Enhanced generalization and specialization of brain representations of semantic knowledge in healthy aging

Pedro Margolles [1] and David Soto [1,2]

[1] Basque Center on Cognition, Brain and Language, San Sebastian, Spain
[2] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Author Note
Correspondence should be addressed to: pedromargolles@gmail.com, d.soto@bcbl.eu, Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd Floor 20009 San Sebastian.

Abstract

Aging is often associated with a decrease in cognitive capacities. However, semantic memory appears relatively well preserved in healthy aging. Both behavioral and neuroimaging studies support the view that changes in brain networks contribute to this preservation of semantic cognition. However, little is known about the role of healthy aging in the brain representation of semantic categories. Here we used pattern classification analyses and computational models to examine the neural representations of living and non-living word concepts. The results demonstrate that brain representations of animacy in healthy aging exhibit increased similarity across categories, even across different task contexts. This pattern of results aligns with the neural dedifferentiation hypothesis that proposes that aging is associated with decreased specificity in brain activity patterns and less efficient neural resource allocation. However, the loss in neural specificity for different categories was accompanied by increased dissimilarity of item-based conceptual representations within each category. Taken together, the age-related patterns of increased generalization and specialization in the brain representations of semantic knowledge may reflect a compensatory mechanism that enables a more efficient coding scheme characterized by both compression and sparsity, thereby helping to optimize the limited neural resources and maintain semantic processing in the healthy aging brain.

Keywords: semantic memory, aging, semantic representation, MVPA, decoding, RSA, neural dedifferentiation

Introduction

Aging is typically associated with a decline in performance across different cognitive domains including executive functions (Turner & Spreng, 2012), memory (Balota et al., 2000), attention (Madden, 2007) and general speed of processing (Salthouse, 2000), thereby impacting our quality of life. However, contrary to the decline observed in various cognitive domains, research has consistently shown that semantic memory tends to be relatively well preserved in healthy aging.

The preservation of semantic memory in aging can be attributed to the accumulation of knowledge and experiences acquired through lifetime. As individuals age, their conceptual richness and vocabulary size tend to remain stable or even improve (Rönnlund et al., 2005; Spaniol et al., 2006; Verhaeghen, 2003). Notably, behavioral studies observed that an expanded vocabulary is also associated with global modifications in the structure of semantic memory networks (Kenett & Thompson-Schill, 2020; Wulff et al., 2019; Yee & Thompson-Schill, 2016). These alterations may render the semantic networks less flexible and efficient, becoming more segregated and modular compared to younger individuals (Cosgrove et al., 2021, 2023; Dubossarsky et al., 2017; Wulff et al., 2019, 2022), thereby leading to changes in semantic knowledge use (e.g., Brosseau & Cohen, 1996; Kintz & Wright, 2017; Morrow & Duffy, 2005; Verheyen et al., 2019). Moreover, previous studies have also indicated that compensatory brain mechanisms contribute to maintaining semantic cognition in aging. Older adults undergo a shift from a left-lateralized semantic network to a more evenly distributed bilateral activation pattern, along with a transition from domain-specific to domain-general neural resources (see Hoffman & Morcom, 2018 for a meta-analysis).

Despite significant progress in understanding the relatively preserved nature of semantic memory in aging and the compensatory brain mechanisms in language and domain-general cognition, there remains a significant gap in our understanding of how healthy aging shapes the brain representation of semantic knowledge. The previous neuroimaging studies have primarily relied on univariate measures of brain activation. However, these measures, which average activity across individual voxels within a specific brain region, overlook the voxel covariance and the multivariate nature of brain activity patterns, potentially resulting in cases where distinct mental states generate indistinguishable univariate activation (Hebart & Baker, 2018; Ritchie et al., 2019). In contrast, multivariate pattern analysis (MVPA) methods (Haxby et al., 2001; Haynes & Rees, 2006; Norman et al., 2006; Tong & Pratte, 2012) offer higher specificity and sensitivity when exploring information encoded within distributed patterns of neural activity.

Here, we leverage the power of machine learning pattern classifiers and computational models to investigate how aging influences the brain representation of semantic categories related to animacy, based on words associated with living and non-living objects. Additionally, the study

addresses the role of the depth of processing in the brain representation of semantic knowledge. In this study, the depth of processing was operationalized using two conditions. The first condition involved silent reading, which required covertly reading the words and represented shallow processing. The second condition involved mental simulation of the concepts' properties, which required actively imagining sensory-specific representations and represented deep processing. This condition required a more profound comprehension of the words' meanings compared to the shallow processing that primarily focused on phonological aspects (see Methods and Figure 1 for an illustration of the experimental procedure). Previous neuroimaging studies in healthy young adults have primarily focused on examining how the brain represents semantic knowledge by using tasks in which participants had either to attend to the items (Buchweitz et al., 2012; Shinkareva et al., 2011; Simanova et al., 2014), think about the stimuli (Mitchell et al., 2008), or make semantic judgments (Fernandino et al., 2016). However, the influence of task demands on the neural representation of semantic knowledge remains to be fully elucidated (Soto et al., 2020). Understanding the impact of task demands on the brain's representation of semantic categories is crucial for a comprehensive understanding of the neurocognitive effects of aging. Older adults may employ compensatory mechanisms to sustain cognitive performance as a function of task demands (Hoffman & Morcom, 2018), which, in turn, may impact the brain representation of semantic knowledge relative to younger adults.

Our analysis approach entailed training a machine learning classifier on the activation patterns of a subset of experimental words, allowing it to predict the semantic categories of novel living/non-living words, not seen during the training phase. We conducted this analysis across 15 key regions of the left and right semantic network (see Methods and Figure 2), as identified in a previous meta-analysis conducted by Binder et al. (2009). According to the neural dedifferentiation hypothesis that proposes that aging is linked to a decline in the distinctiveness of brain activity patterns and the ability to efficiently allocate neural resources according to task goals (Koen & Rugg, 2019), here we predicted decoding accuracy would be lower in older adults compared to younger adults. Specifically, we hypothesized that this decline in semantic decoding with age would be particularly prominent during the mental simulation task because (i) it demands a greater involvement of cognitive resources (Haitas et al., 2021; Reuter-Lorenz & Cappell, 2008) and (ii) according to grounded cognition perspectives (Barsalou, 1999, 2008; Barsalou et al., 2003) age-related declines in physical and sensorimotor abilities could impact the integration of information necessary to support concept representation and mental simulation (Costello & Bloesch, 2017; Kuehn et al., 2018; Mille et al., 2021; Vallet, 2015). Moreover, based on the meta-analysis of mass-univariate fMRI studies in healthy young adults showing that the left hemisphere is more active during semantic tasks (Binder et al.,

2009), here we predicted a better decoding of semantic categories within the left hemisphere. However, we expected a decoding performance reduction in the left-hemisphere and change towards increased decoding in the right-hemisphere in older adults. This prediction follows from the meta-analysis by Hoffman & Morcom (2018) according to which older adults could display a compensatory mechanism characterized by a reduction in left hemisphere engagement during semantic tasks and an increased activity in the right hemisphere.

Second, we investigated how aging affected the generalization of brain representations of semantic categories across different levels of processing imposed by the task by employing cross-classification analysis (Kaplan et al., 2015). This approach aimed to assess the accuracy of a MVPA decoder trained on brain activation patterns of a subset of experimental words from one condition (e.g., deep processing during mental simulation) in predicting the semantic categories of words from the other condition (e.g., shallow processing during reading), and vice versa. We reasoned that if aging influences the generalization of brain representations across different tasks, then we would expect better classification performance (i.e., more generalizable representations) in the cross-condition classification analyses in the older compared to the younger group.

Third, we used model-based representational similarity analysis (RSA) to pinpoint the underlying structure of the brain representations associated with the experimental concepts using a language model based on word embeddings (Cardellino, 2016), and a computer vision model (Simonyan & Zisserman, 2014) based on images of living and non-living object referents associated with the words (e.g., Anderson et al., 2015, 2019; Deng et al., 2021; Devereux et al., 2018; Martin et al., 2018). The decision to use a computer vision model was motivated by the cognitive demands of the mental simulation condition, which required participants to simulate the sensory properties of each concept. By integrating both language and computer vision models into the RSA analysis, our goal was to offer a comprehensive understanding of the similarities and differences in the brain representation of semantic categories across different age groups.

Methods

Participants

Twenty-seven young adults (20-33 years old, mean±SD age: 24±3 years, 10 males) from a previous study (Soto et al., 2020), and 29 healthy older adults (see below for neuropsychological data; 66-80 years old, mean±SD age: 70.89±3.95 years, 12 males) participated in the research and received monetary compensation in return. All participants were right-handed, native Spanish speakers, with normal physical and neurological examinations. Moreover, participants reported to have normal or corrected-to-normal vision. Inclusion criteria for the older adult group required

participants to be at or above the Spanish retirement age (≥65 years old). Participants were recruited from the BCBL participants database. The experimental procedures were carried out following the principles stated in the Declaration of Helsinki and received approval from the BCBL Research Ethics Board. Written informed consent was obtained from all participants.

Neuropsychological Assessments

To ensure that the level of cognitive performance of older adults fell within the normal range for their age, a neuropsychological assessment was administered using standardized instruments adapted for the Spanish population. Cognitive impairment was evaluated through the administration of the following tests: the Mini-Mental State Examination (Mean = 29.21, SD = 1.01; Folstein et al., 1975; Lobo et al., 1980), the Montreal Cognitive Assessment (Mean = 26.66, SD = 1.91; Nasreddine et al., 2005), and the Repeatable Battery for the Assessment of Neuropsychological Status (Randolph et al., 1998). The Repeatable Battery evaluates five cognitive domains: immediate memory, visuospatial/constructional skills, language, attention, and delayed memory. For our participants, the Global Index Score had a mean of 112.31 with a standard deviation of 13.98. Moreover, the Percentile rank has a mean of 76 with a standard deviation of 24.29. Additionally, handedness was assessed using the Edinburgh Handedness Inventory, with all participants identified as right-handed (Oldfield, 1971). Furthermore, the older adults' cognitive reserve was evaluated utilizing the Cognitive Reserve Questionnaire (Rami et al., 2011) and the Cognitive Reserve Scale (León et al., 2011, 2014, 2016). The Cognitive Reserve Questionnaire is composed of 8 self-report test items measuring subject participation in cognitively stimulating activities throughout life (educational background, work occupation, musical training, language proficiency, and more) and provides a scoring of participants' cognitive reserves within four levels: lower-range (0 participants), low-medium range (0 participants), medium-high range (6 participants), and upper range (23 participants). The mean score was 16.9, with a standard deviation of 3.75. The Cognitive Reserve Scale is designed to assess individuals' engagement in cognitive stimulating activities across three different life stages: youth, adulthood, and maturity. It consists of 24 self-report items, categorized into four groups: activities of daily living, training-information, hobbies, and social life. To obtain the total score, the direct scores of each item from all life stages are added together (Mean = 173,66; SD = 31,70). Moreover, the scale also gathers data on participants' educational and occupational achievements, which are then classified into three distinct levels each [Educational attainment: Basic studies (6 participants), Mid studies (12 participants), High studies (11 participants); Occupational Attainment: Low (9 participants), Mid (12 participants), High (8 participants)].

MRI settings

Functional images were acquired from each participant using a SIEMENS Magnetom Prisma-fit whole-body MRI scanner with a 3T magnet and a 64-channel head coil. During each fMRI run, a multiband gradient-echo echo-planar imaging sequence was employed to capture 530 3D volumes encompassing the entire brain. The sequence incorporated a multi-band acceleration factor of 6, resulting in a voxel resolution of 2.4 * 2.4 * 2.4 mm^3. The acquisition parameters included a time-to-repetition of 850 ms, a time-to-echo of 35 ms, and a bandwidth of 2582 Hz/Px. The acquired volumes covered 66 slices with a field of view of 210. After the completion of the initial four functional runs, a high-resolution T1-weighted structural MPRAGE scan was obtained with a voxel size of 1 * 1 * 1 mm^3. This scan was conducted to aid in anatomical localization and to ensure accurate alignment across participants. The visual stimuli were presented on a screen located outside the scanner bore, and participants viewed them through a mirror attached to the head coil. To minimize the influence of head movements during imaging, participants were instructed to maintain their heads in a stable position. The head coil of the MRI scanner was padded with foam for added comfort. Additionally, participants were provided with cushioned headphones to attenuate background scanner noise and enable communication with experimenters via a compatible microphone. After each run, participants had the option to request a brief 1–2-minute break.

Experimental task

The experimental procedure consisted of 8 fMRI runs, each lasting approximately 8 minutes. At the beginning of each trial, a red fixation asterisk was displayed for 250 ms, followed by a 500 ms blank screen. Subsequently, a Spanish word was randomly chosen from a selection of 18 living words and 18 non-living words. The chosen word was then presented at the center of the screen for a duration of 1 second.

The living words category consisted of the following animal names: tigre (tiger), gallo (rooster), caballo (horse), oveja (sheep), cerdo (pig), gorila (gorilla), burro (donkey), yegua (mare), ardilla (squirrel), conejo (rabbit), gallina (hen), ballena (whale), tortuga (turtle), pantera (panther), camello (camel), elefante (elephant), canguro (kangaroo), perro (dog); while the non-living words category included the following tool names: llave (wrench), lápiz (pencil), tijera (scissors), aguja (needle), pinza (clamp), sierra (saw), clavo (nail), pincel (brush), alicate (pliers), tuerca (nut), navaja (knife), cepillo (brush), taladro (drill), soplete (blowtorch), tornillo (screw), cuchara (spoon), martillo (hammer), cuchillo (knife). The main properties of each word, such as word frequencies (per million of words), number of letters, number of phonemes, number of syllables, orthographic neighborhood size (per million of words), phonological neighborhood size (per million of words), familiarity, imageability, and concreteness were obtained from the EsPal database (bcbl.eu/databases/espal/;

Duchon et al., 2013). Independent samples t-tests, adjusted for multiple comparisons, were conducted to compare the properties of living and non-living words. No significant differences were observed between the word categories for any of the properties. Although 12 words lacked some of the familiarity, imageability, and concreteness ratings in the EsPal database, no significant differences were found between the experimental categories for the words that did have available ratings. A summary of the characteristics of living and non-living words can be found in Table S1, which is provided in the Supplementary materials.

