

# An Experiential Account of Semantic Category Organization in the Brain

By  
Stephen Mazurchuk

*A Dissertation Submitted to the Faculty of  
the Graduate School of Biomedical Sciences of the  
Medical College of Wisconsin in  
Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy*

Advisor:  
Jeffrey R. Binder, MD

Department of Biophysics  
Medical College of Wisconsin  
Milwaukee, Wisconsin, 53226, U.S.A

2024

## Abstract

There have been extensive efforts to understand how concepts are represented in the brain. One phenomenon that has received much attention is the loss of knowledge seemingly restricted to a single category of items. This clinical observation, combined with many functional magnetic resonance imaging studies that find differential activation across the cortex depending on the category of the stimuli, has led to a general view of category related functional organization across the cortex. One class of models put forward to explain this phenomenon are experiential models of cognition. These models posit that those regions of the cortex involved in experiencing an object (e.g., perceiving an object) are also involved in representing semantic knowledge about the object. The idea that some perceptual, or experiential, information (e.g., color, sound, shape, touch, etc.) is differentially important across categories of concepts, combined with the well accepted view that some regions of the brain preferentially process a particular type of perceptual information, has long been held as a potential explanation for the category related organization of the cortex. However, while this hypothesis has motivated the development of several models of semantic cognition, the ability of a specific experiential model to explain category related organization of the cortex has hitherto not been explicitly examined. Experiential models, like the one examined in this dissertation, propose that, ultimately, all semantic knowledge is grounded in what is experienced.

This dissertation first describes and validates a dataset that is well suited for addressing some of the core predictions of experiential models of cognition. Following data validation, evidence for the categorical organization of the semantic system was found via both univariate and multivariate analysis. Then, encoding models constructed

for each point on the cortical surface demonstrated that experiential features can be used to predict both activation category contrasts, as well as activation patterns for individual stimuli across categorical boundaries. It was further observed that the ability of a comparable non-experiential semantic model to generalize across categories is largely accomplished using the same information that is in the experiential model.

The subsequent chapter examines a relatively understudied yet distinctly represented category of concepts: body parts. We found that a robust network of regions was significantly more involved in the representation of body part concepts compared to concepts from other concrete object noun categories. The implicated set of areas was shown to be predicted by experiential features using several analysis techniques. The distribution of involved regions also provides possible insight as to why semantic representations of body part concepts are often spared relative to concepts from other categories.

Lastly, feature sensitivity maps across the cortex were generated for a set of 14 experiential features. These maps were observed to hierarchically converge as the distance from canonical primary sensory cortices increased. The maps are also individually discussed with reference to prior literature. Results were mixed with regard to how well individual feature maps matched expectations derived from prior literature, and possible reasons for this are discussed.

Taken together, the results of this dissertation offer strong support for many of the core predictions of an experiential account of concept representation.

## Table of Contents

|   |           |
|---|-----------|
| <b>Abstract.....</b>  | <b>I</b>  |
| <b>Acknowledgements .....</b>   | <b>X</b>  |
| <b>Chapter 1: Introduction .....</b>  | <b>1</b>  |
| Purpose of the Study .....  | 1         |
| Scientific Relevance .....  | 1         |
| Clinical Relevance.....   | 2         |
| Historical Background.....  | 3         |
| Ancient Beginnings .....  | 3         |
| Enlightenment.....  | 6         |
| Modern Times.....   | 7         |
| Evidence of Category Related Semantic Deficits (CRSDs).....   | 10        |
| Theoretical Accounts of CRSDs .....   | 13        |
| Sensory Functional Hypothesis .....   | 14        |
| Grounded Cognition.....   | 15        |
| Domain Specificity .....  | 18        |
| Explanatory ability .....   | 21        |
| Operationalized models.....   | 25        |
| Concept representation as experiential attributes .....   | 25        |
| Distributional Models .....   | 28        |
| Approach .....  | 28        |
| <b>Chapter 2: Stimulus Repetition and Sample Size Considerations .....</b>                          | <b>33</b> |
| Introduction .....  | 33        |
| Materials and Methods .....   | 38        |
| Results .....   | 50        |
| Discussion .....  | 56        |
| <b>Chapter 3: A Shared Representational Basis Across Four Dissociable Semantic Categories .....</b> | <b>62</b> |
| Introduction .....  | 62        |
| Materials and Methods .....   | 66        |
| Results .....   | 71        |
| Discussion .....  | 79        |
| Conclusions .....   | 82        |

---

|  |            |
|--|------------|
| <b>Chapter 4: The neural representation of body part concepts.....</b> | <b>84</b>  |
| Introduction .....   | 84         |
| Materials and Methods .....  | 90         |
| Results .....  | 95         |
| Discussion .....   | 106        |
| Conclusions .....  | 117        |
| <b>Chapter 5: Selected Experiential Feature Maps .....</b>             | <b>118</b> |
| Introduction .....   | 118        |
| Generating Feature Maps.....   | 121        |
| Feature Selection .....  | 122        |
| Materials and Methods: .....   | 134        |
| Results: .....   | 136        |
| Discussion .....   | 151        |
| Conclusions .....  | 164        |
| <b>General Discussion.....</b>   | <b>166</b> |
| Chapter Specific Future Directions .....                               | 167        |
| General Future Directions .....  | 169        |
| <b>General Conclusion.....</b>   | <b>171</b> |
| <b>References.....</b>   | <b>172</b> |

## List of Tables

|   |     |
|---|-----|
| <b>Table 1: Aristotle's Categories.</b> Substance is further divided into two subcategories (Shields 2023). The entries of the table are also taken from Shields (2023).....  | 5   |
| <b>Table 2: Characteristics of previous RSA studies.</b> Number of participants, stimuli, and presentations used in previous RSA studies.....   | 35  |
| <b>Table 3: Participant demographics and study assignment.</b> Button-Press ICC refers to the consistency, assessed by a consistency-based intraclass correlation coefficient, of the participant's rating across different presentations of the stimuli. A minimum ICC greater than 0.2 was used as inclusion criteria for chapters 3, 4, and 5 of this dissertation. ....   | 40  |
| <b>Table 4: Stimuli for Study 1</b> .....   | 40  |
| <b>Table 5: Stimuli for Study 2</b> .....   | 40  |
| <b>Table 6: Body part stimuli groupings.</b> Body parts were divided into subcategories based on consensus agreement amongst study authors.....   | 92  |
| <b>Table 7: Peak coordinates corresponding to labels corresponding to the univariate analysis shown in Fig. 15.</b> Coordinates refer to MNI 152 template space and parcel names are taken from the surface-based atlas in Glasser et al (2016). For the univariate analysis, values correspond to the smallest bold difference between body parts and any other category. For the SVM analysis, values correspond to the average difference in cross-validated accuracy between category comparisons containing body parts, and category comparisons that do not contain body parts. IFG = inferior frontal gyrus, IPS = intraparietal sulcus, paraIns = parainsular cortex, pEF = premotor eye field, PGs = superior part of Von Economo and Koskinas area PG, pMFG = posterior middle frontal gyrus, pMTG = posterior middle temporal gyrus, preCS = precentral sulcus, preCun = precuneus, RSC = retrosplenial cortex, SMG = supramarginal gyrus..... | 98  |
| <b>Table 8: Pair-wise Pearson correlations between different unthresholded conjunction maps.</b> .....  | 102 |
| <b>Table 9: Correlation between feature ratings.</b> The table shows the Pearson correlation coefficients between each pair of the fourteen features included in the analysis. The color scale is centered on 0 .....   | 136 |
| <b>Table 10: Variance inflation factors for selected experiential features.</b> .....   | 137 |

