that bypass multiple layers, (Fig. 3a, right), enabling more direct backpropagation of the loss function to the early layers.

All architectures were trained over 5000 epochs using stochastic gradient descent (SGD) on a mean squared error (MSE) 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 on `https://github.com/pandamt/MVPN`.

2.7 Representational similarity analysis

We used representational similarity analysis (RSA) [55, 41] to study the representational structure of integration hubs and to investigate how it differs from the representational structure in category-selective regions. RSA 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 [56].

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 each of the four sets of category-selective ROIs, response patterns in each of the regions in the set were concatenated, yielding four RDMs for each of the 14 participants. Additionally, RDMs were calculated for each of the integration hubs. RDMs were then averaged across participants, and the standard error of the mean (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. Note that 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.

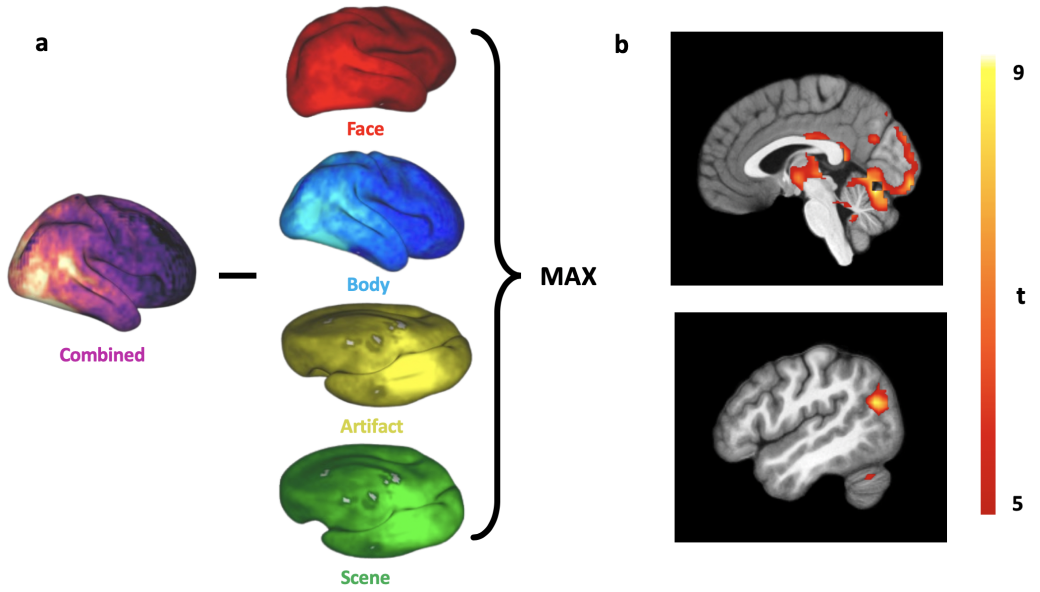

Figure 2: **a.** The MVPN analysis approach to detect hubs integrating information from multiple category-selective regions. **b.** Statistical t-maps computed from $\Delta\,\mathrm{varExpl}$ across subjects. $\Delta\,\mathrm{varExpl}$ is the voxelwise difference between the variance explained by the combined model and the variance explained by the single best predicting category. T-values were calculated across subjects using the $\Delta\,\mathrm{varExpl}$ results from the 5-Layer Dense MPVN model. Maps are thresholded at $t = 5$ for visualization purposes.

To further quantify the differentiation in the representational structure between integration hubs and category-selective regions, we proposed two hypotheses on the animacy of different object categories as the bases to evaluate the animate/inanimate distinction in each region. *Hypothesis 1:* Face and body stimuli share the same degree of animacy as animate objects, while artifact and scene stimuli share the same degree of inanimacy as inanimate objects. *Hypothesis 2:* The animacy of face, body, object, and scene stimuli decreases in sequence. We assigned a numerical rank value to each category of objects based on the degree of animacy in each hypothesis, i.e. *Hypothesis 1:* rank(face) = rank(body) = 1, rank(artifact) = rank(scene) = 2; *Hypothesis 2:* rank(face) = 1, rank(body) = 2, rank(artifact) = 3, rank(scene) = 4. RDMs derived from the two hypotheses were calculated by using the euclidean distance between the animacy order of each category of objects. The representational geometry in each hub and each set of category-selective ROIs were then compared with two hypothetical ones in terms of RDMs respectively (computed based on Spearman rank correlation).

3 Results

3.1 Identification of category integration hubs

To identify brain regions that integrate information from multiple categories of objects, we calculated for each voxel the difference between the proportion of variance explained by a model using all category-selective regions as predictors, and the proportion of variance explained by a model using as predictor regions selective for the single best predicting category. This subtraction revealed significant integration effects in five regions ($p < 0.05$ corrected with SnPM): (1) the posterior medial thalamus, (2) the middle cingulate gyrus, (3) the posterior cingulate gyrus, (4) the angular gyrus, and (5) the cerebellum (Figure 2, see also Supplementary Figure S1). All regions showed significant integration effects for all three network architectures tested, with the exception of the angular gyrus, which showed significant integration effects in two of the three architectures: the one layer network and the five layer dense network. The failure to identify angular gyrus in the five layer model without dense connections is likely due to the overall poor predictive accuracy of that model (see Figure 3, blue bars). MNI coordinates for the peaks of the SnPM t values for each of the hubs and each of the network architectures are reported in Table 1.

Table 1: **Integration hubs of multi-category object information in human brain.** For each MVPN model, every integration hub is defined as a 6mm radius sphere with 33 voxels inside. The locus of the hub is expressed in MNI coordinates thresholding at p < 0.05, FWE-corrected for multiple comparisons. # vox indicates the total number of voxels in the union hub of three MVPN models with duplicates removed.