During each scanning run, participants performed a total of 36 trials, with each trial corresponding to a randomly selected living or non-living word (Figure 1). The stimuli were presented in uppercase Arial font and displayed in white against a black background. Following the presentation of each word, participants were allotted a response window of 4 seconds. From the onset of the word, participants were instructed to either covertly read the word, pay attention to it, and mentally repeat it (shallow processing condition); or to engage in mental simulations related to the word's sensorimotor properties, such as its shape, color, sound, action, or context (deep processing condition). The two experimental conditions were alternated between runs. In order to improve the signal-to-noise ratio (SNR) in MVPA analyses, each word was shown a total of 8 times: four times during the shallow processing sessions (once per run), and likewise, four times during the deep processing condition runs (once per run). In the deep processing condition, participants were instructed to employ similar mental simulations when encountering the same word again. Finally, a red asterisk appeared in the center of the screen during the inter-trial period. To ensure that the BOLD response was separated between trials, the interstimulus interval period was jittered between 6 and 8 seconds. The duration of the ITI was determined using a pseudo-exponential distribution, resulting in different intervals for each trial. Specifically, 50% of the trials had a 6-second interval, 25% had a 6.5-second interval, 12.5% had a 7-second interval, and so on. Therefore, the time between word presentations ranged from 10.75 to 12.75 seconds across trials. Participants were instructed to relax and await the start of the next trial during this interval. To maintain participants' attention, a maximum of two catch trials were randomly included within each run. In these catch trials, number words (e.g., CERO, UNO, TRES) were presented instead of living and non-living words, and participants were instructed to press any of the four buttons on an fMRI compatible response pad. The number of catch trials was consistent across the different experimental tasks. We employed Psychopy (version 1.83.04, Peirce, 2007) for managing the display of stimuli during the experiment.

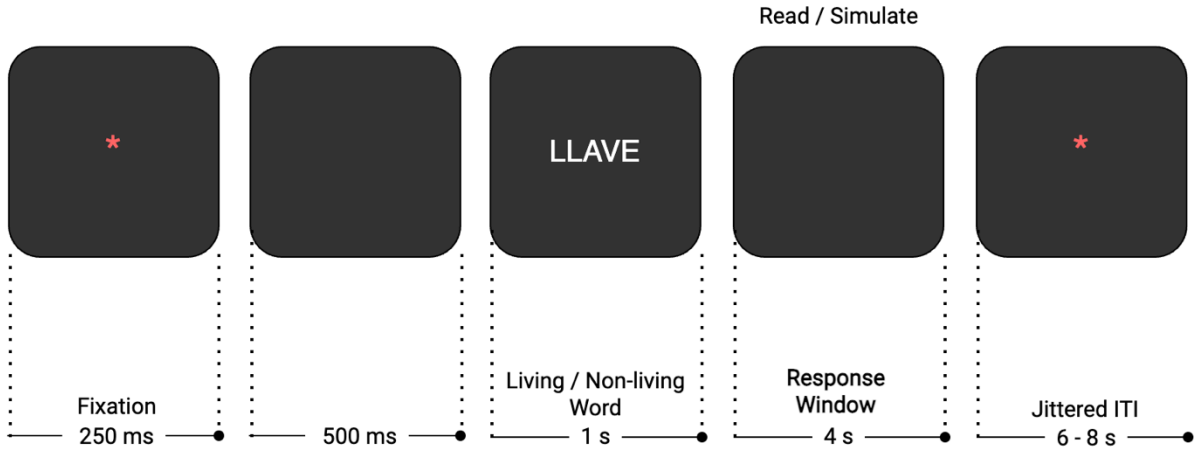

Figure 1. Illustration of an experimental trial in both depth of processing conditions (i.e., shallow, and deep).

Multivariate pattern analysis for decoding

MRI data preprocessing

The MRI data was first converted from DICOM to NIfTI format using MRIConvert (J. Smith & Theobald, 2013). Subsequently, a sequence of preprocessing steps was implemented on each fMRI run utilizing FSL (Jenkinson et al., 2012; S. M. Smith et al., 2004) and its *FEAT tool* (Woolrich et al., 2001, 2004). To ensure image stabilization, the first 9 volumes were discarded, and head-motion correction was carried out using the MCFLIRT algorithm (Jenkinson et al., 2002). Functional brain extraction was employed to eliminate non-brain voxels from data, followed by a minimal spatial smoothing with a Full-Width-Half Maximum (FWHM) Gaussian Kernel of 3 mm. The data was then linearly registered to the full brain structural image using a full search and the Boundary-Based Registration cost function. Furthermore, nonlinear registration was performed to align the data with the Montreal Neurological Institute (MNI) standard space employing a warp resolution of 10 mm and linear affine transformations comprising 12 parameters. The reference template used was MNI-152, with a voxel size of 2*2*2 mm^3. Co-alignment of runs 2-8 was accomplished by aligning each run to a reference volume derived from the preprocessed first run. Motion-induced signal variations were eliminated using the ICA-AROMA algorithm (Pruim et al., 2015), and a high-pass temporal filtering was applied with a cutoff of 60 seconds to further process the data.

Regions of interest (ROIs)

Based on a meta-analysis of the semantic system (Binder et al., 2009) a predetermined set of 15 ROIs was established. These ROIs included the inferior parietal lobe, inferior temporal lobe, middle temporal lobe, precuneus, fusiform gyrus, parahippocampal gyrus, superior frontal gyrus, posterior cingulate gyrus, pars opercularis, pars triangularis, pars orbitalis, frontal pole, medial

orbitofrontal cortex, lateral orbitofrontal cortex, and temporal pole (Figure 2). Both left and right hemisphere ROIs were extracted. The extraction process was initiated with an automatic segmentation of the T1-weighted structural image (excluding the neck) using FreeSurfer's v.6.0 recon-all brain reconstruction algorithm (Fischl, 2012). Subsequently, the mri_binarize command from FreeSurfer was utilized to independently generate anatomical masks for the 15 ROIs. This process involved extracting the cortical label indices from the resulting aparc + aseg file, which contained detailed information about the cortical regions, and carefully matching these indices with the corresponding entries in the FreeSurferColorLUT text file (which can be accessed at surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/AnatomicalROI/FreeSurferColorLUT). The resulting anatomical masks were examined in FSLView tool and then transformed into the functional space of each preprocessed run using FLIRT in FSL (Greve & Fischl, 2009; Jenkinson et al., 2002; Jenkinson & Smith, 2001), employing a 7 degrees-of-freedom (DOF) linear registration. Finally, the masks were binarized using the fslmaths command.

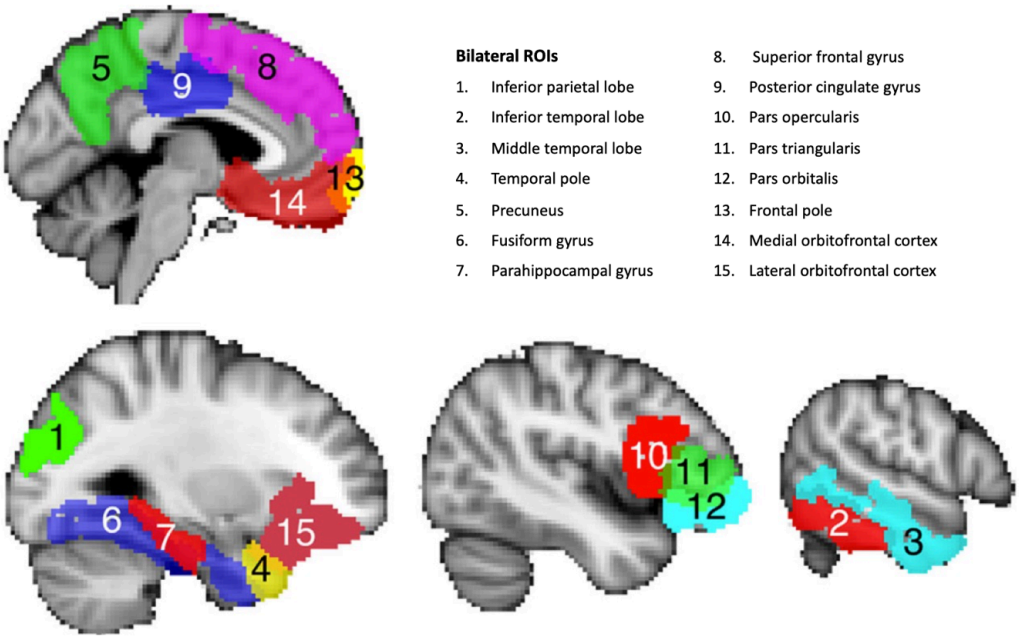

Figure 2. Illustration of the anatomical ROIs used in the decoding analyses based on Binder et al. (2009) meta-analysis.

Decoding analyses

MVPA decoding was performed using Python 2.7, utilizing Scikit-learn (Pedregosa et al., 2011) and PyMVPA (Hanke et al., 2009). The functional data within each brain region were extracted using the 15 binarized masks for each hemisphere, which were applied to all fMRI runs utilizing

PyMVPA. The generated behavioral data from Psychopy provided labels for time points within each preprocessed fMRI run mask, including information such as word, category, and condition. Data labeling was performed with consideration of the removal of the first 9 fMRI volumes of each run for image stabilization. The data from all runs were stacked separately for each ROI, while removing invariant features (i.e., voxels with zero variance across sessions) to avoid numerical issues during Z-scoring of dataset features. Subsequently, the data underwent run-wise linear detrending and Z-score standardization. To account for the hemodynamic lag and increase the SNR for decoding each trial, a single example was generated for every individual trial. This was achieved by averaging the mean voxel intensities from 3.4s to 8.6s following the onset of the word, effectively encompassing around six volumes. To assess whether multi-voxel patterns of brain activation contained relevant information regarding the semantic category of the word (living and non-living), a supervised machine learning classification approach was employed. Specifically, a linear support vector machine (SVM) classifier (Fan et al., 2008) with default parameters from the Scikit-learn library (l2 regularization, C = 1.0, tolerance = 0.0001) was employed for decoding in two contexts: within-condition and cross-condition. For each condition and participant, the following procedures were independently repeated for each ROI in both hemispheres:

Within-condition generalization. The objective was to obtain an unbiased estimate of generalization and assess the performance of our decoding model on unseen living and non-living words. Initially, the preprocessed data from each condition was divided into two balanced sets: the training set and the test set. The test set comprised all instances of one living concept (approximately 4 instances) and one non-living concept (approximately 4 instances). On the other hand, the training set comprised all instances of the remaining 34 concepts (approximately 94.44% of the data for a specific condition) to maximize the information available for a robust algorithm generalization. Each example was initially represented by a single feature vector with a length equal to the number of voxels in a ROI. To mitigate overfitting and reduce data dimensionality, a feature extraction process was implemented using Principal Component Analysis (PCA) on the training set. PCA transforms a correlated set of features into a smaller number of uncorrelated variables, known as principal components, which capture a significant portion of the original features' information. The PCA implementation utilized the default parameters of the PCA function from the Scikit-learn library. To maintain consistency across all ROIs, the number of components yielded by the PCA was adjusted to match the number of available instances in the training set. This approach ensured that each ROI had an equal number of components. It is noteworthy that the transformation to a subspace spanned by the instances was information-preserving, as no components were excluded (Mourão-Miranda et al., 2005). These new features obtained through PCA were used to train the decoding model.

Importantly, the PCA was applied solely to the training set, and the resulting dimensions were subsequently projected onto the test set. The decoding model's performance was evaluated by assessing its accuracy in predicting the semantic category of unseen instances for both one living and one non-living concept present in the test set. The leave-pair-words-out cross-validation method was applied for each condition, covering all possible combinations of training and test sets (18 living words * 18 non-living words = 324 potential test sets). The average and standard deviation of the generalization accuracies were then calculated over these 324 folds for each experimental condition, participant, and specific ROI (Figure 3).

Cross-condition generalization. To examine the generalization of brain representations across different levels of processing (i.e., shallow, and deep), a cross-condition decoding procedure was implemented (Kaplan et al., 2015). In this approach, the process was similar to the previous one, with the distinction that the classifier was trained using all instances of 34 concepts within one condition and tested on all instances of the two remaining living and non-living concepts from the other experimental condition. As a result, two potential cross-condition directions emerged: from shallow-to-deep and from deep-to-shallow processing for every pair of test concepts. The performance of the cross-condition model was assessed by measuring its accuracy in predicting the semantic category of previously unseen instances from each of the two concepts within the test condition (one living and one non-living concepts). The leave-pair-words-out cross-validation procedure was iterated 324 times, covering all possible combinations of training and test sets. The average and standard deviation of cross-condition generalization accuracies were then computed over these 324 folds for each cross-condition direction, participant, and specific ROI (Figure 4).

We conducted two mixed effects ANOVAs to compare the decoding accuracy across various factors in both within and cross-condition generalization analyses. For the within-condition generalization, the ANOVA involved 15 regions of interest (ROIs), 2 conditions (shallow vs. deep), 2 hemispheres (left vs. right), and 2 age groups (older vs. young). Similarly, in the cross-condition generalization analysis, the ANOVA incorporated 15 ROIs, 2 cross-condition directions (shallow-to-deep vs. deep-to-shallow), 2 hemispheres (left vs. right), and 2 age groups (older vs. young). The ANOVAs were conducted using JASP 0.12.2 (JASP Team, 2023), with Huynh-Feldt corrected DOF. Post-hoc pairwise comparisons were performed using the Python Scipy and Statsmodels modules (Seabold & Perktold, 2010; Virtanen et al., 2020) to examine main effects and interaction effects identified in the ANOVAs. Statistical p values in the pairwise comparisons were corrected for multiple comparisons using the False Discovery Rate (FDR) method.

Representational Similarity Analysis (RSA)

Word embedding model

We utilized the Word2Vec word embedding model in our study, specifically selecting it for its established performance (Fernandino et al., 2022; Martin et al., 2018; Wang et al., 2018), as well as the availability of a pre-trained model trained on the Spanish Billion Word Corpus (Cardellino, 2016). This model was obtained from the GitHub repository located at github.com/dccuchile/spanish-word-embeddings. Word embedding models capture the semantic and syntactic representation of words within a given corpus and represent them as dense vectors of numerical values (embeddings). Similar words tend to have similar embedding features. Following the standard practice in the natural language processing community, we obtained a vector representation of 300 dimensions (300-D) for each of the 36 words by extracting the embeddings from the final layer of the pre-trained Word2Vec model. The feature representations of words were standardized by subtracting their mean and dividing them by their standard deviation. Then we computed the representational dissimilarity between pairs of words using the Pearson correlation distance as the dissimilarity metric. This calculation was performed using the Scipy library (Virtanen et al., 2020) with the pdist function and specifying 'correlation' as the distance metric. To ensure robustness in subsequent analyses, correlations of words with themselves (e.g., dog - dog) and symmetrical pairs (e.g., dog - hammer | hammer - dog) were excluded. Figure 7A showcases a visualization of the Representational Dissimilarity Matrix (RDM) resulting from the pairwise correlation of the embeddings.