## List of Figures

|   |    |
|---|----|
| <b>Figure 1: Illustration of CREA model. a)</b> An “example query screen” provided to the Amazon Mechanical Turk respondents. <b>b)</b> An illustration of a “mean attribute vector” for a concrete object noun category. For ease of interpretation, the “attributes are grouped and colour-coded by general domain and similarity” (Binder et. al., 2016). Both of the illustrations in this figure are taken directly from the text of Binder et al. (2016). ....  | 27 |
| <b>Figure 2: FMRI Mapping example.</b> This figure uses data from the HCP S1200 release. Motor homunculus adapted from Wikimedia commons and licensed under Creative Commons 4.0 ( <a href="https://commons.wikimedia.org/wiki/File:Motor_homunculus.svg">https://commons.wikimedia.org/wiki/File:Motor_homunculus.svg</a> ).....   | 29 |
| <b>Figure 3: Schematic of experimental task.</b> .....  | 42 |
| <b>Figure 4: Overview of RSA method. a.</b> The mask used to define the main region of interest in the present study. <b>b.</b> The model RDMs used in the analyses, with words sorted by semantic category. <b>c.</b> Group-averaged neural RDMs from the two data sets analyzed. ....   | 45 |
| <b>Figure 5: Reliability as a function of the number of stimulus presentations and the number of participants.</b> Each point, except for the ones furthest to the right on each plot, represents the average Cronbach's alpha calculated from 1,000 random resamples. The fitted line represents the Spearman-Brown (SB) Prophecy formula applied to the two-way random effects single measurement calculated using neural RDMs from all participants in each presentation combination. ....   | 51 |
| <b>Figure 6: Relation between mean RSA values and Cronbach's alpha.</b> Using RDMs derived from all 6 presentations, each point on the plot, except for the two points with the highest reliability (i.e., furthest to the right), are generated from the average of 1,000 resamples at a particular sample size. The horizontal axis represents Cronbach's alpha, and the vertical axis represents the average RSA correlation between an averaged set of neural RDMs and an RDM derived from a semantic model. The points with the highest reliability are based on the full set of participants and did not involve any resampling. The figure demonstrates that the average RSA correlation is proportional to the square-root of the reliability. This relation, along with the ability to estimate reliability at different sample sizes using the Spearman-Brown formula, allows for estimating group averaged RDM correlations to a model at different sample sizes. .... | 52 |
| <b>Figure 7: Spearman correlation with experiential model for each stimulus presentation.</b> Each point represents the correlation between a participant's neural RDM and an RDM derived from the experiential semantic model. The bars show the average RSA values, and the error bars show the 95% confidence intervals based on 1,000 bootstraps. The across-session and within-session main effects were highly significant, and an interaction effect was also significant in study 1. ....   | 53 |
| <b>Figure 8: Neural RDM correlations as a function of selected presentation combinations.</b> Each point represents the Spearman correlation between a single participant's neural RDM and the experiential model. The bars show the average RSA values, and the error bars show the 95% confidence intervals based on 1,000 bootstraps. The neural RDMs were derived using only the functional scans for the presentations noted on the horizontal axis. The data indicate that excluding the second presentations in sessions two and three only modestly affects the observed RSA correlation. ....  | 54 |

- Figure 9: Upper- and lower-bound noise ceiling estimates by number of participants.** Each point in this plot represents a noise ceiling estimate for a randomly sampled group of participants, with 1,000 estimates for each sample size. Also plotted is the expected asymptotic value of the upper- and lower-bound noise ceilings, as well as an analytic estimate of the lower-bound noise ceiling as a function of the number of participants extended out to a sample size of 60 participants. The figure shows a relative plateauing of the noise ceiling estimates at around  $N = 25$  in both studies..... 55
- Figure 10: Pairwise univariate contrasts.** 12 contrast maps were created. Each cell shows vertices where the average activation for the category listed in the corresponding row header was greater than the average activation listed for the corresponding column heading. All vertex values are average difference in bold signal normalized to percent signal change, and are significant by a Wilcoxon sign-rank test with FDR corrected  $p < .01$ ..... 72
- Figure 11: Multivariate clustering of data.** **Top:** A pseudo F-score was calculated for each participant's neural data based on the 4 categories of the stimuli. The pseudo F-score is a ratio of between cluster dispersion over within cluster dispersion, and a null distribution of F-scores was calculated by shuffling the category labels 10,000 times for each participant. The range of each violin plot indicates the largest and smallest F-value observed in any of the permutations. Pearson correlation distances were used to quantify distances between activation patterns. **Bottom:** A dimensionality reduction technique (UMAP) is applied to semantic-network derived activation-patterns, experiential ratings, and distributional representations. To visualize the neural data, neural RDMs were created for each participant using Pearson correlation distances. These neural RDMs were then averaged prior to visualizing the pairwise distances using UMAP. The two-dimensional projections highlight the intrinsic correlational structure of the data..... 73
- Figure 12: Vertex-wise encoding – Category Difference.** **Top:** Vertex-wise encoding models were used to predict category contrasts. Each box above the diagonal contains the Spearman correlation coefficient between the predicted difference map and the observed difference map. Significance was determined by permuting the training labels 10,000 times. Values below the diagonal are equivalent Z-scores for the Spearman correlations that were estimated through the null distribution. **Bottom:** Visualization of observed and predicted average category differences for two category comparisons. .... 77
- Figure 13: Vertex-wise encoding – Individual Words.** Encoding models were used to predict individual stimuli in the labelled left out category. All left out patterns were correlated with the predicted pattern, and the average percentile of the target map was calculated for each participant, and these percentiles were then averaged across participants. Significance was determined by shuffling the training data 1,000 times.... 77
- Figure 14: Variance partitioning.** Each 'dot' represents the unique variance attributed to a given feature space for a given participant. In all plots, error bars represent the 95% confidence interval of the mean calculated from 1,000 bootstraps. .... 78
- Figure 15: Mass univariate and multi-voxel pattern classifier analyses.** **Top:** The top portion of the figure shows where BOLD signal was greater for body part concepts than concepts from other categories based on standard mass univariate contrasts. All individual contrasts are significant at FDR corrected  $p < .01$  using a 1-tailed Wilcoxon signed-rank test. The conjunction map shows where body parts produced greater activation in all comparisons. Colors in the individual comparisons represent average