Regions	MNI Coordinates												# vox
	1-Layer				5-Layer				5-Layer Dense				
	x	y	z	t	x	y	z	t	x	y	z	t	
Thalamus	-12	-27	14.4	9.05	-15	-24	14.4	7.74	-18	-21	21	7.38	77
Middle Cingulate	-6	-27	30.9	8.85	3	-24	30.9	6.39	-6	-30	27.6	6.97	81
Posterior Cingulate	3	-45	21	9.53	0	-45	17.7	6.23	3	-45	47.4	8.76	65
Angular Gyrus	-48	-66	24.3	8.54	-	-	-	-	-51	-66	24.3	9.61	46
Cerebellum	0	-69	-12	9.39	-6	-72	-25.2	5.94	0	-69	-12	10.14	66

3.2 Predictive performance of different network architectures

After having identified category integration hubs, we sought to test the predictive performance of different network architectures in each of the hubs. For each of the hubs and each network architecture, we selected the voxels within a 6mm radius sphere (33 voxels in total) surrounding the peak of the SnPM t contrast map in each MVPN model. Comparing the three network architectures within regions selected with one of the architectures might be circular, potentially leading to an advantage for the network that was used to define the region. To overcome this issue, we constructed regions of interests by computing the union of the spherical integration regions defined with the three networks. The total resulting number of voxels for each region is reported on Table 1, right column.

Architectures that yield more accurate models of the interaction between brain regions could lead to greater variance explained both when using as predictor with the combination of responses in all category-selective regions and when using as predictor with the responses in the regions selective for the single best predicting category. As a consequence, the difference in variance explained between the two is not a good indicator of the accuracy of a given architecture. Therefore, to evaluate the predictive power of different network architectures, we used the prediction performance of each architecture using the combination of all four category-selective regions. In each integration hub, we applied a one-sided sign test for each pair of MVPN models in terms of the combined variance explained to test for consistent differences of the predictive power between pairs of models. Specifically, we computed the non-parametric sign test statistic S as the number of subjects where one model's variance explained by the combined category prediction is greater than that of the other model.

Fig. 3b shows the predictive accuracy of each model in each integration hub by all category-selective regions combined and the corresponding p-values (Bonferroni-corrected for the 3 possible comparisons within each region). Overall, the 1-Layer MVPN model and the 5-Layer Dense MVPN model significantly outperformed the 5-Layer MVPN model without dense connections in most of the integration hubs 3. The 1-Layer MVPN model showed a significant higher combined variance explained than the 5-Layer MVPN model in four out of five integration hubs (n = 14 subjects, Thalamus: S = 12, p = 0.019; Posterior Cingulate: S = 13, p = 0.003; Angular Gyrus: S = 14, p = 0.000; Cerebellum: S = 13, p = 0.003). Similarly, the 5-Layer Dense MVPN model showed a significant higher combined variance explained than the 5-Layer MVPN model in all the integration hubs except for the Middle Cingulate (n = 14 subjects, Thalamus: S = 12, p = 0.019; Posterior Cingulate: S = 13, p = 0.003; Angular Gyrus: S = 14, p = 0.000; Cerebellum: S = 13, p = 0.003). In addition, the 1-Layer MVPN model significantly outperformed the 5-Layer Dense MVPN model in the Middle Cingulate (n = 14 subjects, S = 13, p = 0.003). Whereas in the Posterior Cingulate, the Angular Gyrus, and the Cerebellum, the 5-Layer Dense MVPN model significantly outperformed the 1-Layer MVPN model (n = 14 subjects,

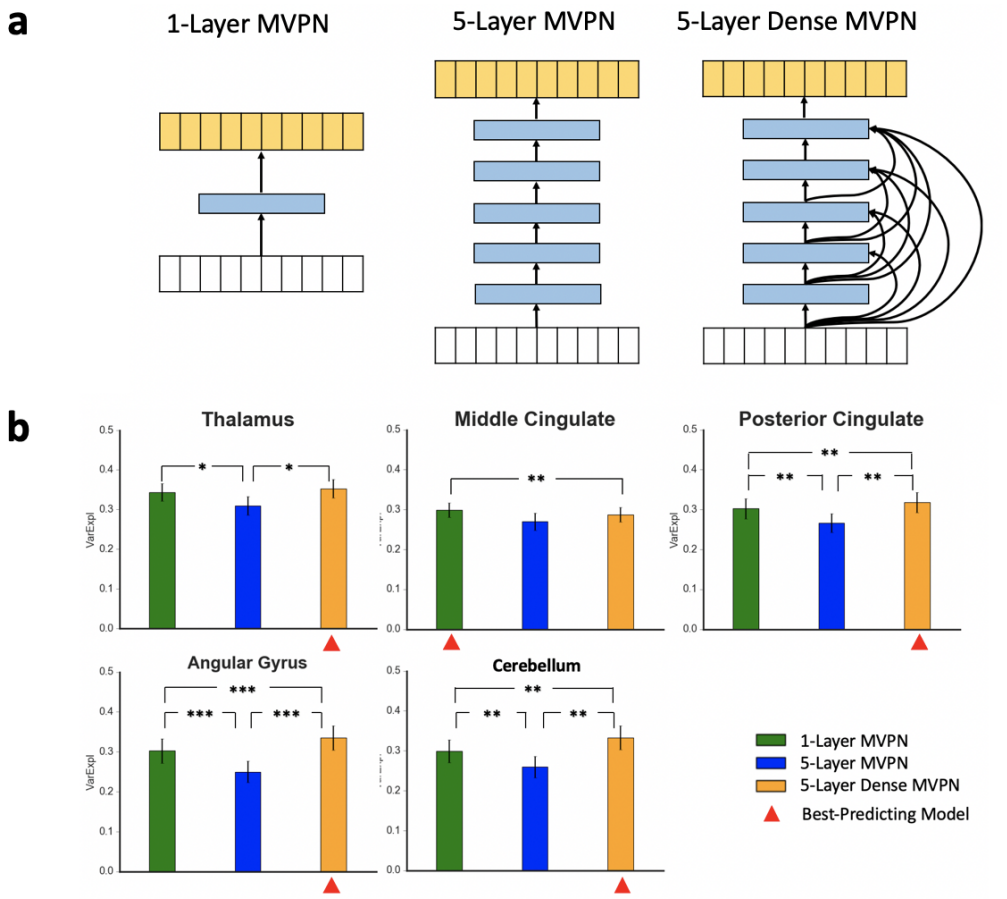

Figure 3: **a.** Different MVPN architectures: 1-Layer MVPN (left); 5-Layer MVPN (middle); 5-Layer Dense MVPN (right). **b.** Different predictive power of MVPN models across integration hubs and network architectures by the combination of response patterns from all category-selective regions. Error bars indicate the standard error of the mean (SEM). Stars above bars indicate significant differences between models (one-sided sign test, ***p < 0.001; **p < 0.01; *p < 0.05; Bonferroni-adjusted). The red triangle points to the model showing the highest variance explained across networks in each integration hub.