Computer vision model

We chose the VGG-19 model (Simonyan & Zisserman, 2014), a pre-trained computer vision model with a shallower architecture of 26 layers, as a well-established baseline in previous research (Deng et al., 2021; He et al., 2016; Naspi et al., 2022; Ren et al., 2015; Russakovsky et al., 2015). This convolutional neural network is specifically designed to map images to dense feature vectors and generate category probability estimates, making it a suitable choice for our study. The model was pre-trained on the ImageNet dataset (imagenet.org) and obtained from the Keras Python Library (Chollet, 2015). It is important to note that the dimensionality of the computer vision models differs from that of the word embedding models. Therefore, while the VGG-19 represents images with 512-D features, we adjusted the dimensionality of the computer vision model to match the 300-D dimensionality of the word embedding model to ensure comparability. This was achieved by adding a new layer on top of the pre-trained model structure before fine-tuning (Mesnil et al., 2012; Yosinski et al., 2014).

During the pre-training phase, this model was trained to predict the category of an input image from the ImageNet dataset, which consists of one thousand different categories. By leveraging this pre-training on ImageNet, the model acquired valuable information that could be further refined and applied to our specific image dataset and classification task. Consequently, to enhance the performance of the VGG-19 with our specific word-image dataset, we conducted a fine-tuning process. The fine-tuning process utilized the extensive dataset from the Caltech 101 database (Fei-Fei et al., 2007), which consists of 101 different object categories. This database provided a diverse collection of images covering a wide range of visual concepts. The images can be accessed directly through the following URL: data.caltech.edu/records/mzrjq-6wc02. The image categories in the Caltech 101 database encompass varying numbers of images, ranging from 45 to 400 images. Before fine-tuning, we excluded six categories that lacked semantic relevance to the living and non-living categories investigated in our experiment. These excluded categories were “BACKGROUND google”, “Faces”, “Faces easy”, “Brain”, “Stop sign”, and “Trilobite”. To ensure an equal distribution of samples across categories, we implemented a stratified approach by randomly selecting 30 images from each image category. Out of these 30 images, 24 were allocated to the training set, while the remaining 6 were assigned to the test set. Prior to fine-tuning, all images underwent individual normalization following the specific preprocessing steps outlined in the Keras documentation. Additionally, they were resized to dimensions of 128 * 128 px^2. Lastly, we augmented the existing model architecture by adding a classification layer with 96 categories on top of the existing 300-D layer. This modification ensured that the model was well-equipped to handle our specific classification task.

The fine-tuning process involved several steps to adapt the computer vision model to our experimental settings. Firstly, we froze the convolution pooling blocks of the pre-trained VGG-19 during training and validation, allowing only the weights of the newly added layer to be updated. To enhance the model's performance, augmentation techniques such as rotation, shifting, zooming, and flipping were applied to the image sets. For the 300-D layer, we utilized the Self-Normalizing activation function (Klambauer et al., 2017), and the weights were initialized using the LeCun normal initialization procedure (LeCun et al., 2012) as recommended in the TensorFlow documentation (Abadi et al., 2016). The Adam optimizer with a learning rate of 0.0001 was employed, and the loss function used was categorical cross-entropy (Kingma & Ba, 2014). No additional regularization techniques were applied. The model was trained for a maximum of 3000 epochs, with a batch size of 16 images per back-propagation step. If the model's performance on the validation set did not improve for 5 consecutive batches, the training process would halt.

After the fine-tuning of the VGG-19 model, it was employed to extract abstract representations of image referents linked to the experimental words. For each of the 18 living and 18

non-living words, a set of 10 images depicting the objects or entities that these words represent were collected from the Internet. These images were cropped with the object centered on a white background and their values were normalized based on the chosen model. Subsequently, these images were fed as input into the fine-tuned computer vision model. The penultimate layer, which was the newly added 300-D layer, provided the output vector that represented the features associated with the image referent of each word. To obtain a shared representation, the vector representations of all 10 images associated to each word were averaged. Like the word embedding models, the 300-D feature representations of the images corresponding to each concept were standardized and their pairwise correlations were computed to construct the RDM (Figure 7B).

RSA searchlight

The whole-brain pre-processed MRI data, which was previously used for ROI-based decoding, was employed for conducting RSA and searchlight analyses. Using this data, we labeled the time points within each preprocessed fMRI run with relevant properties, including word, category, and condition. The labeling process was performed based on the behavioral data generated by Psychopy, and it accounted for the removal of 9 fMRI heat-up volumes. The fMRI data from all runs were combined by stacking them together. Subsequently, run-wise linear detrending and Z-score standardization were applied to the entire brain data at the voxel level. To account for the hemodynamic lag inherent in the fMRI signal and increase the SNR, a single example per trial was created by averaging the mean voxel intensities between 3.4s and 8.6s after the onset of each word, which encompassed approximately 6 volumes.

For each experimental condition and computer model, searchlight RSA analyses were conducted using the Brain Imaging Analysis Kit (BrainIAK v. 0.9.1; http://brainiak.org; Kumar et al., 2020, 2021) and Python 3.6 in the subject's native and functional space. The searchlight analysis involved the generation of a binarized mask that identified the voxels within the brain where the searchlight analysis would be performed, excluding non-brain voxels where the fMRI activity was consistently zero across all pre-processed runs. A searchlight was defined as a spherical cluster with a radius of 2 voxels (approximately 4.6 mm) around each brain voxel, which was chosen as the target. This searchlight consisted of 32 voxels in total. Then, whole-brain voxel activations of all instances for each word, were averaged together into a single vector obtaining in total 36 features vectors, representing 18 living and 18 non-living words. The dissimilarity between the average brain activations of different word pairs was computed within each searchlight. The Pearson correlation distance was used as the dissimilarity metric, employing Scipy library (Virtanen et al., 2020) and the pdist function with the 'correlation' distance metric specified. As mentioned earlier for the computational models' dissimilarity matrices, we excluded correlations between words with

themselves (e.g., dog - dog) and symmetrical pairs (e.g., dog - hammer | hammer - dog). The RDM based on the neural data within each searchlight was then correlated with the RDM based on the computational representations of the stimuli using non-parametric Spearman rank correlation. This approach was recommended by Carlin et al. (2011) and Kriegeskorte et al. (2008) as rank-ordering is suitable for comparing RDMs from different data sources without assuming a linear relationship. The resulting Spearman ρ correlation coefficients resulting from the comparison of the RDMs were assigned to the center voxel of each searchlight, generating a whole-brain similarity map that revealed the correlation between brain activity patterns and the computational model. To ensure a normal distribution of the correlation coefficients, the Spearman correlation values for each subject were transformed using Fisher-Z transformation. This transformation stabilizes the variance of the correlation coefficients and facilitates subsequent statistical analyses that assume normality. The Fisher-transformed correlation coefficient maps were aligned from the subject's native space to the MNI-152 2*2*2 mm^3 standard space using FSL.

To account for the unknown null distribution of imaging data, a non-parametric approach was employed for the statistical contrasts. Within each age group, the whole-brain MNI correlation maps from all participants were merged into a single 4D image file using the fslmerge tool. A non-parametric one-sample t-test was conducted on the generated 4D image files to identify brain regions where the Z-transformed correlation coefficients significantly differed from zero. The t-test contrasts were computed using FSL RANDOMISE with Threshold-Free Cluster Enhancement (TFCE cluster correction) and 10,000 random permutations across subjects (Nichols & Holmes, 2002; S. M. Smith & Nichols, 2009; Winkler et al., 2014). To enhance the statistical power of the analyses, variance smoothing was applied using a spatial FWHM Gaussian Kernel of 5 mm. Multiple comparisons were corrected using Family-wise Error (FEW) correction. FSL RANDOMISE generated a 1 - p corrected statistical image, which was subsequently thresholded to display only statistically significant clusters (p-values ≤ 0.05).

A similar non-parametric procedure was followed for the comparisons between age groups. Participants' Spearman correlation maps between the brain and the model were merged into a 4D NifTi file, including data from both groups. The General Linear Modeling visual interface for FSL was used to set up the design matrix for the statistical analyses, maintaining the same order of participants as in the 4D file. Two two-sample unpaired T contrasts were defined: young group > older group and older group > young group. Using the generated 4D image, the design matrix, and the defined contrast files as inputs, a non-parametric permutation test was performed across subjects using FSL RANDOMISE with the same configuration as for the within-group analyses. This

analysis allowed us to identify clusters showing statistically significant correlation differences between the age groups.

To improve visualization, we independently created a global average mask for each age group. This was achieved by averaging the whole-brain correlation maps of participants within each group, resulting in a separate global average mask for the young and older age groups. In addition, we generated two additional masks by comparing the global average masks between the age groups (young group > older group; older group > young group). The average mask values were converted back from Fisher-transformed correlation coefficients to Spearman correlation coefficients. These converted values were then overlaid onto the significant cluster maps obtained previously. The resulting correlation maps, illustrating both within-group and between-group significant correlations, are presented below.

Results

Within-condition decoding results

A leave-pair-words-out cross-validation approach was used within each condition (i.e., shallow, and deep processing) to assess decoding accuracy for unseen word pairs' instances (corresponding to 1 living and 1 non-living concepts). The model was trained on a subset of 34 semantic concepts and tested on all possible pairs. This approach facilitated the assessment of decoding accuracy and the examination of how age, task condition, anatomical ROIs, and hemispheres influence the results using a mixed ANOVA analysis.

Decoding analyses within each level of processing (shallow, and deep processing) did not reveal any significant effect of aging or interactions with other factors. Hence, both age groups showed similar decoding patterns of semantic categories across the tested ROIs in both shallow and deep conditions, and across both hemispheres. However, we observed that decoding performance was better in the left hemisphere compared to the right hemisphere. Furthermore, the overall decoding accuracies were significantly greater in the deep processing condition than in the shallow processing condition across both hemispheres in line with the observation reported in a recent study involving young adults (Soto et al., 2020; see also the Supplementary results for a detailed account of the statistics unrelated to the effects of age). Figure 3 provides an overview of the decoding results across ROIs, experimental conditions, hemispheres, and age groups.

Cross-condition decoding results

We then examined the impact of aging on the specificity and generalization of brain representations of semantic categories across the different levels of processing. We utilized a

leave-pair-words-out cross-validation method, where a model was trained using data from 34 out of 36 semantic concepts in one task condition (e.g., deep), and then assessed on all potential pairs of concepts that were excluded (1 living concept, 1 non-living concept) in the other task condition (e.g., shallow). This resulted in two cross-condition directions: shallow-to-deep and deep-to-shallow. Figure 4 illustrates the cross-condition decoding results across various ROIs, cross-condition direction, hemispheres, and age groups. Decoding accuracies were compared utilizing a mixed ANOVA across ROIs, cross-condition directions, hemispheres, and age groups.

Similar to the within-condition decoding results, the cross-condition decoding analysis revealed significant main effects and also interaction effects that were unrelated to aging. These results are presented in the Supplementary results. Here, we report the most critical results regarding the effects of aging on cross-condition decoding.

The ANOVA revealed a significant main effect of age [F(1, 54) = 4.315, p = .043, η_p^2 = .074], indicating that cross-condition decoding was higher in the older group (Mean = .524, SE = .004) compared to the younger group (Mean = .513, SE = .004). We also observed an interaction effect of age and ROIs [F(12.650, 683.082) = 2.269, p = .007, η_p^2 = .04]. To unravel the source of this interaction, post-hoc pairwise comparisons were conducted, averaging over hemisphere and cross-condition direction levels. Within each ROI, unpaired two-tailed t-tests, with p-values corrected for multiple tests using the FDR method, were performed to compare the two age groups. Notably, a significant effect was found in the middle temporal lobe, with cross-condition decoding being significantly higher in the older group (Mean = .542, SE = 6.4e-3), compared to the young group (Mean = .514, SE = 5.3e-3) [t(54) = 3.300, p = .026].

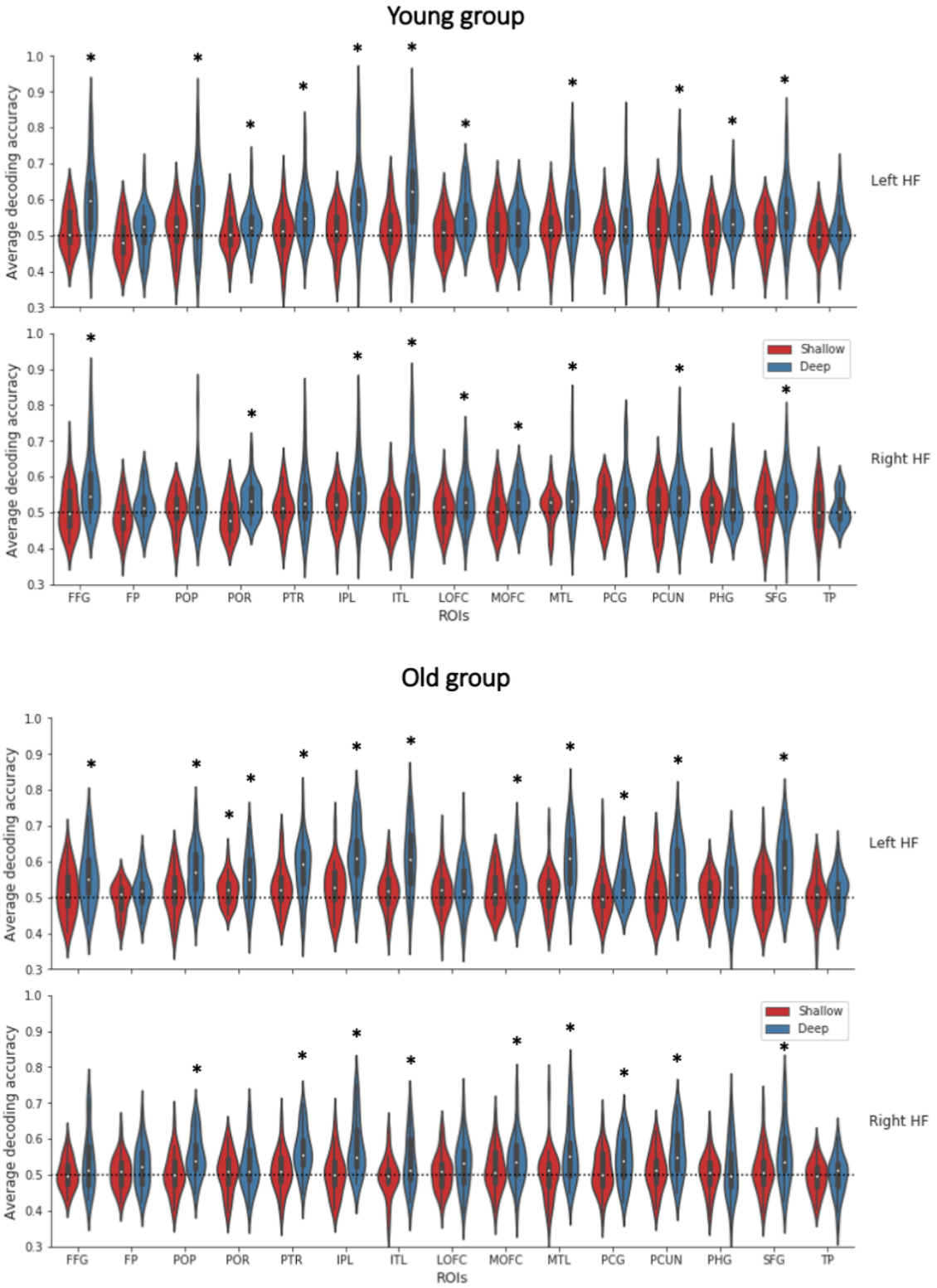

Figure 3. Within-condition decoding accuracies by ROI, condition, hemisphere, and age-group represented with violin plots. The black dotted line represents chance level (0.5). ROIs in which decoding performance was significantly above chance within each group in a one-sample t-test (FDR corrected p-values) are marked with an asterisk.