|  |     |
|--|-----|
| difference in beta-values. <b>Bottom:</b> The lower portion shows average classification accuracies for held out data in cross-validated searchlight SVM category classification. Classes were balanced, and 50% indicates chance accuracy. Shown in the center is the difference in average accuracy when one of the categories was body parts compared with when body parts are not a target category. Abbreviations are identified in the caption for table 2.....  | 96  |
| <b>Figure 16: Figure 16: Subgroup Univariate Contrasts.</b> Pairwise univariate contrasts were computed for subcategories of the body part category based on the groupings in supplementary table 1. Conjunctions were defined vertex-wise as the minimum value of the set [subcategory>animal , subcategory>artifact, subcategory>plant/food ]. Units are in percent BOLD signal change.....  | 99  |
| <b>Figure 17: Location contrasts.</b> Conjunction analysis with body part groups and other concrete object noun categories. All colored vertices are significant at FDR corrected $p < .01$ and a minimum cluster size of 40 mm <sup>2</sup> was used to aid in visualization.....   | 100 |
| <b>Figure 18: ICC and RSA comparisons.</b> ICC and RSA maps were generated for each of four categories of nouns. Results for the body part category are shown in the top row, and contrasts between body part and other categories are shown in subsequent rows. The bottom row shows conjunction maps of the categorical comparisons. Colors in the ICC conjunction map represent the difference between the body part ICC value and the highest upper limit of the confidence interval for any of the other three categories. The RSA conjunction maps were generated by testing the mean of paired differences against zero using a Wilcoxon signed-rank test. The conjunction maps show vertices where body part RSA values were larger than for any other category using an FDR corrected $p < 0.01$ threshold. The conjunction map color shows the mean paired difference between body part RSA values and whichever other category had the highest RSA value for that vertex..... | 104 |
| <b>Figure 19: Vertex-wise encoding results.</b> Shown on the left is the difference between the average of all body part concepts and average of all other concrete concepts. On the right is the result of a vertex-wise encoding model trained on 150 concrete concepts to predict beta values for the 50 body part concepts. Shown is the average difference between the average of the 50 predicted values and the 150 observed values along with vertices that were significantly larger for body parts at FDR corrected $p < .01$ using a one-tailed Wilcoxon signed rank test.....  | 105 |
| <b>Figure 20: HCP body part activations.</b> <b>a)</b> Select HCP parcels are labelled. The coloring corresponds to that used in the original report by Glasser et al. (2016). <b>b)</b> Body part word > other category univariate results overlaid with HCP parcel boundaries. <b>c)</b> Body part pictures > other category during a working memory task, univariate results from the HCP dataset. <b>d)</b> Overlap of the maps generated by body part words relative to other categories in the current study and body part images relative to other categories in the HCP dataset. The current results are displayed in Cohen's D units, the same units as the HCP data, and thresholded so that only values greater than 0.3 are displayed for both maps. <b>e)</b> Resting-state fMRI connectivity map using seed-based correlation, with a seed placed in left PHT applied to data from 1003 participants in the HCP S1200 release. ....  | 108 |
| <b>Figure 21: Experiential feature overlap map.</b> This figure illustrates the number of overlapping experiential features at each vertex across the vertex. All colored vertices are significant at FDR corrected $p < .01$ .....  | 138 |

---

|   |     |
|---|-----|
| <b>Figure 22: Complexity feature map .....</b>  | 139 |
| <b>Figure 23: Color feature map .....</b>   | 140 |
| <b>Figure 24: Shape feature map.....</b>  | 141 |
| <b>Figure 25: Scene feature map .....</b>   | 142 |
| <b>Figure 26: Sound feature map .....</b>   | 143 |
| <b>Figure 27: Taste feature map.....</b>  | 144 |
| <b>Figure 28: Upper Limb feature map.....</b>   | 145 |
| <b>Figure 29: Head feature map .....</b>  | 146 |
| <b>Figure 30: Number feature map.....</b>   | 147 |
| <b>Figure 31: Social feature map.....</b>   | 148 |
| <b>Figure 32: Valence feature map.....</b>  | 149 |
| <b>Figure 33: Follow-up univariate feature maps.....</b>  | 150 |
| <b>Figure 34: Comparison of feature overlaps to prior work. a)</b> Shows the number of features that were significant at each vertex (reproduction of Figure 21). <b>b)</b> A reproduction of Figure 5 from Fernandino et. al. (2016). The colored areas correspond to “areas associated with all 5 semantic attributes in the conjunction analysis” (Fernandino 2016). The figure is licensed under Oxford University Press and permission to reproduce the figure here was provided by Copyright Clearance Center. <b>c)</b> A visualization of the first principal gradient from Margulies et. al. (2016). <b>d)</b> A visualization of the general semantic mask reported in Binder et. al. (2009). All colored vertices are significant at $p < .01$ ..... | 151 |
| <b>Figure 35: Select feature overlays.</b> Several feature maps are overlaid that demonstrate either conceptually similar or distinct sets of features.....   | 160 |

## Acknowledgements

I always enjoy reading the acknowledgment section for any large piece of work.

Sometimes, I think, the question occurs: "will I be mentioned?" To those of you in this group, yes, I likely thought about you when writing this section.

There are so many people to thank for their friendship and support. Many research mentors and teachers that have instilled in me a love for knowledge and critical thinking. I'm grateful to Dr. Bill Bauer for accepting me into the BioXFEL program as an undergraduate and to Dr. Peter Schwander for teaching me about fascinating methods like diffusion mapping.

I've also had the support of many friends along the way. My roommate Andrew, friends across Milwaukee, neighbors, medical school friends, and my MSTP class. I am thankful for all of them.

Perhaps the most direct reason I've had the great good fortune of being able to live in Milwaukee, write a dissertation, and spend most days in the past seven years learning something new is because of the MSTP at MCW. I'm incredibly indebted to them for their support and belief in me as a scientist. It is not lightly that I recognize the treasure the leadership at MCW has entrusted me with.

I'm not convinced that I would have found a deep interest in cognitive neuroscience if not for Dr. Jeffrey R. Binder accepting me into his lab and providing me the environment and support to learn these past several years. Included in the environment, I'm thankful for conversations with Tony, Dr. Anderson, Anna, and members of the LiL. I particularly want to thank Dr. Fernandino and Dr. Binder for teaching me that a paper is not ready until it is ready, for all of their careful edits, and

---

most importantly, for their teaching and sharing of insight. I'm also thankful for my committee members and their help.

I'm saddened to think of Dr. Lisa Conant not being at my dissertation defense. She had an amazing breadth of knowledge that she only ever humbly shared, a passion for research, and a kind heart.

I have also been blessed to have the support and care of Fiona, who is a reminder to me of how lucky we can sometimes be.

I'm grateful to my family. My grandparents, Dorothy, John, Bettina, and Michael. My dad for teaching me more than he could know and for all our conversations. My mom for her constant love, and Joe. My brother Michael, and my sister Maria.

Lastly, I'm thankful to God for all the good he has placed in my life, for each day I'm provided with, and for anything worthwhile that I might accomplish.

# Chapter 1: Introduction

## Purpose of the Study

This project was undertaken to thoroughly test how category-related organization of the cortex can be accounted for using a model of semantic meaning developed by Binder and colleagues (2016). To accomplish this, I will first cover general background and then apply the semantic model to open problems in the literature and investigate the neural substrates of the information represented in the model.

## Scientific Relevance

Understanding how the cognitive abilities of the mind arise from neural substrates has been a topic of epistemic interest since the writings of the earliest philosophers. While the topic of how word meanings are stored and accessed in the brain is actively debated, it is both implicitly and explicitly informed by centuries, and arguably millennia, of debate. For example, even the implicit notion that semantic cognition can be understood through scientific reductionism can be readily traced back to the empiricists and earlier. In modern times, the advent of functional neuroimaging techniques has led to a resurgence in the study of the neural substrate of cognitive function, and debates around the nature of semantic representation have emerged with new fervor. Even more recently, the rapid growth, availability, and adoption of large language models (LLMs) has led to new interest in questions around how the meaning of concepts can be represented. The growing interest in the neural implementation of language is reflected in the large number of publications related to the topic.

Our motivation for explicitly testing the claims of a prominent model of concept representation is linked to two purposes: First, models must provide parsimonious accounts of seemingly disparate phenomena. Scientific advancement depends on consensus building and requires explicit testing of predictions to advance discussions. Second, understanding the neurobiological basis of cognitive functions can lead to improved therapies for pathologies that disrupt normal brain functioning and technology that can improve health even in the absence of known pathology.