Posterior Cingulate: $S = 13$, p = 0.003; Angular Gyrus: $S = 14$, p = 0.000; Cerebellum: $S = 13$, p = 0.003). Among all the 15 comparisons, only three of them did not show significant differences across subjects (Thalamus: 9 subjects showed higher variance explained in the 5-Layer Dense MVPN model than in the 1-Layer MVPN model, p = 0.636; Middle Cingulate: 11 subjects showed higher variance explained in the 1-Layer MVPN model than in the 5-Layer MVPN model, p = 0.086; 9 subjects showed higher variance explained in the 5-Layer Dense MVPN model than in the 5-Layer MVPN model, p = 0.636).

3.3 Representational geometry of integration hubs

After identifying integration hubs, and finding that deep densely connected networks provide the best model (among those tested) of the statistical dependence between the hubs and category-selective regions, we aimed to characterize the representational similarity between objects in different categories in the hubs and compare it to the representational similarity in category-selective regions. For each hub and each set of category-selective regions, we calculated RDMs based on the correlation distance (Fig.4). Radar charts were used to visualize representational similarity together with its variability across participants (calculated as the standard error of the mean, Fig.4).

In face-, body- and artifact-selective ROIs, animate objects (faces and bodies) and inanimate objects (artifacts and scenes) elicited more similar responses than faces and scenes, faces and artifacts, or bodies and scenes. This similarity structure is reflected in the asymmetrical shape of the radar charts for the category-selective ROIs (Fig.4). In scene-

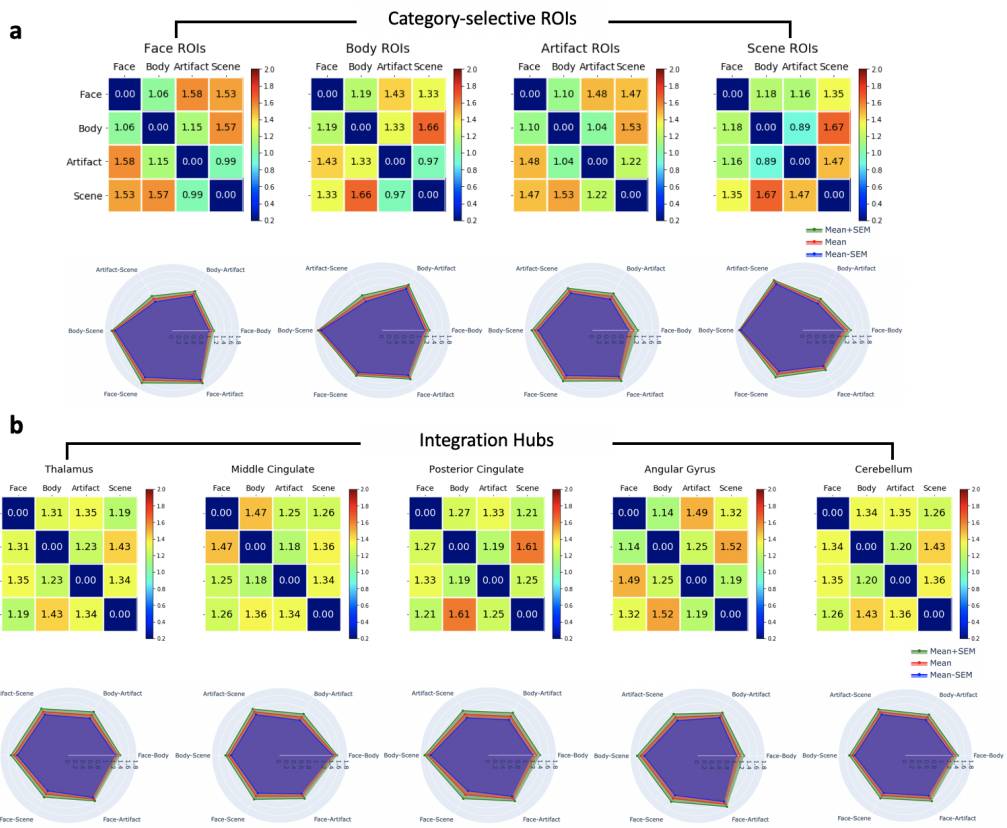

Figure 4: **Using representational dissimilarity matrices (RDMs) and radar charts to discriminate between category-selective ROIs and integration hubs. a.** RDMs (top) and radar charts (bottom) for average brain activities across subjects in four category-selective ROIs: face-selective ROIs, body-selective ROIs, artifact-selective ROIs, and scene-selective ROIs; **b.** RDMs (top) and radar charts (bottom) for average brain activities across subjects in five integrate hubs: 1) Thalamus, 2) Middle Cingulate, 3) Posterior Cingulate, 4) Angular Gyrus, and 5) Cerebellum.

selective ROIs, response patterns to scenes showed high dissimilarity from the responses to all other object categories. Whereas this animacy effect was not observed in the integration hubs (with the possible exception of the angular gyrus).

We then quantified the extent to which the animate/inanimate distinction drives the representational geometries of integration hubs and category-selective ROIs. For each region, we computed the Spearman rank correlation between its RDM with the RDM derived from each of the two animacy hypotheses (Fig.5a). The results (Fig.5b) clearly show that the similarity structure in all category-selective ROIs except the scene-selective ROIs has a generally high correlation with both animacy hypotheses, indicating an evident animacy effect in these regions. Whereas all the integration hubs do not correlate well with either animacy hypothesis except for the angular gyrus showing a slightly higher correlation with two animacy hypotheses.