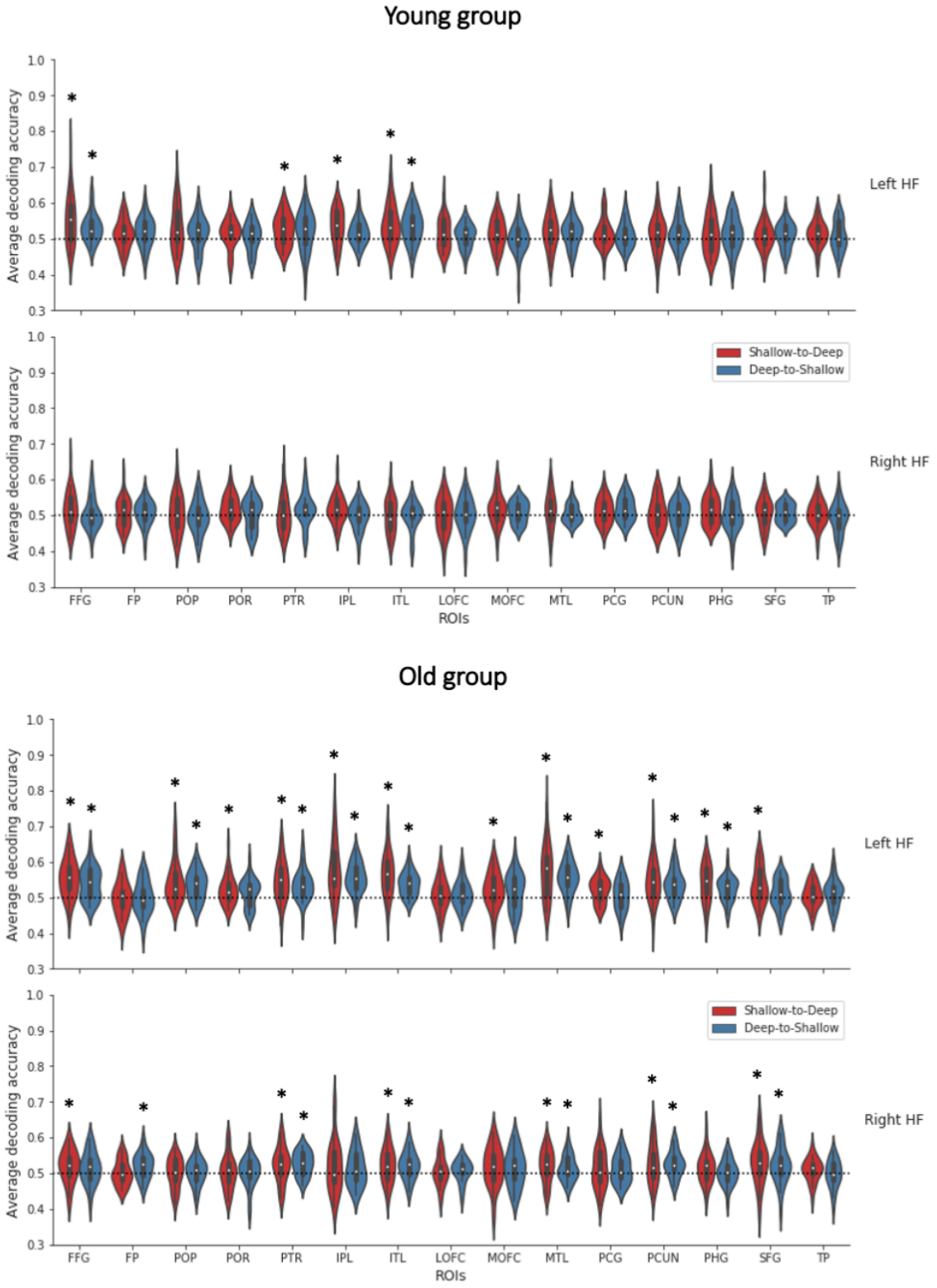

Figure 4. Cross-condition decoding accuracies by ROI, cross-condition direction, hemisphere, and age-group represented with violin plots. The black dotted line represents chance level (0.5). ROIs in which decoding performance was significantly above chance within each group in a one-sample t-test (FDR corrected p-values) are marked with an asterisk.

We further observed a significant interaction effect among age, ROIs, and hemisphere [F(11.562, 624.340) = 1.862, p = .038, η_p^2 = .033] indicating that the main effect of aging on cross-condition decoding depended on the hemisphere and ROI. To further investigate this interaction, two supplementary mixed-effects ANOVA analyses were conducted independently in the left and right hemispheres, utilizing a 15 (ROIs), 2 (cross-condition direction: shallow-to-deep vs. deep-to-shallow), 2 (age: older group vs. young group) design with Huynh-Feldt corrected DOF.

The supplementary ANOVA of the right hemisphere showed no significant main effect of age or interactions with other factors. However, in the left hemisphere supplementary ANOVA, a main effect of age was observed (F(1, 54) = 5.273, p = .026, η_p^2 = .089). Specifically, the older group (Mean = .533, SE = .004) exhibited larger cross-condition decoding compared to the younger group (Mean = .519, SE = .004). Additionally, an age by ROIs interaction was found (F(12.056, 651.035) = 2.962, p < .001, η_p^2 = .052). To explore this interaction effect, post-hoc pairwise comparisons were performed, collapsing across cross-condition decoding direction levels. Within each ROI in the left hemisphere, unpaired two-tailed t-tests with p-values corrected using the FDR method were conducted to compare the two age groups. Notably, significant differences were observed in the left inferior parietal lobe [Older: Mean = .566, SE = .010; Young: Mean = .528, SE = 6.4e-3; t(54) = 3.092, p = .024] and the left middle temporal lobe [Older: Mean = .566, SE = 9.9e-3; Young: Mean = .518, SE = 7.5e-3; t(54) = 3.780, p = .006] (Figure 5).

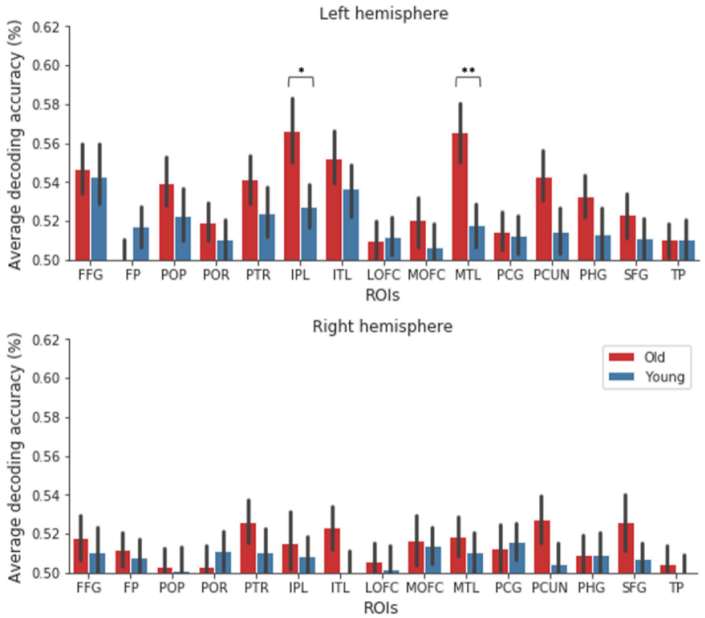

*Figure 5. Cross-condition decoding accuracy averaging over cross-condition direction levels (i.e., deep-to-shallow, and shallow-to-deep) by ROI, hemisphere, and age-group. The error bars represent a 95% confidence interval. * p < .05, ** p < .01.*

Model-based RSA

We utilized searchlight RSA to model the brain representation of animacy relative to the representational spaces given by the language model and the computer vision model (namely, Word2Vec and VGG-19 models), and examine differences across the age groups.

Firstly, to assess the significance of the correlations between each model and brain representational dissimilarity matrices, we performed one-sample permutation t-tests on the searchlight RSA maps for each age group separately, aiming to identify brain clusters where the correlations were significantly above zero (whole-brain corrected, $p < 0.05$ against 10000 permutations). In the shallow processing condition, significant correlation clusters were observed only in the older group. These clusters were located primarily in the left middle temporal gyrus for both the computer vision and word embedding models. In contrast, in the deep processing condition, significant correlation clusters were found for both the young and older groups. In keeping with the decoding analyses conducted within each level of processing task, the strongest correlations between neural data and latent features derived from VGG-19 and Word2Vec were observed in the left hemisphere. The correlations observed in both models were particularly robust during the deep condition, encompassing regions across the ventral visual pathway (including the occipitotemporal cortex, middle temporal gyrus, and inferior temporal gyrus), the dorsal visual pathway (such as the angular gyrus, supramarginal gyrus, superior parietal lobule, postcentral gyrus, parietal operculum cortex, and precentral gyrus), as well as frontal regions like the middle frontal gyrus, inferior frontal gyrus, and frontal pole.

We then compared the RSA maps across the age groups. Significant differences were observed in the deep processing condition. Specifically, the younger group exhibited higher correlations with the models compared to the older group in the bilateral postcentral gyrus and surrounding regions. These differences were most pronounced in the left hemisphere and were consistent for both the VGG-19 and Word2Vec models (Figure 6; the model RDMs are presented in Figure 7A and 7B). No significant differences were found between age groups in the shallow processing condition.

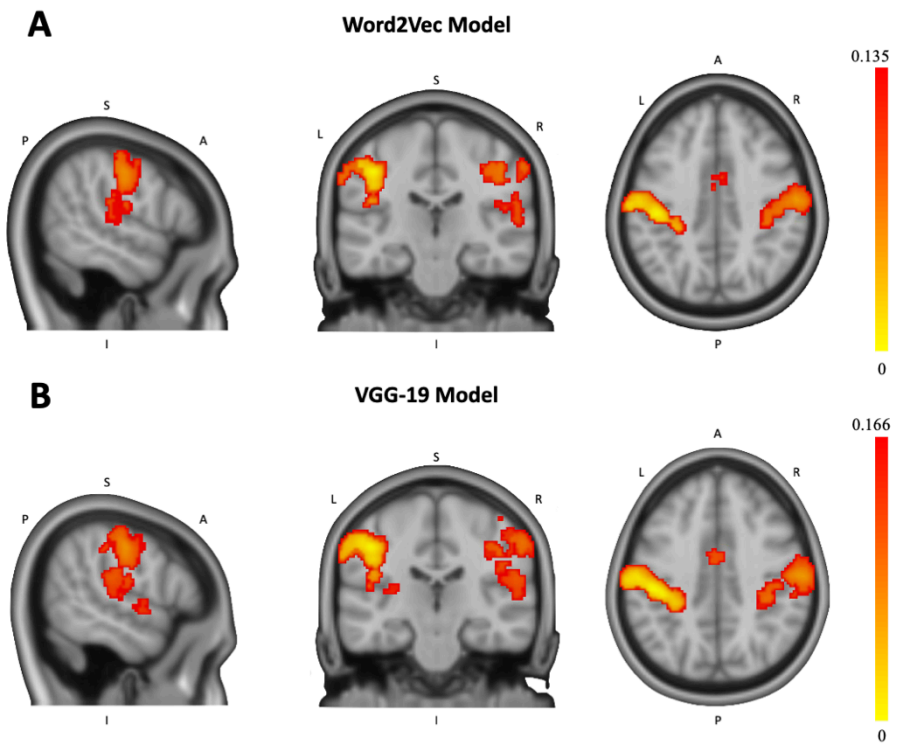

Figure 6. Deep processing condition. Searchlight RSA correlation differences between groups in the young > older direction using the word embedding model (A) Word2Vec; and the computer vision model (B) VGG-19.

However, we note that addressing the precise nature of the age differences at the representational level required further examination. It remains unclear whether the neural representation of the different animacy categories in the postcentral gyrus cluster become more similar with aging, or whether the similarity between the neural representations of different items within each category is also impacted by aging. The data from all participants were transformed into MNI space to extract information from the significant postcentral gyrus cluster. For each participant, we calculated the average Pearson dissimilarity scores (1 - ρ) for all pairs of concepts within each living and non-living category, as well as between categories. The correlation distance values for each subject were subjected to the Fisher-Z transformation to achieve normality. A mixed ANOVA with category as a within-subjects factor (living, non-living, average within-categories, and between categories concepts dissimilarity) and aging group as between-subjects factor (young vs. older adults) was conducted. There were significant main effects of category and group factors, and a significant interaction effect between both factors [$F(3, 162) = 12.051$, $p < .001$, $\eta_p^2 = .182$]. Post-hoc t-tests, with Holm-Bonferroni correction for multiple comparisons, revealed that older adults exhibited greater dissimilarity among concepts within the living [Mean Difference = .041, SE = .011, $t(55) = 3.736$, $p = .004$], within the non-living categories [Mean Difference = .035, SE = .011, $t(55) = 3.250$, $p = .016$], and also for the average dissimilarity of both categories (i.e., average within-category dissimilarity)

[Mean Difference = 0.038, SE = .011, $t(55) = 3.493$, $p = .009$]. However, the older group showed higher similarity between the living and non-living concepts relative to the young group [Mean Difference = -.036, SE = .011, $t(55) = -3.356$, $p = .012$]. These results are depicted in Figure 7C and 7D, and Figure 8A.