### Clinical Relevance

The *raison d'être* of the National Institutes of Health (NIH), the primary funding source for the results presented in this dissertation, is to support research that improves medical treatment. Language disorders are a common feature of many neurological diseases. For example, aphasia, defined as “the loss or impairment of language caused by brain damage” (Berthier 2005), is estimated to be present in 21-38% of acute stroke patients (Berthier 2005). While aphasia is common, there is still a considerable amount about the disorder that is unknown, and there is active debate over the best therapy for regaining language function following insult to the brain. The primary topic of this thesis relates to understanding where and how the brain stores and accesses the meanings of words. A better understanding of this topic promises to lead to better prognostication and improved therapies for recovery following damage to cortex.

Although the field of functional neuroimaging has had prolific expansion since its origin, much work remains in explicitly testing the predictions made by different theories. While many theories have been put forward about how semantics might be operationalized by the brain, little has been done to translate general theory development

into clinically informative tools. With the notable exception of pre-surgical language mapping, fMRI has had surprisingly limited clinical impact given its expansive research utility. With semantic impairment implicated in increasingly prevalent neurological disorders such as epilepsy, stroke, and dementia, it is critical to build consensus around the roles of regions of the cortex so that current tools can be effectively leveraged, and new tools developed to improve care to patients suffering from language disorders.

## Historical Background

It is by them (our perceptions) also that we think.

(Epicurus in Cicero's De Finibus<sup>1</sup>)

### Ancient Beginnings

It is difficult to frame the works and writings of the earliest philosophers into today's lexicon. Moreover, the field of philosophy uses several interrelated terms to discuss the ideas surrounding what is today termed semantic memory and what is colloquially meant by the term 'concept'. For example, 'forms' in the platonic sense can be seen as a precursor to what we mean when we talk about concepts, but discussion of 'forms' by philosophers is often couched separately from 'theories of meaning', which has been broadly considered under the purview of logic in the Western analytical framework.

Given the limited scope of this background section, and to limit the degree of misrepresentation when summarizing prior work, discussion of the pre-modern origins of

---

<sup>1</sup> This quote is lifted from Moseley R, Kiefer M, Pulvermüller F. 2016. Grounding and embodiment of concepts and meaning: A neurobiological perspective. Foundations of embodied cognition: Perceptual and emotional embodiment. New York, NY, US: Routledge/Taylor & Francis Group. p. 93-113., but the original reference is from: Handbook of the History of Philosophy by Dr. Albert Stöckl Part I. Pre-Scholastic Philosophy, translated by T.A. Finlay, S.J., M.A. (1887). This can be accessed online at: <https://www3.nd.edu/~maritain/jmc/etext/hhp44.htm>

semantic memory is limited to an extremely cursory overview of widely known Western philosophy.

Pre-Socratic philosophers recognized that the appearances of objects change and posed the question of what a thing ‘really’ is. For example, a chair could change colors, yet it remains a chair. The proposed answer to ‘what makes a thing’ was substance, and individual items that are members of a category are called particulars, which may have idiosyncratic differences. It is from this starting point that the writings of the Socratic philosophers can be considered. According to Richard Kraut (2022) “Plato (429?-347 B.C.E.) is, by any reckoning, one of the most dazzling writers in the Western literary tradition and [...] one of the most influential authors [...] in all of the Western literary tradition.” He laid the foundation of many areas of study, and arguably, conceptual representation is not an exception. Plato’s basic ideas surrounding what things ‘are’ is known by several names (e.g., Theory of forms, Platonic idealism, or Platonic realism), and is a theory of metaphysics that broadly posits that all objects and matter in the physical world are merely ‘imitations’. According to Kraut (2022), “nearly every major work of Plato is, in some way, devoted to or dependent on this distinction.” Although Plato’s theory of forms would not fit well with any contemporary theory of semantics, it is perhaps a precursor to the ‘symbolic’ viewpoint where concepts are themselves atomic. That is, there is no substructure, and concepts are in some way innately graspable by our minds. While there is extensive exegesis on forms, given the limited nature of this introduction, little else is said here.

Following Plato, Aristotle extended and reformed the thought on forms. Particularly motivated by how universals change, he proposed a modified account. In his

| Category  | Example                           |
|-----------|-----------------------------------|
| Substance | man, horse                        |
| Quantity  | four-foot, five-foot              |
| Quality   | white, grammatical                |
| Relation  | double, half                      |
| Place     | in the Lyceum, in the marketplace |
| Date      | yesterday, last year              |
| Posture   | is lying, is sitting              |
| State     | has shoes on, has armor on        |
| Action    | cutting, burning,                 |
| Passion   | being cut, being burned           |

**Table 1: Aristotle's Categories.** Substance is further divided into two subcategories (Shields 2023). The entries of the table are also taken from Shields (2023).

work *Categories*, he placed every object of human understanding under one of 10 categories (**Table 1**). This theory can be related to contemporary theories that view word meaning through a taxonomic (i.e., hierarchical classification) lens where basic categories emerge because they carry the most information (i.e., they “possess the highest category cue validity, and are, thus, the most differentiated from one another” (Rosch et al. 1976).

Although not explicitly in this grouping, another view that is typically found in survey articles about the history of concepts is what is called ‘the classical view’. Under this view, lexical concepts (i.e., concepts expressed by a single word) are composed of ‘simpler’ concepts, where something is a particular concept if it meets the required conditions. The canonical example that is often provided is the term ‘bachelor’ which is defined with reference to being 1) a man that is 2) unmarried. In terms of what this means for concept use, it is the implication that when we are determining whether something is a

'bachelor' we are testing the item against the necessary conditions (Johnston and Leslie 2019). This 'classical view' would come to be generally rejected due to the extensive work on natural categories and typicality effects by Eleanor Rosch and colleagues (Rosch et al. 1976)

Contemporary philosophers group the view held by Plato and Aristotle under the term 'realism' because of their appeal to forms or universal ideals. That is, realists hold that universal ideals, which transcend particulars, are in some sense 'real'. This view was challenged in the Middle Ages by the view contemporary philosophers call Nominalism. Nominalists hold that ideals/universals are simply names, that is, human creations (e.g., there is no such thing as 'blue', only blue things). Nominalism represents a shift away from 'realism', the idea that transcendental/universal forms exist, and becomes a starting point for the views generally endorsed by subsequent empiricists.

### Enlightenment

Following the nominalists, empiricism is the dominant school of philosophy that began in the 15<sup>th</sup> century and forms the foundation of much modern thought. British empiricists such as John Locke (1632-1704), George Berkeley (1685-1753), and David Hume (1711-1776), generally viewed the representations that underlie cognition as imagistic (Barsalou 1999). Quoting Margolis and Laurence (2023):

"Empiricists have argued that all concepts derive from sensations. Concepts were understood to be formed from copies of sensory representations and assembled in accordance with a set of general-purpose learning rules [...] On this view, the content of any concept must be analyzable in terms of its perceptual basis."

Another way of phrasing this idea is that the concept ‘dog’ might be akin to an average of the experiences one has had with dogs. Although empiricists have a range of views within their own circles, there is a general value placed on what is ‘sensed’ and a repudiation of realism (i.e., that all things have a ‘form’ in the platonic sense). The empiricist viewpoint is strongly represented in contemporary neuroscience and implicit across much of the natural sciences.