We conducted a one-sample one-tailed t-test across all subjects to determine whether the Fisher-transformed Spearman rank correlation coefficient for each animacy hypothesis was significantly greater than zero in each region. To correct for multiple comparisons, we adjusted the p-values by multiplying the number of regions tested (four category-selective ROIs, five integration hubs). Results for each integration hub showed no significant correlations with either animacy hypothesis (Hypothesis 1, Thalamus: mean = 0.015, $t(13) = 0.102$; Middle Cingulate: mean = -0.148, $t(13) = -1.179$; Posterior Cingulate: mean = 0.148, $t(13) = 0.884$; Angular Gyrus: mean = 0.266, $t(13) = 1.906$; Cerebellum: mean = -0.118, $t(13) = -0.678$; all $p > 0.198$; Hypothesis 2, Thalamus: mean = 0.082, $t(13) = 0.689$; Middle Cingulate: mean = -0.040, $t(13) = -0.387$; Posterior Cingulate: mean = 0.143, $t(13) = 1.346$; Angular Gyrus: mean = 0.247, $t(13) = 2.041$; Cerebellum: mean = 0.004, $t(13) = 0.031$; all $p > 0.155$), demonstrating that the representational geometries in these regions were not driven by discriminating animacy.

In contrast, our results showed significant correlations between the RDMs for face- and body-selective ROIs with RDMs from two animacy hypotheses (Hypothesis 1, face-selective ROIs: mean = 0.562, $t(13) = 7.066$, $p = 1.694 \times 10^{-5}$;

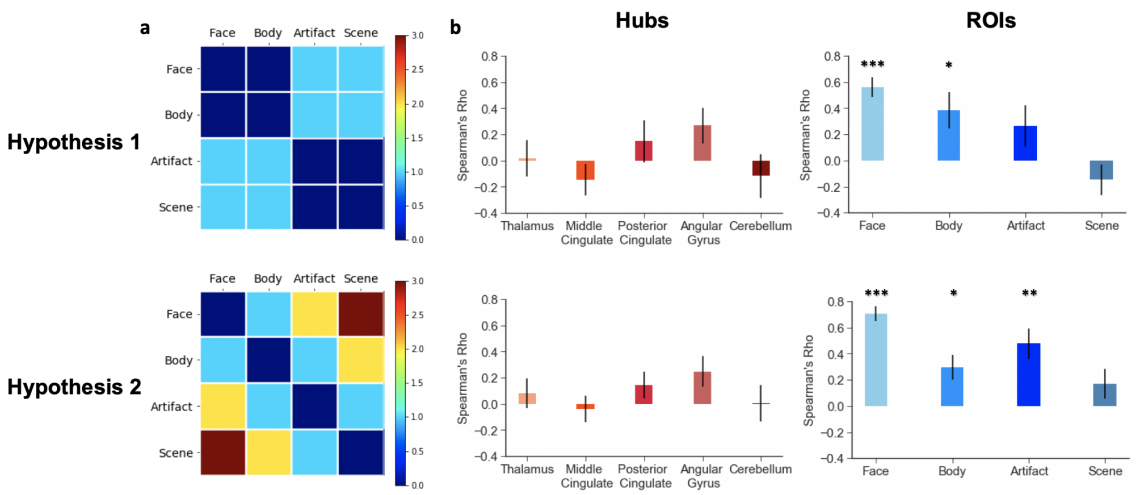

Figure 5: **a.** Representational dissimilarity matrices (RDMs) for two hypotheses on the animacy of objects in different categories. Hypothesis 1 (top): ranking order of animacy - face = body > artifact = scene. Hypothesis 2 (bottom): ranking order of animacy - face > body > artifact > scene; Each RDM is computed by using euclidean distance with regard to the numerical rank value of each object category. **b.** Spearman rank correlation of the RDMs for integration hubs and category-selective regions versus RDMs derived from two animacy hypotheses, respectively. Error bars indicate the standard error of the mean(SEM). Stars above bars indicate the statistical significance (one-sample one-tailed t-test on Fisher-transformed correlation coefficients, ***p < 0.001; **p < 0.01; *p < 0.05; Bonferroni-adjusted)

body-selective ROIs: mean = 0.384, t(13) = 2.645, p = 0.040; Hypothesis 2, face-selective ROIs: mean = 0.705, t(13) = 12.135, p = 3.655×10^{-8}; body-selective ROIs: mean = 0.298, t(13) = 3.050, p = 0.019). We additionally found that the RDM for artifact-selective ROIs significantly correlate with the RDM from the animcay Hypothesis 2 (mean = 0.476, t(13) = 4.006, p = 0.003). No other significant correlations can be observed (Hypothesis 1, artifact-selective ROIs: mean = 0.266, t(13) = 1.633, p = 0.253; scene-selective ROIs: mean = -0.148, t(13) = -1.215, p = 0.492; Hypothesis 2, scene-selective ROIs: mean = 0.168, t(13) = 1.415, p = 0.361).

4 Discussion

We have identified five candidate hubs for multi-category integration: 1) angular gyrus, 2) posterior cingulate, 3) middle cingulate, 4) thalamus, and 5) cerebellum. In each of these five regions, responses are better predicted by response patterns across all category-selective regions combined than by the best-predicting single category. Representational geometry in the candidate hubs does not appear to be driven by animacy, making a clear distinction from regions selective for faces, bodies and artifacts. Four of the five candidate hubs for multi-category integration we identified (angular gyrus, posterior cingulate, middle cingulate and thalamus) lie within areas previously reported to show high network centrality and vulnerability in analyses of structural and functional connectivity [57, 58, 59, 35].

Previous research has found that distinct subregions of posterior cingulate show differential responses to people and places across both words and pictures [60, 26]. At first glance, these findings might appear in contrast with the proposal that posterior cingulate integrates information across multiple object categories. However, knowledge about a specific category of objects can be inferred by integrating perceptual inputs from multiple category-selective regions. For instance, in the example used in the Introduction of this article, person-specific knowledge (someone's intention) was inferred by integrating information about the object that was taken (a cereal box) and the place (a store vs. one's own kitchen). Posterior cingulate might be organized by the category of the objects about which it represents knowledge, and at the same time it could infer that knowledge by integrating perceptual inputs across multiple category-selective regions.