Lastly, we compared the neural pattern of dissimilarity between categories across groups by subtracting the within-categories concept dissimilarity from the between-categories dissimilarity for each group (Carp, Park, Hebrank, et al., 2011; Deng et al., 2021; Haxby et al., 2001). An independent sample t-test demonstrated that the older group [Mean = .026, SE = .006] exhibited enhanced category similarity in neural patterns compared to the young group [Mean = .100, SE = .021] [$t(54) = -3.492$, $p < .001$] (Figure 8B).

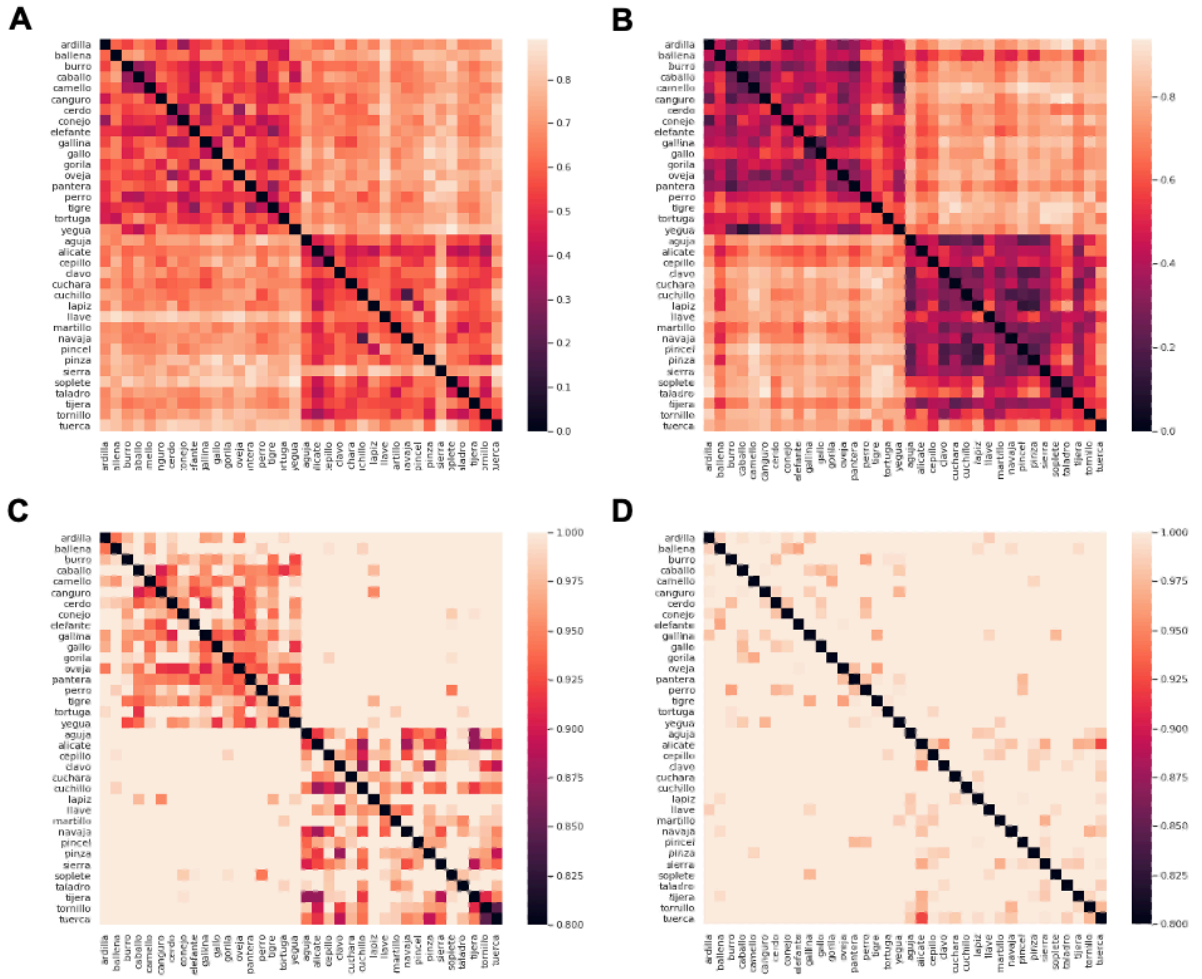

Figure 7. (A) Representational dissimilarity matrix (RDM) of the fine-tuned Word2Vec model embeddings. (B) RDM of the fine-tuned VGG-19 model embeddings. (C) Average RDM of brain activation patterns from the searchlight analysis in the postcentral gyrus cluster for the young group. (D) Average RDM of brain activation patterns from the searchlight analysis in the postcentral gyrus cluster for the older group. RDMs represent concepts for each living and non-living category. In the figures, the top-left squares illustrate the dissimilarity among concepts within the living condition. The bottom-left squares represent the dissimilarity among concepts encompassing both the living and non-living categories. Lastly, the bottom-right squares demonstrate the dissimilarity among concepts within the non-living condition. A darker shade of red indicates a higher degree of representational similarity between concepts (ρ), while a lighter shade of red suggests a greater dissimilarity (1- ρ).

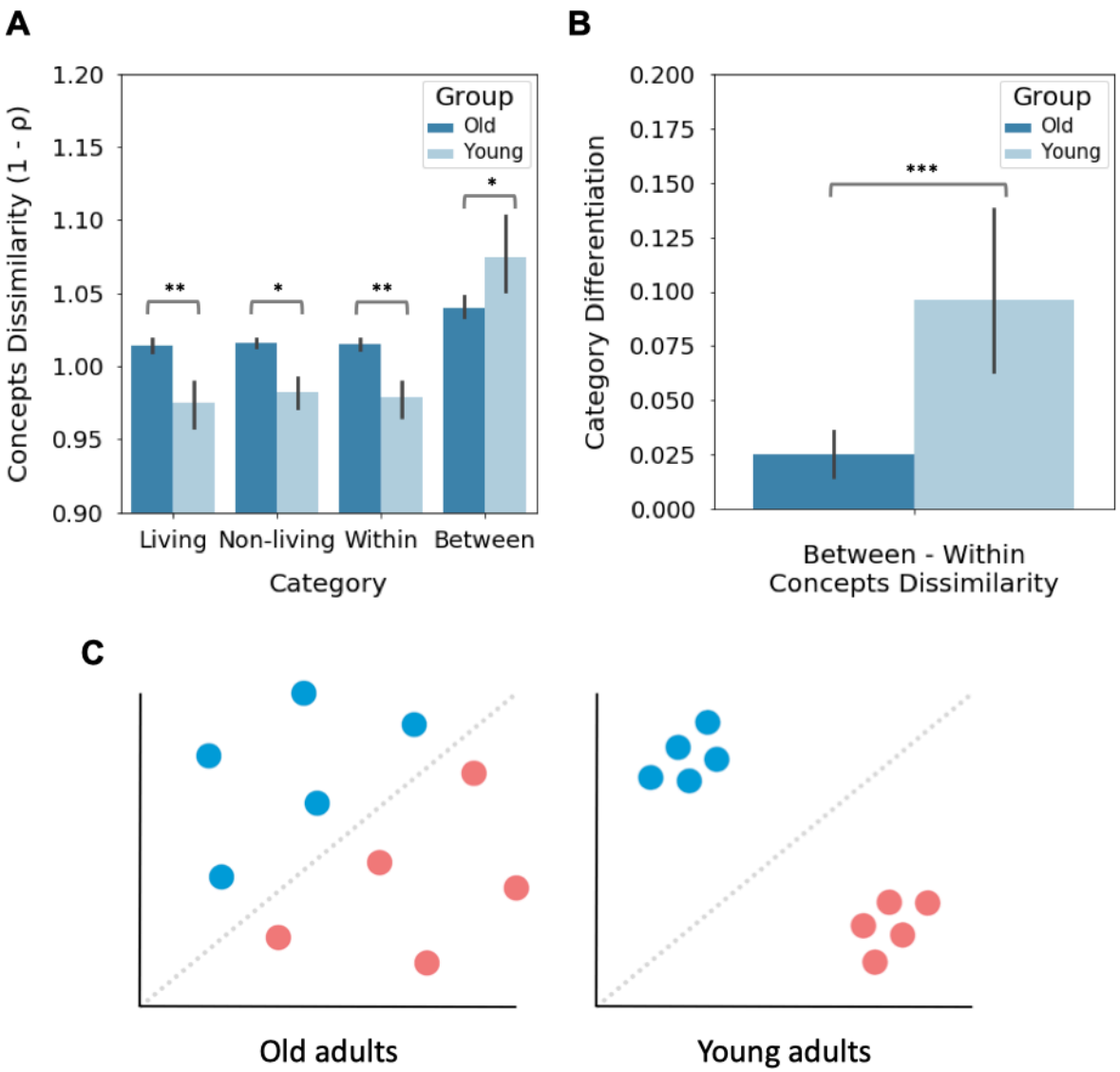

*Figure 8. (A) Comparison of average dissimilarity in conceptual representation categorized by living, non-living, and combined categories, as well as between different category concepts. (B) Quantification of category distinctiveness within each group by computing the difference between dissimilarity among concepts between categories and dissimilarity among concepts within categories. The error bars indicate the 95% confidence interval, which was adjusted using the Cousineau-Morey method (Cousineau, 2005; Morey, 2008). * $p < .05$, ** $p < .01$, *** $p < .001$. (C) The figure illustrates the hypothetical representational structure derived from RSA results for each group, depicting the dissimilarity of concepts within and between categories. Blue dots correspond to concept representations from a specific category (e.g., living), whereas red dots represent concept representations from the other category (e.g., non-living). Notably, in older adults, there is enhanced dissimilarity in conceptual representations within each category, accompanied by an increased similarity between concepts from different categories. Conversely, in young adults, the opposite pattern is observed.*

Discussion

The present fMRI study used a comprehensive approach including pattern classification, computational models, and RSA analyses with the goal of defining the impact of aging on the brain representation of living and non-living semantic categories and how this varies across different levels of processing associated with the task.

While young and older groups did not differ in the level of decoding accuracy within each task, we observed that decoding accuracy in the cross-condition generalization analyses was significantly higher in older adults compared to the young group. Furthermore, model-based RSA indicated that the brain representations related to animacy in the young group closely resembled the fine-grained distinction of living and non-living concepts according to the language model and computer vision model. However, older participants displayed a reduction in pattern similarity relative to the models' RDMs in the superior parietal during the deep processing condition.

Both cross-condition decoding, and RSA results can be interpreted in light of the neural dedifferentiation hypothesis of aging (Koen & Rugg, 2019), according to which aging leads to a reduction in the distinctiveness of brain activity patterns. Previous work used MVPA methods to assess age-related changes in neural representations in visual processing (Carp, Park, Polk, et al., 2011), motor responses (Carp, Park, Hebrank, et al., 2011), episodic memory (Abdulrahman et al., 2017) and working memory (Carp et al., 2010). These investigations have shown that, compared to younger individuals, older adults tend to display less specific brain activity patterns for task relevant representations in various domains (e.g., for visual categories such as faces and houses; Park et al., 2010).

The cross-condition decoding results indicate that the brain representations of semantic categories become more similar with aging across different task conditions which differ in the depth of processing, specifically in the left inferior parietal and middle temporal lobes that are known to be implicated in the storage of lexico-semantic information during language processing (Binder & Desai, 2011). These results indicate that aging leads to generalized neural representations of semantic knowledge. However, it may be argued that the higher cross-condition generalization observed in older adults could be attributed to carry-over effects between conditions. Previous studies have shown that older adults tend to exhibit reduced attentional control and executive functions (Paxton et al., 2008; Verhaeghen et al., 2003; Verhaeghen & Cerella, 2002; Wasylyshyn et al., 2011). Consequently, one could make the case that older adults unintentionally engaged in mental simulation during the covert reading task, and vice versa, due to task-switching costs across the different fMRI runs. However, this explanation seems unlikely. First, it should be noted that, although we included catch trials, due to the covert nature of our tasks, we could only monitor the quality of

participants' performance by means of neural decoding. However, we implemented several strategies to ensure participants engaged in the tasks in an active manner. For example, participants received thorough instructions and engaged in practice sessions before scanning to familiarize themselves with the task requirements. Following each fMRI run, regular check-ins and debriefings were held to respond to any queries from participants and to evaluate their engagement. Furthermore, the inclusion of catch trials should increase participants' compliance with the instructions and the data indeed showed this to be the case. Second, if carry-over effects were present, they would have been noticeable in the shallow processing condition due to switching costs from the previous deep processing run. This scenario should have resulted in significant decoding accuracies among older adults in the shallow condition during within-condition decoding analyses. However, these expected results were not evident. Older adults did not exhibit significant differences in within-condition analyses when compared to young adults, in both shallow and deep processing conditions, indicating that both age groups likely utilized similar mental strategies during task performance. Moreover, our analysis of catch trials did not reveal any significant differences between age groups, suggesting similar levels of attention and engagement in the specific conditions of our experiment. An additional potential concern relates to the fact that the different concepts reappeared across runs and the possibility of age-related differences in neural repetition suppression (Bergerbest at al., 2009). It may be argued that these age-related differences in neural repetition suppression could impact on the decoding results. However, the absence of age-related differences in decoding accuracy within the shallow or deep processing conditions indicates that the age-related differences observed in the cross-condition analyses are unlikely to be mediated by differences in neural repetition suppression effects.

Regarding the RSA, a recent fMRI study by Deng et al. (2021) investigated age-related differences in visuo-perceptual and mnemonic representations of complex scenes. They compared the initial and the penultimate hidden layers of a computer vision model (VGG-16; Simonyan & Zisserman, 2014) with the participants' brain representations of the images. The results indicated that brain activity patterns in the early visual cortex were less aligned with the image model based on the initial layer, in keeping with the neural dedifferentiation hypothesis. In contrast, Deng and colleagues observed a heightened correlation between brain activity patterns in the ventral anterior temporal lobe and the higher-order layer of the computer vision model in the aging group compared to young adults (see also Naspi et al., 2021, 2022). This finding was interpreted as neural hyperdifferentiation, indicating a greater dissimilarity of neural representations. On the other hand, our RSA and cross-condition decoding results provide evidence for neural dedifferentiation in the semantic domain, but we did not observe any evidence for neural hyperdifferentiation. The

differences between our RSA results and the study by Deng et al. and Naspi et al. may well relate to differences in the cognitive processes and stimuli targeted by the different studies. In their studies, participants were specifically tasked with encoding and visually recalling images of complex scenes and living and non-living images, placing a strong emphasis on visual processing and episodic memory.