### Modern Times

Up to this point, the discussed theories have focused on philosophical grounds for different theories. However, work in the nineteenth century would shift to focus on a ‘scientific’ understanding of the brain. Little regarding cerebral localization was known at the beginning of the nineteenth century. Excepting that the brain was long held to be the “seat of consciousness” or rationality (e.g., Alcmaeon of Croton<sup>2</sup>, Galen<sup>3</sup>), the dominant idea in the early 19<sup>th</sup> century was that the brain was, broadly, one large undifferentiated organ. One prominent neurologist of the time, Flourens (1794 - 1867), who is “one of the pioneers in the usage of experimental methods in neuroanatomy” (Yildirim and Sarikcioglu 2007), is famous for his experimental method of carrying out “ablation and stimulation methods and many experimental investigations [...] especially on rabbits and pigeons” (Yildirim and Sarikcioglu 2007). The purpose of Flourens’ experiments was to investigate localizationism (i.e., to test whether different parts of the brain had different functions). Although “he was the first to show experimentally that different parts of the nervous system have different functions” (Tizard 1959), he was unable to find specific

---

<sup>2</sup> “He was the first to identify the brain as the seat of understanding” - <https://plato.stanford.edu/entries/alcmaeon/>

<sup>3</sup> “These are three: the rational or leading-part, located, Galen argues, in the brain” - <https://plato.stanford.edu/entries/galen/>

regions for memory and cognition. This lead him to put forward a “theory of equipotentiality within the hemispheres” (Tizard 1959) believing that memory and cognition are represented in a diffuse manner across the brain (Kanwisher 2010).

Flourens argued that “all sensations, all perceptions, and all volition occupy concurrently the same seat in [the brain]. The faculty of sensation, perception, and volition is then essentially one faculty” (Finger 2001)<sup>4</sup>. He taught that all parts of the brain served that same function, and that if one part was damaged, another could take over the function (Nielsen 1962). This view, however, would see radical change and would no longer be in vogue as the 19<sup>th</sup> century progressed<sup>5</sup>.

A common point that many consider the start of modern aphasiology (i.e., the study of language disorders) are the now famous two cases reported by Paul Broca (1824-1880) in 1861. This report claimed that the faculty of language was located at the base of “the third left frontal convolution” (Berker et al. 1986). From here, the work of neurologists such as Wernicke, Lichtheim, and others led to information processing models of brain functioning (Heilman 2006). Although Freud is generally more remembered for his work on psychoanalysis, his early work on aphasia appears to have influenced his later writings (Forward to English edition; Freud 1891). While the latter half of the 19<sup>th</sup> century has a significant amount of work discussing the different association and language centers of the brain, a history of aphasiology is outside the scope of this introduction. What is important is that from the mid-19<sup>th</sup> century and

---

<sup>4</sup> This quote was originally found and lifted from the supplementary material of Kanwisher (2010)

<sup>5</sup> As Wernicke wrote: “The Flourens concept of the intellect as a single unity, claiming equivalence of all brain areas, has long proved untenable” (Wernicke. (1874). *Der Aphasische Symptomencomplex.*). For a longer review and discussion see: Theories of Brain Localization from Flourens to Lashley by Barbara Tizard (1959)

onwards, language localization was studied inside a broader discussion about, according to Freud, “the idea of ‘localization’, i.e., of the restriction of nervous functions to anatomically definable areas, which pervades the whole of recent neuropathology” (Freud 1891). While the early 20<sup>th</sup> century would see competing models of language representation, deficits restricted to a single category of objects would not be reported until later.

Prior to covering the first reported semantic deficit along categorical lines in the 1930’s, there were reports of a different phenomenon called agnosia. While there are many specific types of agnosia, they all refer to a deficit in recognition that cannot be attributed to low-level visual or sensory deficits (Coslett 2018). The term ‘agnosia’ is generally attributed to Freud in “his discussion of aphasia and related disorders” (Biran and Coslett 2003). A more in-depth description and theoretical framework of agnosia was offered in a now famous account by Lissauer, another German neurologist, who described “associative” agnosia where patients have “an inability to link an adequate percept to stored knowledge indicating its name, function, size, and so forth” (Biran and Coslett 2003). Quoting a translation of Lissauer’s report (Lissauer and Jackson 1988): “In fact it became immediately obvious that this man, who could give a good account of everything around him, was quite incapable of visually recognizing the most common objects, although he could recognize everything by touch or hearing.” There are many types of agnosia such as agnosia for faces (prosopagnosia), agnosia for words, and agnosia for colors, amongst others. While these examples are primarily visual, agnosias are not limited to this modality as “visual agnosia, tactile agnosia and auditory agnosia have all been reported as relatively isolated disabilities” (Warrington 1974). The

commonality across different types of agnosia is that they represent a modality-specific recognition deficit.

### Evidence of Category Related Semantic Deficits (CRSDs)

Prior to beginning discussion of CRSDs, one term that could stand to benefit from clarification is what is meant by the term ‘semantic’. Attempting to rigorously define the term ‘semantic’ presents a formidable challenge (for discussion, see Reilly et al. 2023). However, some working definition ought to be provided. First, the phrase ‘semantic representation’ will be repeatedly used throughout this dissertation, normally in reference to some region of the brain. The reason for this is that the phrase admits at least one straight-forward operationalized definition: a region is said to contain a semantic representation if the activation of the region is modulated by a semantic variable (e.g., category membership or other semantic feature content such as imageability). It is important to stress that the modulation cannot be otherwise accountable by visual or ortho-phonological properties of the stimuli. While this definition might at first appear circular, it hopefully does *some* work on advancing the conversation by touching on a relevant point: things like category membership and attributes (e.g., size, color, texture, as well as abstract traits such as amicability, intelligence, or efficiency) have long been considered ‘semantic variables’. Admittedly, the crux of the definition still depends on the difficult task of providing a sense of what can be admitted as semantic variables. One definition for this term is put forward by Reilly and colleagues:

“Any variable used for differentiating exemplars (e.g., axe vs. spoon) across any given aspect of meaning (e.g., capacity for inflicting harm). Semantic dimensions

are often but not always continuous (e.g., pleasantness vs. animacy).” (Reilly et al. 2023)

Although the above definition might still have some words that could use further clarification, it has the benefit of being endorsed by 95% of the researchers in a “multidisciplinary workgroup (N=53)” (Reilly et al. 2023), and establishes a general gestalt of how ‘semantic’ should be understood throughout this dissertation.

The earliest descriptions of category related semantic deficits go back to case reports by Nielson in the late 1930s. Quoting Nielsen’s report of patient C.H.C:

“A bottle of milk and his drinking glass were unknown objects. He failed to recognize an airplane by sight or by sound. Yet with all this visual disability relative to inanimate objects he recognized and revisualized all living things. He knew his most intimate friend, his doctor, and the six nurses in attendance, even recognizing them by their names as well as by sight alone. When he could not recall owning his estate he nevertheless knew his housekeeper [...]. When it became so evident that he was drawing a sharp line between animate and inanimate objects, both for recognition and revisualization, a flower was presented and he was asked what is was. He immediately names it (a daffodil) correctly.” (Nielsen 1962)<sup>6</sup>

Nielson reports two other cases demonstrating loss along categorical lines and writes again:

---

<sup>6</sup> This quote comes from the Second Revised Edition version of the textbook. However, the first edition that contained this and other cases was first published in 1936.