Angular gyrus and posterior cingulate are implicated in both semantic knowledge and episodic memory retrieval [61, 62]. Semantic knowledge might play a key role to segment experience into meaningful units and their relations, and to support compressed and efficient episodic memory storage. A recent meta analysis found that angular gyrus and posterior cingulate are consistently activated in studies on emotion, Theory of Mind, and morality, and proposed that these regions might play a key role for the representation of both concrete and abstract concepts [63]. Recent work investigating the representation of events used Hidden Markov Models to reveal a hierarchy of event representations,

with angular gyrus and posterior cingulate at the top of the hierarchy [64]. More specifically, angular gyrus and posterior cingulate represented longer events, and showed high consistency in their responses across a movie and an audio narration of the same story [64]. In addition, event segmentation based on neural responses in these regions matched closely with behavioral judgments of event boundaries [64]. Interestingly, participants who had been already exposed to the movie version of the story showed an anticipation effect in the middle cingulate upon hearing the audio narrative [64], suggesting that this region might play a role in the anticipation of future events.

Taken together, the angular gyrus, posterior cingulate and middle cingulate might integrate object representations across multiple category-selective regions, and supplement them with semantic knowledge about those objects and their relations to organize experience into complex events with causal structure. In turn, this inferred causal structure could be used to generate expectations about the future.

Despite several studies have argued for a key role of the ATL as an multimodal and multi-category integration hub [17, 19], we did not find evidence for integration of information across multiple category-selective regions in this area. Thus, our results appear to be in conflict with reports of deficits affecting multiple object categorizing in semantic dementia [17], a condition affecting the ATLs [15]. One possibility is that the quality or resolution of our data was not sufficient to reveal multi-category integration effects in ATL. The proximity of the ATL to nasal sinuses and ear canals [65] might prevent us from detecting reliable results in this region. However, it is also possible that multi-category deficits in the ATL might stem from damage to multiple adjacent category-selective subregions. Patients with herpes simplex encephalitis, a condition that also causes ATL damage, can present with category-selective deficits for people [66]. Even in neurodegenerative diseases, cortical thinning in the right medial ATLs is associated with impairments for naming living things [67]. Moreover, evidence from fMRI studies has revealed the presence of regions within the ATL showing selective responses to faces [68] and animals [69]. Anterior temporal face-selective patches are also well documented in the macaque literature [70]. Additional evidence indicates that subregions of the ATL show differential responses to person knowledge [71, 72] and Theory of Mind [73]. While these findings do not rule out the possibility that some other subregions of the ATL might integrate information across multiple object categories, the evidence for the existence of such a subregion is not conclusive.

Previous work has demonstrated the thalamus in the integration of information from multiple sensory modalities [74, 75, 76]. More specifically, the medial and dorsal divisions of the medial geniculate body (MGB) have been implicated in the integration between visual and auditory information [77]. In addition to MGB, the pulvinar has been reported to respond to visual, auditory, and somatosensory stimuli [78, 79], and retrograde tracer studies identified it as the region with the most extensive overlap between auditory, somatosensory and premotor afferents [80]. Given the results in this article, thalamus might be involved in integrating information not only across sensory modalities, but also across regions selective for different object categories.

Interestingly, we also have identified the cerebellum as one of the candidate hubs, which is rarely mentioned in previous literature related to information integration. Although the cerebellum has traditionally been thought to coordinate motor actions [81, 82, 83], there is compelling evidence over the past three decades showing its involvement in a broad range of cognitive and affective functions, including language processing [84], visual attention [85], working memory [86], emotion perception [87], and social cognition [88]. Moreover, a recent study has identified the existence of distinct functional subregions (i.e. motor, cognitive, and social/affective) in the human cerebellum by employing a multi-domain task-based parcellation of the cerebellar cortex [89]. We suspect that the observed multi-category integrative role of the cerebellum in this study may provide a basis for multiple higher-level cognitive and social functions. Future studies are needed to more rigorously test this hypothesis and determine how the cerebellar circuits might perform these computations.

A possible concern is that using brain responses from multiple category-selective regions as predictors, we might obtain more accurate predictions because increasing the number of predictor variables could 'average out' the noise. However, MVPN trains a model of the interaction between brain regions using part of the data (the 'training set'), and tests how accurate the model is on independent data (the 'testing set'). In this type of generalization test, including additional variables that do not provide novel information can lead to overfitting [90], and to a decrease in the accuracy of predictions in the test set. In addition, if using responses from multiple category-selective regions led to improved predictions thanks to overall noise reduction, we would have expected to find that the combined model outperformed the best of the four single-category models across broad swaths of temporal and parietal cortex. In contrast with this expectation, our results identified five highly-localized candidate hubs.

Multi-category hubs were identified robustly across different neural network architectures. Among different models we tested, densely connected artificial neural networks [54] were the most accurate - we would recommend the use of this architecture for future studies. In the absence of dense connections, networks with multiple layers performed more poorly than networks with a single layer (Fig. 4).

The results of our representational similarity analysis revealed (not surprisingly) that category-selective regions demonstrate an effect of animacy, with faces showing greater similarity to bodies, and artifacts showing greater similarity to scenes (with the exception of scene-selective regions, where we found scenes to be dissimilar from every other category). By contrast, integration hubs do not show this animacy effect. Future studies might help identify key dimensions driving the representational geometry of integration hubs.

In conclusion, this study supports the view that a small set of candidate hubs - including angular gyrus, posterior cingulate and middle cingulate - contributes to the integration of information across multiple category-selective regions. Such integration might contribute to the encoding and retrieval of semantic knowledge and episodic memory, and to the anticipation of future events.

Acknowledgments

We would like to 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., 2018) for their assistance with the fmriprep preprocessing pipeline.

References

[1] Hz Hecaen and R Angelergues. Agnosia for faces (prosopagnosia). *Archives of neurology*, 7(2):92–100, 1962.

[2] Justine Sergent, Shinsuke Ohta, and BRENNAN MACDONALD. Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain*, 115(1):15–36, 1992.

[3] Nancy Kanwisher, Josh McDermott, and Marvin M Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11):4302–4311, 1997.

[4] Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598, 1998.

[5] Paul E Downing, Yuhong Jiang, Miles Shuman, and Nancy Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001.

[6] Alex Martin and Linda L Chao. Semantic memory and the brain: structure and processes. *Current opinion in neurobiology*, 11(2):194–201, 2001.