Our study goes beyond the studies by Deng et al. and Naspi et al. to examine more general semantic processes, allowing participants to freely engage in multimodal mental simulations of the concepts and, therefore, encouraging the integration of information from various sensory modalities (e.g., visual, somatosensory, auditory, memory). The postcentral gyrus, serving as the primary somatosensory cortex, plays a crucial role in processing and integrating sensory information related to touch, proprioception and generating motor responses (Kropf et al., 2018; Longo et al., 2010). As individuals age, changes in body perception and action, such as tactile perception, balance, interoception, proprioception, manual motor control, and walking, are commonly observed (see Kuehn et al., 2018 for a review). These age-related declines in physical and cognitive functions can impact the integration of sensory-motor information in the postcentral gyrus associated with simulation of living and non-living word concepts. This perspective is consistent with previous research on neural dedifferentiation, which found that the distinctiveness of neural representations of motor actions was also significantly higher in young adults compared to older adults in bilateral neighboring regions to the postcentral gyrus, including the primary motor cortex and presupplementary motor area (Carp, Park, Hebrank, et al., 2011).

Intriguingly, the RSA analyses also revealed an age-related reduction in neural pattern similarity (i.e., increased distinctiveness) of item-based conceptual representations within the living and non-living categories. The formulation of neural dedifferentiation in previous studies (e.g., Carp, Park, Hebrank, et al., 2011; Deng et al., 2021) focused primarily on the comparison of neural pattern similarity within and between conditions, as proposed by Haxby et al. (2001). Our study builds upon this conceptualization but, critically, highlights the significance of addressing neural pattern similarity at different levels of semantic analysis, both at the lowest dimensional level (conceptual, item-level) and the higher-dimensional level (category) of analysis. This emphasis is particularly crucial in the semantic domain because concepts serve as meaningful units and contribute to the compositionality of language. Moreover the juxtaposition of decoding and RSA results in our study, underscores the importance of employing diverse analytical methods when investigating semantic neural dedifferentiation. While decoding analyses employing SVM models predominantly aim to maximize the margin between categories, RSA takes into account the fine-grained pattern differences across all concept pairs, offering a more comprehensive view of the representational space.

Together, our study highlights the dynamic nature of semantic representations in the aging brain and suggests that healthy aging is characterized by complex changes in neural patterns, rather than by a uniform loss of neural pattern specificity in semantic representations. Deng et al. (2021) and Naspi et al. (2022) observed that neural dedifferentiation and hyperdifferentiation can co-exist in the representation of visual categories, and that neural hyperdifferentiation may be a compensatory mechanism to counteract the diminished quality of sensory representations in healthy aging. Alternatively, we propose here that neural dedifferentiation in the semantic domain may also represent a beneficial compensatory mechanism (Seider et al., 2021), helping to maintain semantic processing in aging by allowing for a more efficient coding scheme, both compressed and sparse, to make better use of limited neural resources in healthy aging.

Acknowledgments

D.S. acknowledges support from the Basque Government through the BERC 2022-2025 program and from grant PI_2017_1_0025, and also from the Spanish State Research Agency, through the 'Severo Ochoa' Programme for Centres/Units of Excellence in R & D (CEX2020-001010-S).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems* (arXiv:1603.04467). arXiv. https://doi.org/10.48550/arXiv.1603.04467

Abdulrahman, H., Fletcher, P. C., Bullmore, E., & Morcom, A. M. (2017). Dopamine and memory dedifferentiation in aging. *NeuroImage*, *153*, 211–220. https://doi.org/10.1016/j.neuroimage.2015.03.031

Akama, H., Murphy, B., Na, L., Shimizu, Y., & Poesio, M. (2012). Decoding semantics across fMRI sessions with different stimulus modalities: A practical MVPA study. *Frontiers in Neuroinformatics*, *6*, 24. https://doi.org/10.3389/fninf.2012.00024

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Raizada, R. D. S., Lin, F., & Lalor, E. C. (2019). An Integrated Neural Decoder of Linguistic and Experiential Meaning. *Journal of Neuroscience*, *39*(45), 8969–8987. https://doi.org/10.1523/JNEUROSCI.2575-18.2019

Anderson, A. J., Bruni, E., Lopopolo, A., Poesio, M., & Baroni, M. (2015). Reading visually embodied meaning from the brain: Visually grounded computational models decode visual-object mental imagery induced by written text. *NeuroImage*, *120*, 309–322. https://doi.org/10.1016/j.neuroimage.2015.06.093

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In *Memory changes in healthy young and older adults* (pp. 395–409). Oxford University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91. https://doi.org/10.1016/S1364-6613(02)00029-3

Bergerbest, D., Gabrieli, J. D., Whitfield-Gabrieli, S., Kim, H., Stebbins, G. T., Bennett, D. A., & Fleischman, D. A. (2009). Age-associated reduction of asymmetry in prefrontal function and preservation of conceptual repetition priming. *Neuroimage, 45(1)*, 237-246.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex (New York, NY)*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Brosseau, J., & Cohen, H. (1996). The representation of semantic categories in aging. *Experimental Aging Research*, *22*(4), 381–391. https://doi.org/10.1080/03610739608254018

Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., & Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain and Language*, *120*(3), 282–289. https://doi.org/10.1016/j.bandl.2011.09.003

Cardellino, C. (2016). Spanish billion words corpus and embeddings. *Cristian Cardellino*. https://crscardellino.github.io/SBWCE/

Carlin, J. D., Calder, A. J., Kriegeskorte, N., Nili, H., & Rowe, J. B. (2011). A Head View-Invariant Representation of Gaze Direction in Anterior Superior Temporal Sulcus. *Current Biology*, *21*(21), 1817–1821. https://doi.org/10.1016/j.cub.2011.09.025

Carp, J., Gmeindl, L., & Reuter-Lorenz, P. (2010). Age Differences in the Neural Representation of Working Memory Revealed by Multi-Voxel Pattern Analysis. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00217

Carp, J., Park, J., Hebrank, A., Park, D. C., & Polk, T. A. (2011). Age-Related Neural Dedifferentiation in the Motor System. *PLOS ONE*, *6*(12), e29411. https://doi.org/10.1371/journal.pone.0029411

Carp, J., Park, J., Polk, T. A., & Park, D. C. (2011). Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. *Neuroimage*, *56*(2), 736–743.

https://doi.org/10.1016/j.neuroimage.2010.04.267

Chollet, F. (2015). *Keras*. https://keras.io/

Cosgrove, A. L., Beaty, R. E., Diaz, M. T., & Kenett, Y. N. (2023). Age differences in semantic network structure: Acquiring knowledge shapes semantic memory. *Psychology and Aging*, *38*(2), 87–102. https://doi.org/10.1037/pag0000721

Cosgrove, A. L., Kenett, Y. N., Beaty, R. E., & Diaz, M. T. (2021). Quantifying flexibility in thought: The resiliency of semantic networks differs across the lifespan. *Cognition*, *211*, 104631. https://doi.org/10.1016/j.cognition.2021.104631

Costello, M. C., & Bloesch, E. K. (2017). Are Older Adults Less Embodied? A Review of Age Effects through the Lens of Embodied Cognition. *Frontiers in Psychology*, *8*, 267. https://doi.org/10.3389/fpsyg.2017.00267

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Deng, L., Davis, S. W., Monge, Z. A., Wing, E. A., Geib, B. R., Raghunandan, A., & Cabeza, R. (2021). Age-related dedifferentiation and hyperdifferentiation of perceptual and mnemonic representations. *Neurobiology of Aging*, *106*, 55–67. https://doi.org/10.1016/j.neurobiolaging.2021.05.021

Devereux, B. J., Clarke, A., & Tyler, L. K. (2018). Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-28865-1

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, *53*(8), 1560–1570. https://doi.org/10.1037/dev0000347

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*(4), 1246–1258. https://doi.org/10.3758/s13428-013-0326-1

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, *9*, 1871–1874. https://www.jmlr.org/papers/volume9/fan08a/fan08a.pdf

Fei-Fei, L., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, *106*(1), 59–70. https://doi.org/10.1016/j.cviu.2005.09.012

Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., & Binder, J. R. (2016). Heteromodal Cortical Areas Encode Sensory-Motor Features of Word Meaning. *Journal of Neuroscience*, *36*(38), 9763–9769. https://doi.org/10.1523/JNEUROSCI.4095-15.2016

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(6), e2108091119. https://doi.org/10.1073/pnas.2108091119

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state." *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Haitas, N., Amiri, M., Wilson, M., Joanette, Y., & Steffener, J. (2021). Age-preserved semantic memory and the CRUNCH effect manifested as differential semantic control networks: An fMRI study. *PLOS ONE*, *16*(6), e0249948. https://doi.org/10.1371/journal.pone.0249948

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics*, *7*(1), 37–53. https://doi.org/10.1007/s12021-008-9041-y

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), Article 7. https://doi.org/10.1038/nrn1931

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 630–645). Springer International Publishing. https://doi.org/10.1007/978-3-319-46493-0_38

Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *NeuroImage*, *180*, 4–18. https://doi.org/10.1016/j.neuroimage.2017.08.005

Hoffman, P., & Morcom, A. M. (2018). Age-related changes in the neural networks supporting semantic cognition: A meta-analysis of 47 functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *84*, 134–150. https://doi.org/10.1016/j.neubiorev.2017.11.010

JASP Team. (2023). *JASP* (0.17.2) [Computer software]. https://jasp-stats.org

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1016/s1053-8119(02)91132-8

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6

Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience*, *8*, 59. https://doi.org/10.3389/fnhum.2014.00059

Kaplan, J. T., Man, K., & Greening, S. G. (2015). Multivariate cross-classification: Applying machine learning techniques to characterize abstraction in neural representations. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00151

Kenett, Y., & Thompson-Schill, S. (2020). *Novel conceptual combination can dynamically reconfigure semantic memory network*. https://doi.org/10.31234/osf.io/crp47

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*. https://doi.org/10.48550/arXiv.1412.6980

Kintz, S., & Wright, H. H. (2017). Semantic knowledge use in discourse: Influence of age. *Discourse Processes*, *54*(8), 670–681. https://doi.org/10.1080/0163853X.2016.1150652

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-Normalizing Neural Networks. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/hash/5d44ee6f2c3f71b73125876103c8f6c4-Abstract.html

Koen, J. D., & Rugg, M. D. (2019). Neural Dedifferentiation in the Aging Brain. *Trends in Cognitive Sciences*, *23*(7), 547–559. https://doi.org/10.1016/j.tics.2019.04.012

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kropf, E., Syan, S. K., Minuzzi, L., & Frey, B. N. (2018). From anatomy to function: The role of the somatosensory cortex in emotional regulation. *Brazilian Journal of Psychiatry*, *41*, 261–269. https://doi.org/10.1590/1516-4446-2018-0183

Kuehn, E., Perez-Lopez, M. B., Diersch, N., Döhler, J., Wolbers, T., & Riemer, M. (2018). Embodiment in the aging mind. *Neuroscience and Biobehavioral Reviews*, *86*, 207–225. https://doi.org/10.1016/j.neubiorev.2017.11.016

Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., Chen, P.-H. C., Ellis, C. T., Henselman-Petrusek, G., Huberdeau, D., Hutchinson, J. B., Li, Y. P., Lu, Q., Manning, J. R., Mennen, A. C., Nastase, S. A., Richard, H., Schapiro, A. C., Schuck, N. W., … Norman, K. A. (2021). BrainIAK: The Brain Imaging Analysis Kit. *Aperture Neuro*, *1*(4).

https://www.doi.org/10.52294/31bb5b68-2184-411b-8c00-a1dacb61e1da

Kumar, M., Ellis, C. T., Lu, Q., Zhang, H., Capotă, M., Willke, T. L., Ramadge, P. J., Turk-Browne, N. B., & Norman, K. A. (2020). BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis. *PLoS Computational Biology*, *16*(1), e1007549. https://doi.org/10.1371/journal.pcbi.1007549

LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. (2012). Efficient BackProp. In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural Networks: Tricks of the Trade: Second Edition* (pp. 9–48). Springer. https://doi.org/10.1007/978-3-642-35289-8_3

León, I., García, J., & Roldán-Tapia, L. (2011). Construcción de la escala de reserva cognitiva en población española: Estudio piloto. *Revista de Neurología*, *52*, 653. https://doi.org/10.33588/rn.5211.2010704

León, I., García-García, J., & Roldán-Tapia, L. (2014). Estimating Cognitive Reserve in Healthy Adults Using the Cognitive Reserve Scale. *PLOS ONE*, *9*(7), e102632. https://doi.org/10.1371/journal.pone.0102632

León, I., García-García, J., & Roldán-Tapia, L. (2016). Escala de Reserva Cognitiva y envejecimiento. *Anales de Psicología / Annals of Psychology*, *32*(1), Article 1. https://doi.org/10.6018/analesps.32.1.182331

Lobo, A., Escobar, V., Ezquerra, J., & Seva Díaz, A. (1980). "El Mini-Examen Cognoscitivo" (Un test sencillo, práctico, para detectar alteraciones intelectuales en pacientes psiquiátricos). [The "Mini-Examen Cognoscitiuo": A simple and practical test to detect intellectual dysfunctions in psychiatric patients.]. *Revista de Psiquiatría y Psicología Médica*, *14*, 39–57.