“these three cases then suggest that animate objects may be recognized and revisualized by the left occipital lobe while the same functions for inanimate objects proceed through functional activity of the right lobe” (Nielsen 1962).

This observation keenly suggests that semantic information may be spatially localized along categorical lines. More testing of this hypothesis, however, would not occur until a seminal work by Goodglass et al. (1966). In their work “Specific semantic word categories in aphasia” (1966), the authors state that some clinically encountered conditions, amongst others, include “loss of object naming without impairment of letter naming or of number naming” and “impairment of color naming, without other anomic difficulties”. They state that:

“Up to the present time there has been no systematic collection of data on the problem of differential capacity for word-finding and understanding of various semantic categories.”

The authors study a few select categories and in summary state:

“It is concluded that the pattern of differences in naming and auditory discrimination among words of various semantic categories varies predictably with the major clinical types of aphasia.”

Although the results of Goodglass et al. (1966) are interesting, the study of CRSDs would not gain much attention until a series of case reports in the 1980’s.

Much of the modern study of the organization of the semantic system started with the work of Warrington and Shallice and their study of CRSDs (Warrington and Shallice 1984). This seminal work was a case series that examined 4 patients who demonstrated

selective impairment of “animals, plants, and foods as compared with inanimate objects” (McCarthy and Warrington 2015).

Following Warrington’s case studies, many other dissociations have since been reported such as for fruits and vegetables and body parts. Further, many fMRI studies have found regions with category-dependent activation suggesting category localization (For review see: Mahon and Caramazza 2009). However, a parsimonious explanation of how and why this occurs remains to be had. The following section outlines prominent proposals that have been put forward.

### Theoretical Accounts of CRSDs

It is likely that every researcher who studies the neurobiology of language has their own thoughts about how concept representation is implemented in the brain. Although the following sections will appear to make sharp boundaries between groups of theories, and present groupings as homogeneous, there is variability within the walls used to separate viewpoints. This cursory review intends only to cover the structure that might remain when viewing the different proposals from afar and grouping modes of thinking together. It starts by laying out several (if not historically) dominant modes of thinking, and then focuses on the specific model of semantics tested throughout this dissertation.

Following a distinction put forward by McCarthy and Warrington (2015), one can make a division between at least two large classes of theories: those derivative of domain specificity, and those that are derivative of the sensory-functional hypothesis (SFH). However, this distinction is not necessarily canonical as, for example, Caramazza and Mahon (2009) divide theories into two different classes, described later. Yet another

classification can be seen in the review by Frisby et. al. (2023). While researchers have grouped theories together using different criteria, there are some general ‘camps’ that are readily apparent in the neurobiology of language literature.

### Sensory Functional Hypothesis

One large group of theories are those derivative of the SFH that is described in the seminal Warrington and Shallice (1984) paper, which broadly described an “animate-inanimate” distinction. Although not discussed in detail here, their proposed explanation draws upon some prior literature related to optic aphasia (i.e., the inability to name visually presented objects, but with preserved naming through other modalities)<sup>7</sup>, which was taken by some as evidence of a close connection between perceptual systems and semantic systems. The explanation of CRSDs put forward in the paper by Warrington and Shallice (1984) can be somewhat summarized by saying that animate (i.e., biological) objects depend more on visual information, and tools/artifacts derive their meaning primarily from how a person interacts with the object, which is to say the meaning is defined from an object’s function. Quoting from their paper:

“Inanimate objects unlike, say, most animals and plants, have clearly defined functions. [...] We would suggest that identification of an inanimate object crucially depends on determination of its functional significance, but that this is irrelevant for identification of living things.”

Under the SFH, categories of concepts that depend on visual information would be expected to be impaired when knowledge of animals is impaired. This view is in line

---

<sup>7</sup> Optic aphasia differs from visual agnosia in that visual agnosia is a problem of *recognition*, not naming. In optic aphasia, patients recognize the visual objects, but they are unable to name the object.

with an implicit hypothesis regarding the semantic role that primary sensory and motor cortices, along with their adjacent association cortices, play in representing concepts.

Namely, what is assumed is that there is no central “language region” and that conceptual information is broadly distributed across primary and association cortices distributed across the cortex.

### *Grounded Cognition*

The SFH is a prototypical form of what is now commonly brought under the umbrella of “embodied” or “grounded” cognition. Given the central role this term has in contemporary discussion, it deserves some elaboration. Two important references that can help build an understanding of what is meant by the term “grounding” are the highly impactful papers of Arthur Glenberg (What memory is for; 1997) and Lawrence Barsalou (Perceptual symbol systems; 1999). Both papers draw significant attention to the symbol grounding problem (Harnad 1990). According to Harnad, the problem is:

“How can the semantic interpretation of a formal symbol system be made *intrinsic* to the system, rather than just parasitic on the meanings in our heads?  
[...] The problem is analogous to trying to learn Chinese from a Chinese/Chinese dictionary alone.”

Both Glenberg and Barsalou argue that the core problem with newer theories is that they think of concepts as amodal symbols that are *arbitrarily* associated with their real-word referents. That is, according to Barsalou, the “similarities between amodal symbols are not related systematically to similarities between their perceptual states”. In so much as amodal symbols might become associated with perceptual states in memory, Barsalou argues that the amodal symbol then become redundant and unnecessary.

One of Barsalou's main theses is that the study of cognition and perception should not be separate and that the two don't represent independent 'modular' systems in the brain. Instead, he argues "cognition is inherently perceptual, sharing systems with perception at both the cognitive and neural levels." His central idea is that memory and language depend upon activating "perceptual symbols" in the sensory-motor systems of the brain, which are viewed as generalizations of sensory-motor experiences. Barsalou's viewpoint helped to herald a shift in neuroscience from the modular frameworks back to what is called embodied or grounded cognition.

As a point of clarification, although 'grounded', 'embodied', and 'experiential' are sometimes used somewhat interchangeably, for consistency throughout this dissertation, I will primarily adopt and use 'experiential'. Some authors make distinctions between these theories, such as the following one given by Reilly and colleagues (2023):

"In contrast [to embodied cognition], grounded cognition theories are broader and often incorporate internal perceptual modalities, such as introspection, emotion, and mentalizing (Kiefer and Harpantner, 2020; Vigliocco et al., 2014). Moreover, many grounded cognition theories do not restrict the conceptual system to modality-specific areas but allow for the additional involvement of cross-modal brain regions that integrate modality-specific features into more abstract conceptual representations."

A general property of grounded theories is that they involve many regions of the cortex that are strongly implicated in other perceptual processes. Strong forms of the grounded cognition hypothesis require the *same* structures to be involved in perception and recognition, weaker forms include primary sensory adjacent regions (for discussion, see

Rugg and Thompson-Schill 2013). Throughout this dissertation, ‘experiential’ is used to encompass not only what is meant by grounded, but more broadly “any information encoded in the brain that originates from experience” (Fernandino and Binder 2024).