[7] Aina Puce, Truett Allison, Maryam Asgari, John C Gore, and Gregory McCarthy. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *Journal of neuroscience*, 16(16):5205–5215, 1996.

[8] Isabel Gauthier, Michael J Tarr, Jill Moylan, Pawel Skudlarski, John C Gore, and Adam W Anderson. The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of cognitive neuroscience*, 12(3):495–504, 2000.

[9] Michael S Beauchamp, Kathryn E Lee, James V Haxby, and Alex Martin. Fmri responses to video and point-light displays of moving humans and manipulable objects. *Journal of cognitive neuroscience*, 15(7):991–1001, 2003.

[10] Rebecca F Schwarzlose, Chris I Baker, and Nancy Kanwisher. Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience*, 25(47):11055–11059, 2005.

[11] Linda L Chao, James V Haxby, and Alex Martin. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience*, 2(10):913–919, 1999.

[12] Bradford Z Mahon, Shawn C Milleville, Gioia AL Negri, Raffaella I Rumiati, Alfonso Caramazza, and Alex Martin. Action-related properties shape object representations in the ventral stream. *Neuron*, 55(3):507–520, 2007.

[13] Russell A Epstein and Chris I Baker. Scene perception in the human brain. 2019.

[14] Antonio R Damasio. The brain binds entities and events by multiregional activation from convergence zones. *Neural computation*, 1(1):123–132, 1989.

[15] John R Hodges, Karalyn Patterson, Susan Oxbury, and Elaine Funnell. Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, 115(6):1783–1806, 1992.

[16] Gorana Pobric, Elizabeth Jefferies, and Matthew A Lambon Ralph. Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rtms in normal participants. *Proceedings of the National Academy of Sciences*, 104(50):20137–20141, 2007.

[17] Karalyn Patterson, Peter J Nestor, and Timothy T Rogers. Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12):976, 2007.

[18] Karalyn Patterson and Matthew A Lambon Ralph. The hub-and-spoke hypothesis of semantic memory. In *Neurobiology of language*, pages 765–775. Elsevier, 2016.

[19] Matthew A Lambon Ralph, Elizabeth Jefferies, Karalyn Patterson, and Timothy T Rogers. The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18(1):42, 2017.

[20] Jeffrey R Binder and Rutvik H Desai. The neurobiology of semantic memory. *Trends in cognitive sciences*, 15(11):527–536, 2011.

[21] Jeffrey R Binder, Rutvik H Desai, William W Graves, and Lisa L Conant. Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12):2767–2796, 2009.

[22] Amy R Price, Michael F Bonner, Jonathan E Peelle, and Murray Grossman. Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *Journal of Neuroscience*, 35(7):3276–3284, 2015.

[23] James Davey, Piers L Cornelissen, Hannah E Thompson, Saurabh Sonkusare, Glyn Hallam, Jonathan Smallwood, and Elizabeth Jefferies. Automatic and controlled semantic retrieval: Tms reveals distinct contributions of posterior middle temporal gyrus and angular gyrus. *Journal of Neuroscience*, 35(46):15230–15239, 2015.

[24] William W Graves, Jeffrey R Binder, Rutvik H Desai, Lisa L Conant, and Mark S Seidenberg. Neural correlates of implicit and explicit combinatorial semantic processing. *Neuroimage*, 53(2):638–646, 2010.

[25] Scott L Fairhall and Alfonso Caramazza. Brain regions that represent amodal conceptual knowledge. *Journal of Neuroscience*, 33(25):10552–10558, 2013.

[26] Scott L Fairhall, Stefano Anzellotti, Silvia Ubaldi, and Alfonso Caramazza. Person-and place-selective neural substrates for entity-specific semantic access. *Cerebral Cortex*, 24(7):1687–1696, 2014.

[27] Aidas Aglinskas and Scott L Fairhall. Regional specialization and coordination within the network for perceiving and knowing about others. *Cerebral Cortex*, 2019.

[28] Katya Krieger-Redwood, Elizabeth Jefferies, Theodoros Karapanagiotidis, Robert Seymour, Adonany Nunes, Jit Wei Aaron Ang, Vierra Majernikova, Giovanna Mollo, and Jonathan Smallwood. Down but not out in posterior cingulate cortex: Deactivation yet functional coupling with prefrontal cortex during demanding semantic cognition. *Neuroimage*, 141:366–377, 2016.

[29] Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, 2(4):229, 2001.

[30] Peter Lakatos, Joachim Gross, and Gregor Thut. A new unifying account of the roles of neuronal entrainment. *Current Biology*, 29(18):R890–R905, 2019.

[31] Andre M Bastos, Julien Vezoli, and Pascal Fries. Communication through coherence with inter-areal delays. *Current opinion in neurobiology*, 31:173–180, 2015.

[32] Agostina Palmigiano, Theo Geisel, Fred Wolf, and Demian Battaglia. Flexible information routing by transient synchrony. *Nature neuroscience*, 20(7):1014, 2017.

[33] James M Shine, Patrick G Bissett, Peter T Bell, Oluwasanmi Koyejo, Joshua H Balsters, Krzysztof J Gorgolewski, Craig A Moodie, and Russell A Poldrack. The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron*, 92(2):544–554, 2016.

[34] James M Shine, Matthew J Aburn, Michael Breakspear, and Russell A Poldrack. The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. *Elife*, 7:e31130, 2018.

[35] Martijn P van den Heuvel and Olaf Sporns. Network hubs in the human brain. *Trends in cognitive sciences*, 17(12):683–696, 2013.

[36] Stefano Anzellotti and Alfonso Caramazza. Multimodal representations of person identity individuated with fmri. *Cortex*, 89:85–97, 2017.

[37] Isabelle Ripp, Anna-Nora zur Nieden, Sonja Blankenagel, Nicolai Franzmeier, Johan N Lundström, and Jessica Freiherr. Multisensory integration processing during olfactory-visual stimulation—an fmri graph theoretical network analysis. *Human brain mapping*, 39(9):3713–3727, 2018.

[38] Stefano Anzellotti, Alfonso Caramazza, and Rebecca Saxe. Multivariate pattern dependence. *PLoS computational biology*, 13(11):e1005799, 2017.