Longo, M. R., Azañón, E., & Haggard, P. (2010). More than skin deep: Body representation beyond primary somatosensory cortex. *Neuropsychologia*, *48*(3), 655–668. https://doi.org/10.1016/j.neuropsychologia.2009.08.022

Madden, D. J. (2007). Aging and visual attention. *Current Directions in Psychological Science*, *16*(2), 70–74. https://doi.org/10.1111/j.1467-8721.2007.00478.x

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *ELife*, *7*, e31873. https://doi.org/10.7554/eLife.31873

Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I., Lavoie, E., Muller, X., Desjardins, G., Warde-Farley, D., Vincent, P., Courville, A., & Bergstra, J. (2012). Unsupervised and Transfer Learning Challenge: A Deep Learning Approach. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 97–110. https://proceedings.mlr.press/v27/mesnil12a.html

Mille, J., Brambati, S. M., Izaute, M., & Vallet, G. T. (2021). Low-Resolution Neurocognitive Aging and Cognition: An Embodied Perspective. *Frontiers in Systems Neuroscience*, *15*, 687393. https://doi.org/10.3389/fnsys.2021.687393

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, *320*(5880), 1191–1195. https://doi.org/10.1126/science.1152876

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morrow, L. I., & Duffy, M. F. (2005). The representation of ontological category concepts as affected by healthy aging: Normative data and theoretical implications. *Behavior Research Methods*, *37*(4), 608–625. https://doi.org/10.3758/bf03192731

Mourão-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *NeuroImage*, *28*(4), 980–995. https://doi.org/10.1016/j.neuroimage.2005.06.070

Naspi, L., Hoffman, P., Devereux, B., & Morcom, A. M. (2021). Perceptual and semantic representations at encoding contribute to true and false recognition of objects. *Journal of Neuroscience*, *41*(40), 8375–8389. https://doi.org/10.1523/JNEUROSCI.0677-21.2021

Naspi, L., Stensholt, C., Karlsson, A. E., Monge, Z., & Cabeza, R. (2022). Effects of aging on successful object encoding: Enhanced semantic representations compensate for impaired visual representations. *BioRxiv*, 2022–12. https://doi.org/10.1101/2022.12.10.519871

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. https://doi.org/10.1002/hbm.1058

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Park, J., Carp, J., Hebrank, A., Park, D. C., & Polk, T. A. (2010). Neural Specificity Predicts Fluid Processing Ability in Older Adults. *Journal of Neuroscience*, *30*(27), 9253–9259. https://doi.org/10.1523/JNEUROSCI.0853-10.2010

Paxton, J. L., Barch, D. M., Racine, C. A., & Braver, T. S. (2008). Cognitive Control, Goal Maintenance, and Prefrontal Function in Healthy Aging. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(5), 1010–1028. https://doi.org/10.1093/cercor/bhm135

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830. https://scikit-learn.org

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064

Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., & Macrae, C. N. (2011). Puddles, parties, and professors: Linking word categorization to neural patterns of visuospatial coding. *Journal of Cognitive Neuroscience*, *23*(10), 2636–2649. https://doi.org/10.1162/jocn.2011.21628

Rami, L., Valls-Pedret, C., Bartrés-Faz, D., Caprile, C., Solé-Padullés, C., Castellvi, M., Olives, J., Bosch, B., & Molinuevo, J. L. (2011). Cognitive reserve questionnaire. Scores obtained in a healthy elderly population and in one with Alzheimer's disease. *Revista de neurologia*, *52*(4), 195–201. https://doi.org/10.33588/rn.5204.2010478

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary clinical validity. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 310–319. https://doi.org/10.1076/jcen.20.3.310.823

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*. https://proceedings.neurips.cc/paper_files/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current Directions in Psychological Science*, *17*(3), 177–182. https://doi.org/10.1111/j.1467-8721.2008.00570.x

Ritchie, J. B., Kaplan, D. M., & Klein, C. (2019). Decoding the Brain: Neural Representation and the Limits of Multivariate Pattern Analysis in Cognitive Neuroscience. *The British Journal for the Philosophy of Science*, *70*(2), 581–607. https://doi.org/10.1093/bjps/axx023

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, Growth, and Decline in Adult Life Span Development of Declarative Memory: Cross-Sectional and Longitudinal Data From a Population-Based Study. *Psychology and Aging*, *20*(1), 3–18. https://doi.org/10.1037/0882-7974.20.1.3

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A.,

Khosla, A., & Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211–252. https://doi.org/10.1007/s11263-015-0816-y

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*(1–3), 35–54. https://doi.org/10.1016/s0301-0511(00)00052-1

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*, *57*(61), 10–25080. https://doi.org/10.25080/Majora-92bf1922-011

Seider, T. R., Porges, E. C., Woods, A. J., & Cohen, R. A. (2021). Dedifferentiation of functional brain activation associated with greater visual discrimination accuracy in middle-aged and older adults. *Frontiers in Aging Neuroscience*, *13*, 651284. https://doi.org/10.3389/fnagi.2021.651284

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. *NeuroImage*, *54*(3), 2418–2425. https://doi.org/10.1016/j.neuroimage.2010.10.042

Simanova, I., Hagoort, P., Oostenveld, R., & van Gerven, M. A. J. (2014). Modality-Independent Decoding of Semantic Information from the Human Brain. *Cerebral Cortex*, *24*(2), 426–434. https://doi.org/10.1093/cercor/bhs324

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

Smith, J., & Theobald, C. (2013). *MRIConvert* (2.1.0) [Computer software]. lcni.uoregon.edu/downloads/mriconvert

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Soto, D., Sheikh, U. A., Mei, N., & Santana, R. (2020). Decoding and encoding models reveal the role of mental simulation in the brain representation of meaning. *Royal Society Open Science*, *7*(5), 192043. https://doi.org/10.1098/rsos.192043

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 101–117. https://doi.org/10.1037/0278-7393.32.1.101

Tong, F., & Pratte, M. S. (2012). Decoding Patterns of Human Brain Activity. *Annual Review*

of Psychology, *63*(1), 483–509. https://doi.org/10.1146/annurev-psych-120710-100412

Turner, G. R., & Spreng, R. N. (2012). Executive functions and neurocognitive aging: Dissociable patterns of brain activity. *Neurobiology of Aging*, *33*(4), 826.e1-13. https://doi.org/10.1016/j.neurobiolaging.2011.06.005

Vallet, G. T. (2015). Embodied cognition of aging. *Frontiers in Psychology*, *6*, 463. https://doi.org/10.3389/fpsyg.2015.00463

van den Hurk, J., & Op de Beeck, H. P. (2019). Generalization asymmetry in multivariate cross-classification: When representation A generalizes better to representation B than B to A. *BioRxiv*, 592410. https://doi.org/10.1101/592410

van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, *114*(22), E4501–E4510. https://doi.org/10.1073/pnas.1612862114

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience & Biobehavioral Reviews*, *26*(7), 849–857. https://doi.org/10.1016/S0149-7634(02)00071-4

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A meta-analysis. *Psychology and Aging*, *18*(3), 443. https://doi.org/10.1037/0882-7974.18.3.443

Verheyen, S., Droeshout, E., & Storms, G. (2019). Age-Related Degree and Criteria Differences in Semantic Categorization. *Journal of Cognition*, *2*(1), 17. https://doi.org/10.5334/joc.74

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, X., Wu, W., Ling, Z., Xu, Y., Fang, Y., Wang, X., Binder, J. R., Men, W., Gao, J.-H., & Bi, Y. (2018). Organizational Principles of Abstract Words in the Human Brain. *Cerebral Cortex*, *28*(12), 4305–4318. https://doi.org/10.1093/cercor/bhx283

Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. (2011). Aging and task switching: A meta-analysis. *Psychology and Aging*, *26*(1), 15. https://psycnet.apa.org/doi/10.1037/a0020912

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*(100), 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060

Woolrich, M. W., Behrens, T. E. J., & Smith, S. M. (2004). Constrained linear basis sets for

HRF modelling using Variational Bayes. *NeuroImage*, *21*(4), 1748–1761. https://doi.org/10.1016/j.neuroimage.2003.12.024

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Wulff, D. U., De Deyne, S., Jones, M. N., & Mata, R. (2019). New perspectives on the aging lexicon. *Trends in Cognitive Sciences*, *23*(8), 686–698. https://doi.org/10.1016/j.tics.2019.05.003

Wulff, D. U., Hills, T. T., & Mata, R. (2022). Structural differences in the semantic networks of younger and older adults. *Scientific Reports*, *12*(1), 21459. https://doi.org/10.1038/s41598-022-11698-4

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). *How transferable are features in deep neural networks?* (arXiv:1411.1792). arXiv. https://doi.org/10.48550/arXiv.1411.1792

Supplementary materials

Supplementary table

		Living	Non-living
Word Frequencies (per million of words)	Range	1 – 82.49	0 – 66.90
	Mean	16,51	8,42
	SD	23,00	15,73
Number of Letters	Range	5 - 8	5 - 8
	Mean	6,17	6,33
	SD	1,04	1,08
Number of Phonemes	Range	4 - 8	4 - 7
	Mean	5,72	5,89
	SD	1,13	0,90
Number of Syllables	Range	2 - 4	2 - 4
	Mean	2,72	2,67
	SD	0,57	0,59
Orhographic Neighbourhood Size (per million of words)	Range	0 - 17	0 - 9
	Mean	5,05	3,94
	SD	5,24	2,43
Phonological Neighbourhood Size (per million of words)	Range	1 - 42	1 - 21
	Mean	11,72	8,61
	SD	12,15	4,91
Familiarity	Range	5 – 6.39	5 – 6,87
	Mean	5,93	5,98
	SD	0,40	0,54
Imageability	Range	5 – 6.71	6 – 6,615
	Mean	6,18	6,01
	SD	0,46	0,29
Concreteness	Range	5 – 6.69	5 – 6.77
	Mean	6,15	6,09
	SD	0,45	0,46

Table S1. Summary of living and non-living Spanish words' properties gathered from EsPal database (bcbl.eu/databases/espal/; Duchon et al., 2013). There were a few words (12) for which some familiarity, imageability, and concreteness ratings were not found on this database. Reported range, mean and SD of ratings for familiarity, imageability and concreteness were calculated only for the words for which it was found.

Supplementary results

Within and cross-condition decoding analyses

We compared the decoding accuracies obtained from within and cross-condition generalization analyses using ANOVA. These analyses involved investigating the effects of various factors and their corresponding levels. Specifically, for within-condition analyses, the factors included 15 ROIs, 2 conditions (shallow vs. deep), 2 hemispheres (left vs. right), and 2 age groups (young vs. older). In the case of cross-condition analyses, the factors consisted of 15 ROIs, 2 cross-condition directions (shallow-to-deep, deep-to-shallow), 2 hemispheres (left vs. right), and 2 age groups (young vs. older). Although our primary focus revolved around exploring the impact of age as a factor within and across conditions when decoding multi-voxel patterns for unseen word pairs, this section provides a detailed exploration and discussion of the ANOVA results pertaining to additional factors.

Increased decoding of semantic categories in the left hemisphere in both within condition and cross-condition analyses

In the within-condition decoding analysis, we observed a significant main effect of the hemisphere factor [F(1, 54) = 42.905, p < .001, η_p^2 = .443]. Decoding accuracies were significantly higher in the left hemisphere (Mean = .536, SE = .005) compared to the right hemisphere (Mean = .524, SE = .005) (Figure S1). Additionally, we found interactions between the hemisphere factor and other variables: hemisphere * ROIs [F(11.856, 640.216) = 7.057, p < .001, η_p^2 = .116], hemisphere * condition [F(1, 54) = 15.975, p < .001, η_p^2 = .228], and hemisphere * condition * ROIs [F(11.927, 644.039) = 2.342, p = .006, η_p^2 = .042]. In the subsequent paragraph, our primary objective is to delve deeper into the interaction between hemisphere, condition, and ROIs. Our aim is to examine the variations in decoding accuracies between hemispheres within both the shallow and deep conditions for each specific ROI.

We conducted two additional mixed-effects ANOVA analyses, one in the shallow condition and another in the deep condition using 15 ROIs, 2 hemispheres, and 2 age groups as factors. The DOF for these ANOVAs were corrected using Huynh-Feldt adjustment. In both the shallow condition [F(1,54) = 7.240, p = .009, η_p^2 = .118] and the deep condition [F(1, 54) = 45.795, p < .001, η_p^2 = .459], we found a significant main effect of the hemisphere. This indicates that overall decoding accuracies were higher in the left hemisphere [Shallow: Mean = .513, SE = .006; Deep: Mean = .560, SE = .008] compared to the right hemisphere [Shallow: Mean = .507, SE = .006; Deep: Mean = .541, SE = .008]. Furthermore, in the deep condition, we also observed a significant hemisphere * ROIs interaction

[F(12.099, 653.331) = 7.351, p < .001, η_p^2 = .120]. To analyze this interaction effect, we conducted post-hoc pairwise comparisons, averaging over the age levels. Within each ROI, paired two-tailed t-tests were performed to compare decoding accuracies between the left and right hemispheres in the deep condition, with p-values corrected using FDR for multiple comparisons. In the deep condition, decoding accuracies were significantly higher in the left hemisphere compared to the right hemisphere in the following ROIs: fusiform gyrus [Left: Mean = .578, SE = .013; Right: Mean = .557, SE = .012, t(54) = 2.813, p = .017], inferior parietal lobe [Left: Mean = .607, SE = .013; Right: Mean = .570, SE = .011, t(54) = 5.193, p < .001], middle temporal lobe [Left: Mean = .592, SE = .011; Right: Mean = .556, SE = .011, t(54) = 5.065, p < .001], superior frontal gyrus [Left: Mean = .578, SE = .011; Right: Mean = .552, SE = .010, t(54) = 4.986, p < .001], inferior temporal lobe [Left: Mean = .613, SE = .013; Right: Mean = .550, SE = .012, t(54) = 8.40, p < .001], and pars opercularis [Left: Mean = .577, SE = .012; Right: Mean = .545, SE = 9.2e-3, t(54) = 3.980, p < .001] (Figure 3).

In the cross-condition decoding approach, we observed a significant main effect of the hemisphere factor [F(1, 54) = 33.411, p < .001, η_p^2 = .382]. Overall, cross-condition decoding accuracies were higher in the left hemisphere (Mean = .526, SE = .003) compared to the right hemisphere (Mean = .511, SE = .003) (Figure S2). Additionally, we identified two interactions: hemisphere * ROIs [F(11.562, 624.340) = 4.921, p < .001, η_p^2 = .084] and hemisphere * cross-condition direction [F(1, 54) = 6.919, p < .001, η_p^2 = .114].

First, we examined the hemisphere * ROIs interaction effect by conducting post-hoc pairwise comparisons, averaging across the levels of age and cross-condition direction. Paired two-tailed t-tests, with p-values corrected using the FDR for multiple comparisons, were performed within each ROI to compare decoding accuracies between the left and right hemispheres. We found significantly larger cross-condition decoding accuracies in the left hemisphere when compared to the right hemisphere in the fusiform gyrus [Left: Mean = .545 SE = 6.4e-3, Right: Mean = .514 SE = 5.4e-3, t(54) = 4.855, p < .001], inferior parietal lobe [Left: Mean = .548 SE = 6.7e-3, Right: Mean = .512 SE = 6.1e-3, t(54) = 4.750, p < .001], middle temporal lobe [Left: Mean = .543 SE = 7.0e-3, Right: Mean = .515. SE = 4.5e-3, t(54) = 3.742, p < .001], inferior temporal lobe [Left: Mean = .545 SE = 5.8e-3, Right: Mean = .512 SE = 5.4e-3, t(54) = 4.589, p < .001], pars opercularis [Left: Mean = .531 SE = 6.1e-3, Right: Mean = .502 SE = 4.8e-3, t(54) = 3.978, p < .001].

Next, we analyzed the hemisphere by cross-condition direction interaction. Post-hoc pairwise comparisons were conducted by averaging across the levels of age and ROIs. Paired two-tailed t-tests were performed within each ROI to compare differences between hemispheres within each cross-condition direction. Cross-condition decoding was higher in the left hemisphere compared to

the right hemisphere for both cross-condition direction: shallow-to-deep [Left: Mean =.531 SE = 3.8e-3, Right: Mean = .513 SE = 3.3e-3, t(54) = 5.553, p < .001] and deep-to-shallow [Left: Mean = .521 SE = 2.6e-3, Right: Mean = .510 SE = 2.3e-3, t(54) = 5.060, p < .001] (Figure S2).