Although the SFH represents a common ancestor for many current theories, contemporary derivatives of the SFH have changed the details and emphasis in response to new discoveries. For example, one shift, is a broadening of SFH from a primary focus on vision and function to vision, sound, tactile, motor and other kinds of experience (e.g., Allport 1985). While this idea was not novel, as empiricists and many neurologists (e.g., Wernicke, Freud, and others) describe sensory ‘images’ from multiple modalities, Allport helped resurrect the idea of concepts as auto-associated patterns distributed across many sensory and motor domains in the contemporary period. Another proposal that is largely in line with the grounded cognition viewpoint is the one articulated in a seminal paper by Antonio Damasio (1989). In this work, Damasio outlines a framework where neuron ensembles located in primary and first-order sensory association cortices lead to local convergence zones, followed by higher-order non-local convergence zones<sup>8</sup>. What is critical about this proposal, as well as at odds with other models that will be mentioned later, is that it:

“rejects a single anatomical site for the integration of memory and motor processes and a single store for the meaning of entities of events. Meaning is

---

<sup>8</sup> This view is again reminiscent of the description provided by Wernicke: “The cerebral surface is a mosaic of such primary elements whose properties are determined by their anatomical connections to the body-periphery. All processes which exceed these primary functions (such as synthesis of various perceptions into single concepts [...] are dependent upon the fiber bundles connecting different areas of the cortex” (Originally published as: Wernicke, 1874. Der Aphasie-Symptomkomplex. Translation from Eggert 2019)

reached by time-locked multiregional retroactivation [sic] of widespread fragment records”

Since its proposal, the idea of convergence and divergence zones has gained significant traction in neuroscience. However, like any proposal, it is not without its detractors.

The SFH would face criticism as other dissociations, often more specific than animate-inanimate, were found. In response, grounded models of cognition proposed more fine-grained distinctions in types of experiential knowledge. There was also growing uncertainty about whether the SFH entailed separate modality specific semantic systems and, if it did, the extent to which these multiple semantic systems were independent. As imaging systems improved, another finding that seemed to require explanation was the relatively ‘strong selectivity’ some regions appeared to show for certain types of information (e.g., a circumscribed region that appears to specifically process face information). More studies were needed to determine exactly what and where the hierarchical convergence zones are along with a better understanding of the spatial extent of the semantic system.

### Domain Specificity

In response to limitations of the SFH, competing explanations gained more attention in the early 1990’s. One model is known as the organized unitary content hypothesis or OUCH model, which challenged the idea that some semantic content was specific to a modal system (Caramazza et al. 1990). At the same time, there was a resurgent interest in modular viewpoints and the proposal of innate ‘circuits’ in the brain. One family of theories, often grouped under what is called the domain specific hypothesis or DSH, claims that some innate domains of knowledge (e.g., plants, food, conspecifics) have

regions of the brain dedicated to their specific processing. These theories propose that specific domains of knowledge are innate because they are under an evolutionary pressure leading to their spatial confinement. One quote that captures the essence of the DSH can be found in a review by Caramazza and Mahon (Mahon and Caramazza 2011):

“According to the domain-specific hypothesis, there are innately dedicated neural circuits for the efficient processing of a limited number of evolutionarily motivated domains of knowledge.”

Domain specific theories were not at first widely accepted, as even in Warrington's initial description of the category impairment, the deficits did not fall precisely along the broad animate-inanimate distinction. For example, patient J.B.R. had deficits for precious stones along with impairment for animals and plants and food. In time, however, domain theories would gain popularity as problems with the SFH grew. For example, there were concerns that the initial category distinctions might not be the result of semantic categories but might simply reflect familiarity and visual complexity information. That is, normative human ratings demonstrated that inanimate objects are rated as more familiar than animals. Several studies found that semantic category effects disappeared when the proper nuisance variables were controlled for (Funnell and Sheridan 1992; Stewart et al. 1992). This concern eventually abated as double dissociations were found on the same stimuli (Hillis and Caramazza 1991), and reliable differences between semantic categories were still found even with nuisance variables rigorously controlled for (Farah 1996; Gainotti 1996).

Another reason domain centered theories gained prominence centers on a parallel discussion that had been occurring in other areas of neuroscience. To provide some

context for the state of affairs in cognitive neuroscience, one article states: “In the 1960s and 1970s, cognitive scientists almost universally modeled mental activity as rule-governed symbol manipulation” (Rescorla 2023). It was in this context that seminal work by Jerry Fodor revitalized what is known as the “language of thought hypothesis” and the idea of a language module in the brain (Fodor 1975). While not elaborated on here, Fodor describes a view of the brain that consists of “informationally isolated” modules. This view has natural overlap with the idea of domain specific regions of cortex postulated by the DSH.

While so far this introduction has been divided along SFH and DSH criteria, as mentioned, there are other ways to group theories. Caramazza and Mahon divide theories based on a different criterion (Caramazza and Mahon 2006). Quoting from their relevant paper:

“One group of theories, based on the neural structure principle, assumes that the organization of conceptual knowledge is governed by representational constraints imposed by the brain itself. A second group of theories, based in the correlated structure principle, assume that the organization of knowledge in the brain is a reflection of the statistical co-occurrence of object properties in the world.”

In this grouping, Caramazza places both SFH and domain-specific hypotheses into the former camp and feature based theories in the latter camp. In my view, this distinction is similar to the “nature” versus “nurture” distinction in that the neural structure principle focuses on innate characteristics of the brain, and the correlated structure principle emphasizes natural categories as something that simply exist in our experience of the world. According to Caramazza and Mahon, theories that fall under the correlated

structure principle differ in the types of feature properties concomitant with the model. The explanation of how category organization arises from the correlated structure principle is that:

“First, conceptual features corresponding to object properties that often co-occur will be stored close together in semantic space; and second, focal brain damage can give rise to category-specific semantic deficits either because the conceptual knowledge corresponding to objects with similar properties is stored in adjacent neural areas, or because damage to a given property will propagate damage to highly correlated properties.” (Caramazza and Mahon 2006)

This has at times been described as saying that a ‘lumpiness’ is induced in ‘semantic space’ because the underlying features have an underlying correlational structure.

### Explanatory ability

Having covered the general types of models, we can now review some of the historical evidence in agreement or at variance with the models. One early and important development in support of the SFH was a computational implementation of its premise by Farah and McClelland (1991). In their work they demonstrated how damage to a visual domain of knowledge “would have greater impact on living things” (McCarthy and Warrington 2015). However, while promising, several later studies would test the hypothesis that patients’ “verbal knowledge of sensory or functional properties” would reflect impairments with concepts in corresponding semantic categories (McCarthy and Warrington 2015). On this prediction by the SFH, the evidence was mixed; some patients had deficits in knowledge of animals without corresponding impairment of other types of

visual semantic knowledge (Coltheart et al. 1998). For instance, Miceli and colleagues report “two brain-damaged subjects who exhibit [...] loss of object color knowledge, but spared color perception naming” (Miceli et al. 2001). The authors go on to say that although patient I.O.C. is “impaired in accessing object color knowledge, [they do] not exhibit selective difficulties in processing fruits and vegetables, or animals” (Miceli et al. 2001). According to McCarthy and Warrington (2015), this then led “to the suggestion that attribute knowledge was a potentially independent level of semantic representation” where “attribute knowledge” refers to reportable properties like ‘has taste’.

The case against the SFH also grew as other category dissociations were found. For example, while the dissociation between animals and food concepts against non-living items was a relatively robust CRSD (Renzi and Lucchelli 1994), exceptions were found where some patients were impaired for animals but not food (Hart and Gordon 1992) or who were impaired on the category of food but not animals (Hillis and Caramazza 1991; Samson and Pillon 2010). Similar problems of CRSDs not always being paired together (for example difficulty with musical instruments only sometimes being associated with deficits in animal knowledge), contributed to the development of other theories. The variability in CRSDs was problematic for the SFH in its simplest form, as one would expect more consistency in the pattern of impairment across categories.