[39] Stefano Anzellotti and Marc N Coutanche. Beyond functional connectivity: investigating networks of multivariate representations. *Trends in cognitive sciences*, 22(3):258–269, 2018.

[40] Michael Hanke, Nico Adelhöfer, Daniel Kottke, Vittorio Iacovella, Ayan Sengupta, Falko R Kaule, Roland Nigbur, Alexander Q Waite, Florian Baumgartner, and Jörg Stadler. A studyforrest extension, simultaneous fmri and eye gaze recordings during prolonged natural stimulation. *Scientific data*, 3:160092, 2016.

[41] Jörn Diedrichsen and Nikolaus Kriegeskorte. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS computational biology*, 13(4):e1005508, 2017.

[42] Mark W Woolrich, Brian D Ripley, Michael Brady, and Stephen M Smith. Temporal autocorrelation in univariate linear modeling of fmri data. *Neuroimage*, 14(6):1370–1386, 2001.

[43] James V. Haxby, Maria Gobbini, Maura Furey, Alumit Ishai, Jennifer L. Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, 293:2425–30, 10 2001.

[44] Kenneth A Norman, Sean Polyn, Greg Detre, and James V Haxby. Beyond mind-reading: Multi-voxel pattern analysis of fmri data. *Trends in cognitive sciences*, 10:424–30, 10 2006.

[45] Nikolaus Kriegeskorte, Elia Formisano, Bettina Sorger, and Rainer Goebel. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, 104(51):20600–20605, 2007.

[46] Adrian Nestor, David C Plaut, and Marlene Behrmann. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences*, 108(24):9998–10003, 2011.

[47] Stefano Anzellotti, Scott L Fairhall, and Alfonso Caramazza. Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex*, 24(8):1988–1995, 2013.

[48] Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5):543, 2008.

[49] Jorie Koster-Hale, Rebecca Saxe, James Dungan, and Liane L Young. Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences*, 110(14):5648–5653, 2013.

[50] Stefano Anzellotti, Evelina Fedorenko, Alexander JE Kell, Alfonso Caramazza, and Rebecca Saxe. Measuring and modeling nonlinear interactions between brain regions with fmri. *bioRxiv*, page 074856, 2017.

[51] Colin Fyfe. A neural network for pca and beyond. *Neural Processing Letters*, 6(1-2):33–41, 1997.

[52] Hrushikesh Mhaskar, Qianli Liao, and Tomaso Poggio. When and why are deep networks better than shallow ones? In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[53] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116, 1998.

[54] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[55] Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:4, 2008.

[56] Karl J. Friston, Jörn Diedrichsen, Emma Holmes, and Peter Zeidman. Variational representational similarity analysis. *NeuroImage*, 201:115986, 2019.

[57] Gaolang Gong, Yong He, Luis Concha, Catherine Lebel, Donald W Gross, Alan C Evans, and Christian Beaulieu. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral cortex*, 19(3):524–536, 2008.

[58] Dardo Tomasi and Nora D Volkow. Functional connectivity hubs in the human brain. *Neuroimage*, 57(3):908–917, 2011.

[59] Longchuan Li, Xiaoping Hu, Todd M Preuss, Matthew F Glasser, Frederick W Damen, Yuxuan Qiu, and James Rilling. Mapping putative hubs in human, chimpanzee and rhesus macaque connectomes via diffusion tractography. *Neuroimage*, 80:462–474, 2013.

[60] Scott L Fairhall, Stefano Anzellotti, Silvia Ubaldi, and Alfonso Caramazza. Person-and place-selective neural substrates for entity-specific semantic access. *Cerebral Cortex*, 24(7):1687–1696, 2013.

[61] Brian Levine, Gary R Turner, Danielle Tisserand, Stephanie J Hevenor, Simon J Graham, and Anthony R McIntosh. The functional neuroanatomy of episodic and semantic autobiographical remembering: a prospective functional mri study. *Journal of cognitive neuroscience*, 16(9):1633–1646, 2004.

[62] Hongkeun Kim. Default network activation during episodic and semantic memory retrieval: a selective meta-analytic comparison. *Neuropsychologia*, 80:35–46, 2016.

[63] Rutvik H Desai, Megan Reilly, and Wessel van Dam. The multifaceted abstract brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752):20170122, 2018.

[64] Christopher Baldassano, Janice Chen, Asieh Zadbood, Jonathan W Pillow, Uri Hasson, and Kenneth A Norman. Discovering event structure in continuous narrative perception and memory. *Neuron*, 95(3):709–721, 2017.

[65] Joseph T Devlin, Richard P Russell, Matt H Davis, Cathy J Price, James Wilson, Helen E Moss, Paul M Matthews, and Lorraine K Tyler. Susceptibility-induced loss of signal: comparing pet and fmri on a semantic task. *Neuroimage*, 11(6):589–600, 2000.

[66] Elizabeth K Warrington and Tim Shallice. Category specific semantic impairments. *Brain*, 107(3):829–853, 1984.

[67] Simona M Brambati, D Myers, A Wilson, Katherine P Rankin, Stephen C Allison, Howard J Rosen, Bruce L Miller, and Maria Luisa Gorno-Tempini. The anatomy of category-specific object naming in neurodegenerative diseases. *Journal of Cognitive Neuroscience*, 18(10):1644–1653, 2006.

[68] Reza Rajimehr, Jeremy C Young, and Roger BH Tootell. An anterior temporal face patch in human cortex, predicted by macaque maps. *Proceedings of the National Academy of Sciences*, 106(6):1995–2000, 2009.

[69] Stefano Anzellotti, Bradford Z Mahon, Jens Schwarzbach, and Alfonso Caramazza. Differential activity for animals and manipulable objects in the anterior temporal lobes. *Journal of Cognitive Neuroscience*, 23(8):2059–2067, 2011.

[70] Doris Y Tsao, Sebastian Moeller, and Winrich A Freiwald. Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences*, 105(49):19514–19519, 2008.

[71] Roland Zahn, Jorge Moll, Frank Krueger, Edward D Huey, Griselda Garrido, and Jordan Grafman. Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences*, 104(15):6430–6435, 2007.