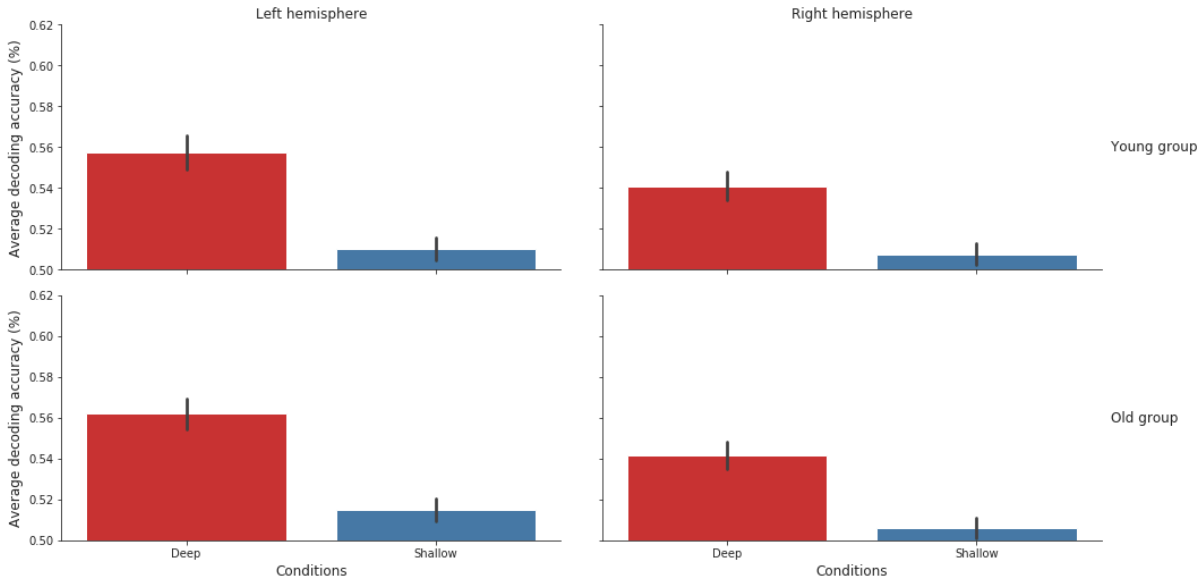

Figure S1. Average decoding accuracy obtained from the within condition analysis by hemisphere, age-group, and condition. The error bars represent a 95% confidence interval.

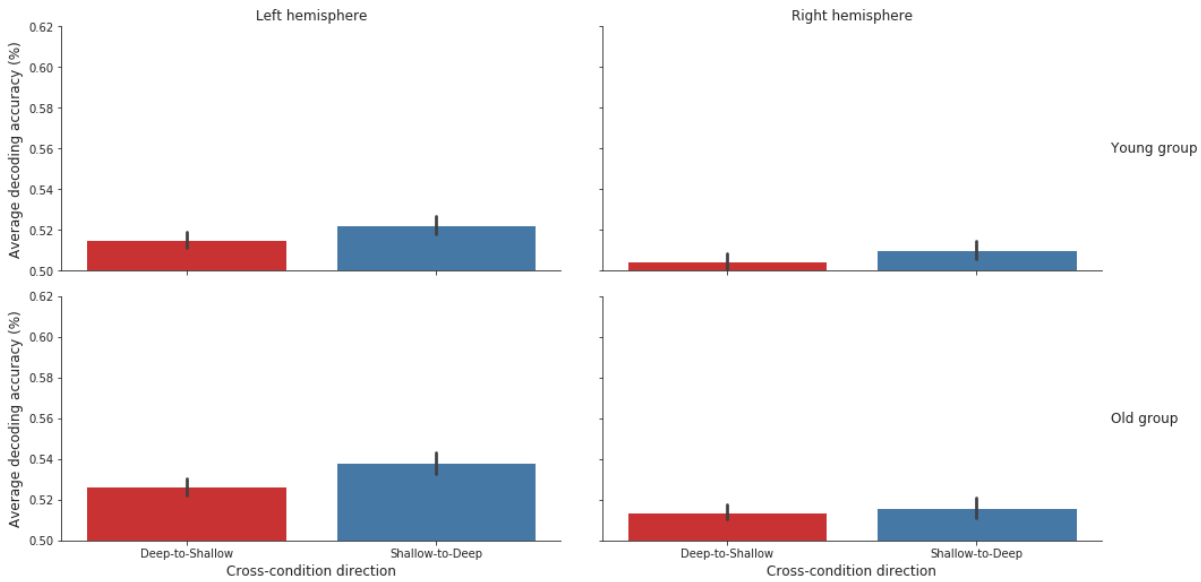

Figure S2. Average decoding accuracy obtained from the cross-condition analysis by hemisphere, age-group, and condition. The error bars represent a 95% confidence interval.

Greater decoding accuracies are observed in the deep compared to the shallow condition in the context of within condition decoding analyses

In the within-condition decoding analysis, we observed a significant main effect of condition [F(1, 54) = 18.553, p < .001, η_p^2 = .256]. Decoding accuracies were significantly higher in the deep (Mean = .551, SE = .007) compared to the shallow condition (Mean = .510, SE = .007) (Figure S1). Additionally, we found interactions between the condition factor and other variables: condition * ROIs [F(6.328, 341.730) = 8.794, p < .001, η_p^2 = .14], condition * hemisphere [F(1, 54) = 15.975, p < .001, η_p^2 = .228], and condition * hemisphere * ROIs [F(11.927, 644.039) = 2.342, p = .011, η_p^2 = .042]. As mentioned in the hemisphere and interaction effects section, we aimed to further explore the hemisphere * condition * ROIs interaction observed in the within-condition decoding analysis. In the following paragraph, our focus is on examining how decoding accuracies differ between conditions within each hemisphere for each ROI.

To investigate this, we conducted two additional mixed-effects ANOVA analyses, one for the left hemisphere and another for the right hemisphere, using 15 ROIs, 2 conditions, and 2 age groups as factors. The DOF for these ANOVAs were corrected using the Huynh-Feldt method. In both the left [F(1, 54) = 24.346, p < .001, η_p^2 = .311] and right [F(1, 54) = 12.736, p < .001, η_p^2 = .191] hemisphere ANOVAs, we found a significant main effect of condition. Overall, decoding accuracies were higher in the deep condition [Left: Mean = .560, SE = .007; Right: Mean = .541, SE = .007] compared to the shallow condition [Left: Mean = .513, SE = .007; Right: Mean = .507, SE = .007] (Figure S1). Furthermore, we observed an interaction between condition and ROIs in both the left [F(7.436, 401.568) = 9.261, p < .001, η_p^2 = .146] and right [F(8.830, 476.797) = 4.062, p < .001, η_p^2 = .070] hemispheres. To explore these interactions, pairwise comparisons were performed within each hemisphere, comparing the shallow and deep conditions in each ROI while averaging over the levels of age. Post-hoc paired two-tailed t-tests with p-values corrected using the FDR for multiple comparisons were conducted.

In the left hemisphere, decoding accuracies were significantly higher in the deep condition compared to the shallow condition in almost all ROIs, except for the middle orbitofrontal gyrus [Deep: Mean = .530 SE = 8.5e-3, Shallow: Mean = .516, SE = 7.9e-3, t(54) = 1.203, p = .233] and the temporal pole [Deep: Mean = .518 SE = 7.3e-3, Shallow: Mean = .502, SE = 7.3e-3, t(54) = 1.348, p = .196]. Similarly, in the right hemisphere, decoding accuracies were significantly higher in the deep condition compared to the shallow condition in all ROIs except for the parahippocampal gyrus [Deep: Mean = .522 SE = 9.7e-3, Shallow: Mean = .507 SE = 6.8e-3, t(54) = 1.210, p = .248], posterior

cingulate [Deep: Mean = .536 SE = 9.4e-3, Shallow: Mean = .514 SE = 7.4e-3, t(54) = 1.980, p = .061], and temporal pole [Deep: Mean = .504 SE = 6.6e-3, Shallow: Mean = .501 SE = 7.3e-3, t(54) = .274, p = .785] (Figure 3).

Asymmetries in the cross-condition direction generalization

We also note that the analysis of cross-condition decoding accuracies using a 15 (ROIs) * 2 cross-condition direction * 2 hemisphere * 2 age mixed-effects ANOVA with Huynh-Feldt corrected DOF revealed significant asymmetries in decoding directionality (van den Hurk & Op de Beeck, 2019). The results showed that decoding accuracies were higher when the decoder was trained on the shallow processing condition and tested on the deep processing condition (i.e., shallow-to-deep; Mean = .522, SE = .003), compared to the reverse direction (i.e., deep-to-shallow; Mean = .515, SE = .003) [F(1, 54) = 9.557, p = .003, η_p^2 = .150] particularly in the left hemisphere (Figure S2). This finding appears counterintuitive given that we did not observe significant decoding of semantic categories above chance level within the shallow condition, and it remains unclear how the trained classifier in the shallow condition could generalize effectively to the deep condition. Furthermore, considering the high decoding accuracy in the deep condition, one might have expected better generalization from deep-to-shallow rather than shallow-to-deep. Similar asymmetries in decoding directionality have been reported in previous studies (e.g., Akama et al., 2012; Johnson & Johnson, 2014; Quadflieg et al., 2011; van den Hurk et al., 2017). Building on the work of van den Hurk & Op de Beeck (2019), we propose that these cross-classification decoding asymmetries may not be directly linked to experimental or neurobiological factors, but rather to differences in SNR ratios between the datasets of shallow processing and deep processing conditions. Van den Hurk & de Beeck demonstrated that a dataset with higher SNR yielded good within-condition classifications but poor generalization to a dataset with lower SNR, and vice versa. They suggested that the orientation of a soft margin SVM hyperplane, trained on a dataset with high SNR, might be more sensitive to adaptations of support vectors in the test set, leading to poor generalization in datasets with higher noise levels. This trend has been observed in other studies as well (Akama et al., 2012; Johnson & Johnson, 2014; Quadflieg et al., 2011; van den Hurk et al., 2017), aligning with our own findings.

Despite the absence of interaction effects between cross-condition direction and age factors in our main ANOVA [F(1, 54) = .015, p = .904], we conducted an additional ANOVA with 15 ROIs * 2 hemispheres * 2 age groups as factors specifically in the deep-to-shallow cross-classification direction in order to address potential biases in our interpretations due to SNR. Importantly, we observed a main effect of age (F(1, 54) = 5.613, p = .021, η_p^2 = .094), indicating that deep-to-shallow

decoding was higher in the older group (Mean = .520, SE = .003) compared to the young group (Mean = .510, SE = .003) (Figure S3).

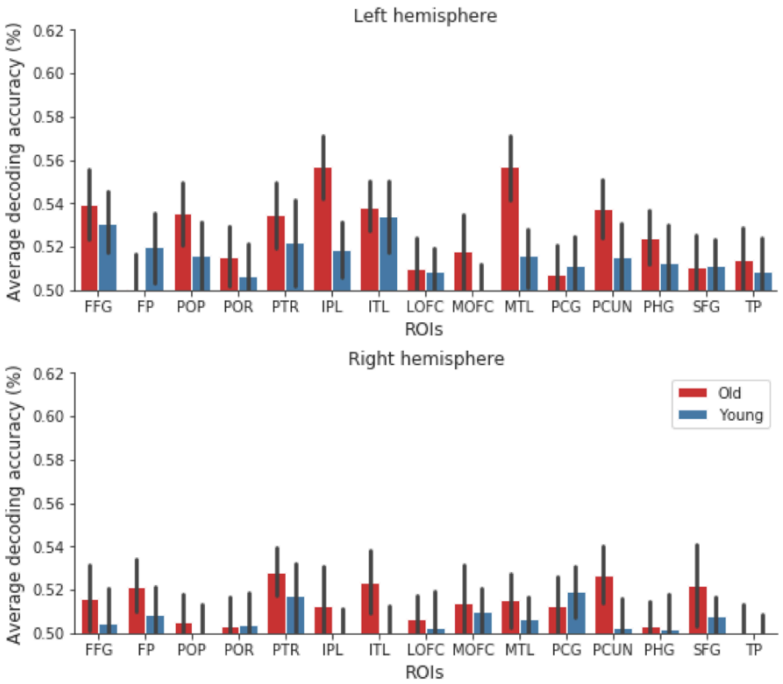

Figure S3. Cross-condition decoding accuracies in deep-to-shallow condition by ROI, hemisphere, and age-group. The error bars represent a 95% confidence interval.

<u>The inferior parietal lobe, middle temporal gyrus and inferior temporal gyrus exhibited the highest decoding accuracies in both within-condition and cross-condition analyses</u>

Given that the previous sections have already provided detailed information on the interactions between ROIs and other factors, we will now provide a brief overview of the overall decoding accuracy across the 15 ROIs, averaged over all levels of other factors (i.e., condition or cross-condition, hemisphere, age). In the within-condition decoding analysis, we observed a significant main effect of ROIs [F(4.682, 253.377) = 2.145, p < .001, η_p^2 = .205]. The ROIs with the highest decoding accuracies were the inferior parietal lobe (Mean = .552, SE = 7e-3), inferior temporal gyrus (Mean = .545, SE = 6.7e-3), middle temporal gyrus (Mean = .543, SE = 6.5e-3), and fusiform gyrus (Mean = .540, SE = 6.8e-3). On the other hand, the ROIs with the lowest decoding accuracies were the parahippocampal gyrus (Mean = .517, SE = 5.1e-3), temporal pole (Mean = .506, SE = 3.7e-3), and frontal pole (Mean = .506, SE = 4.4e-3) (Figure 3).

In the cross-condition decoding analysis, we observed a significant main effect of ROIs [F(4.682, 253.377) = 2.145, p < .001, η_p^2 = .205]. The ROIs with the highest cross-condition decoding

accuracies were the fusiform gyrus (Mean = .540, SE = 5e-3), inferior parietal lobe (Mean = .530, SE = 5.2e-3), middle temporal gyrus (Mean = .529, SE = 4.6e-3), and inferior temporal gyrus (Mean = .528, SE = 4.3e-3). Conversely, the ROIs with the lowest cross-condition decoding accuracies were the frontal pole (Mean = .509, SE = 2.8e-3), lateral orbitofrontal gyrus (Mean = .507, SE = 3.9e-3), and temporal pole (Mean = .506, SE = 3.3e-3) (Figure 4).