[72] W Kyle Simmons, Mark Reddish, Patrick SF Bellgowan, and Alex Martin. The selectivity and functional connectivity of the anterior temporal lobes. *Cerebral Cortex*, 20(4):813–825, 2010.

[73] David Dodell-Feder, Jorie Koster-Hale, Marina Bedny, and Rebecca Saxe. fmri item analysis in a theory of mind task. *Neuroimage*, 55(2):705–712, 2011.

[74] Marcus J Naumer and Jasper JF van den Bosch. Touching sounds: thalamocortical plasticity and the neural basis of multisensory integration. *Journal of neurophysiology*, 102(1):7–8, 2009.

[75] Sascha Tyll, Eike Budinger, and Toemme Noesselt. Thalamic influences on multisensory integration. *Communicative & integrative biology*, 4(4):378–381, 2011.

[76] Yutaka Komura, Ryoi Tamura, Teruko Uwano, Hisao Nishijo, and Taketoshi Ono. Auditory thalamus integrates visual inputs into behavioral gains. *Nature neuroscience*, 8(9):1203, 2005.

[77] John F Smiley and Arnaud Falchier. Multisensory connections of monkey auditory cerebral cortex. *Hearing research*, 258(1-2):37–46, 2009.

[78] Giuliano Avanzini, Giovanni Broggi, Silvana Franceschetti, and Roberto Spreafico. Multisensory convergence and interaction in the pulvinar-lateralis posterior complex of the cat's thalamus. *Neuroscience letters*, 19(1):27–32, 1980.

[79] DB Bender. Retinotopic organization of macaque pulvinar. *Journal of Neurophysiology*, 46(3):672–693, 1981.

[80] Céline Cappe, Anne Morel, Pascal Barone, and Eric M Rouiller. The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cerebral cortex*, 19(9):2025–2037, 2009.

[81] Masao Ito. Mechanisms of motor learning in the cerebellum. *Brain research*, 886(1-2):237–245, 2000.

[82] Amy J Bastian. Learning to predict the future: the cerebellum adapts feedforward movement control. *Current opinion in neurobiology*, 16(6):645–649, 2006.

[83] Zhenyu Gao, Boeke J Van Beugen, and Chris I De Zeeuw. Distributed synergistic plasticity and cerebellar learning. *Nature Reviews Neuroscience*, 13(9):619–635, 2012.

[84] Steven E Petersen, Peter T Fox, Michael I Posner, Mark Mintun, and Marcus E Raichle. Positron emission tomographic studies of the processing of singe words. *Journal of cognitive neuroscience*, 1(2):153–170, 1989.

[85] Greg Allen, Richard B Buxton, Eric C Wong, and Eric Courchesne. Attentional activation of the cerebellum independent of motor involvement. *Science*, 275(5308):1940–1943, 1997.

[86] John E Desmond, John DE Gabrieli, Anthony D Wagner, Bruce L Ginier, and Gary H Glover. Lobular patterns of cerebellar activation in verbal working-memory and finger-tapping tasks as revealed by functional mri. *Journal of Neuroscience*, 17(24):9675–9685, 1997.

[87] Oliver Baumann and Jason B Mattingley. Functional topography of primary emotion processing in the human cerebellum. *NeuroImage*, 61(4):805–811, 2012.

[88] Frank Van Overwalle, Kris Baetens, Peter Mariën, and Marie Vandekerckhove. Social cognition and the cerebellum: a meta-analysis of over 350 fmri studies. *Neuroimage*, 86:554–572, 2014.

[89] Maedbh King, Carlos R Hernandez-Castillo, Russell A Poldrack, Richard B Ivry, and Jörn Diedrichsen. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nature neuroscience*, 22(8):1371–1378, 2019.

[90] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

Supplementary Materials

Supplementary Figures

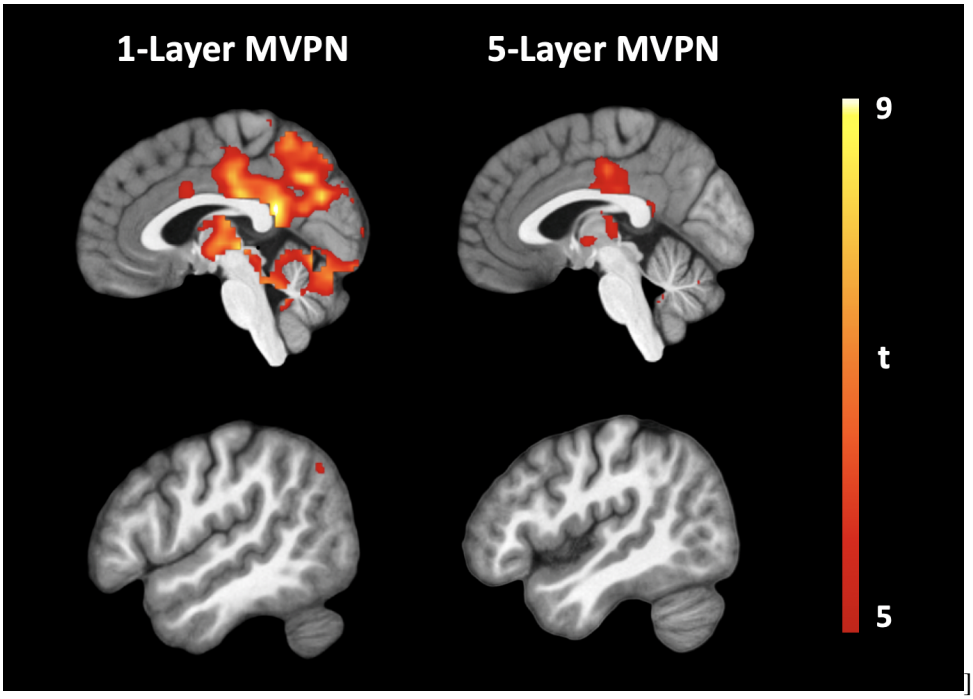

Figure S1: Statistical t-maps computed from Δ varExpl across subjects. Δ varExpl is the voxelwise difference between the variance explained by the combined model and the variance explained by the single best predicting category. Maps are thresholded at $t = 5$ for visualization purposes. Left: results from the 1-layer architecture; Right: results from the 5-layer architecture.