Although not cleanly a derivative of either domain specificity or the SFH, the early 2000s saw the development and growth in popularity of the ‘hub-and-spoke model’. This model is arguably agnostic on the SFH and DSH criteria; the hub-and-spoke model is a theory about the general *neural* structure of the semantic system. The general premise

of the hub and spoke model is that modality-specific ‘spokes’ (e.g., vision, smell, etc.) converge on a central semantic region, putatively placed in the anterior temporal lobe (ATL). One primary reason for considering the ATL the primary ‘hub’ comes from a view argued for in a 1990 paper by Caramazza and colleagues. In this paper, Caramazza et al. (1990) refer to the content of representations and hold to the view that the content of conceptual representation is amodal. This view, combined with evidence from patients with semantic dementia (SD) who show ATL damage, has led to a proposed amodal semantic hub in the ATL. This overall perspective is again articulated in a 2007 review by Patterson et al. (2007).

In some sense, this hub-and-spoke model is a blend of embodied representations that are represented in the ‘spokes’ with amodal representations present in the ‘hub’. In this way it is thought that the ‘hub’ of the model is where the concept is represented, with the spokes being more ancillary. Although similar to the convergence zone model, there are some points of difference between the hub-and-spoke model and the convergence-zone hypothesis. Quoting:

“The convergence-zone hypothesis, however, differs in at least two respects from the distributed-plus hub view. First, it proposes the existence of multiple specialized convergence regions. [...] Second, it suggests that these zones become differentially important for representing different semantic categories.” (Patterson et al. 2007)

Similar to the convergence-zone hypothesis, experiential models, such as the one proposed by Binder and Desai (2011) also differ from the hub-and-spoke model on these two fronts. Experiential theories of semantics predict hierarchical convergence zones

beyond “the unimodal ‘spokes and amodal ‘hubs’ in the hub-and-spoke model” (Fernandino et al. 2016). While many hub-and-spoke theorists place the highest level of abstraction in the temporal pole, the experiential model examined throughout this dissertation more closely aligns with Damasio’s notion of convergence zones in that it predicts that the areas of cortex situated between modal sensory-motor systems contain cross-modal conjunctive representations (Binder and Desai 2011).

Beyond the anatomical considerations of the hub-and-spoke model, one other point that requires clarification is the nature of the representations in the ‘amodal’ hub, apart from simply defining amodal as differing from the “formats of perceptual and motor representations” (Machery 2016).

As already referenced, the central tenet of grounded and experiential models is that the same brain structures that are important for perception are also important for representing conceptual information. As it relates to apparent category localization in the cortex, there are at least two key premises that an experiential explanation rests on. First is that some areas of the cortex preferentially process certain sensory, motor, affective and other experiential phenomena. This is to say that the representation of experiential information (e.g., touch, sound, taste, vision, etc.) has a degree of spatial variation. These representations are not expected to be completely segregated, but instead support a hierarchically organized representational system. Following from this, experiential models predict that concepts differentially depend on the computations done in unimodal, bimodal, and multimodal association areas. In this way experiential models make strong claims about the connection between semantics and the computations performed in distributed association cortices.

The second premise is that certain types of experiential information are differentially important across categories, similar to the original SFH model. For example, it is thought that motor action patterns are more important for tools and other manipulable artifacts than for animals (Grafton et al. 1997). From these premises the general explanation for CRSDs offered by experiential accounts is that categories have a localization as a result of spatial constraints imposed by where experiential features are processed. These premises are the same ones that motivated the SFH, and these newer models are essentially expansions of the dichotomous sensory-functional distinction of the SFH to a high-dimensional space comprising many more experiential modalities.

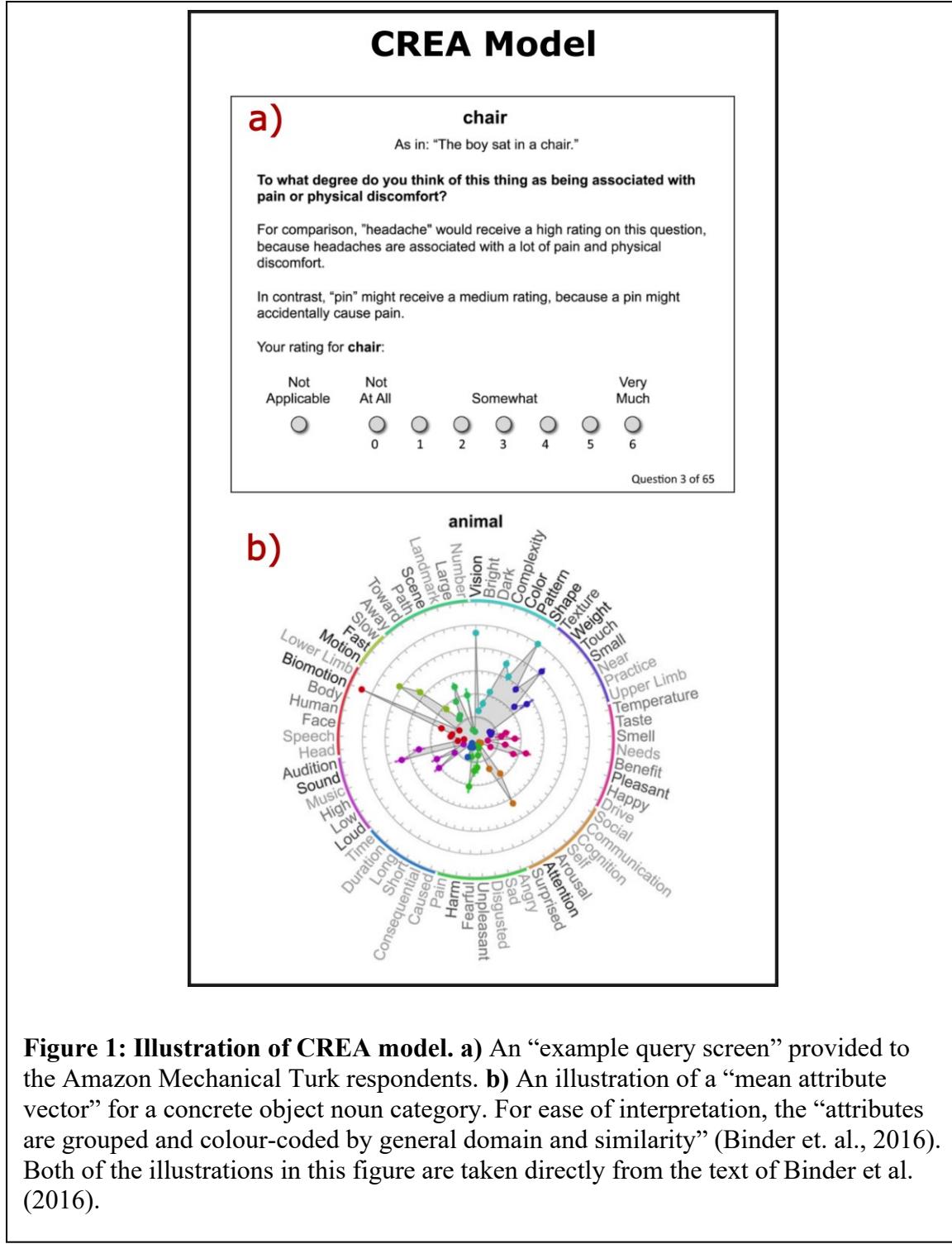
Although both amodal and experiential models have supporting evidence, there are critical gaps in precisely pinning down the exact implementational details of each type of model. Arguably, one of the most critical predictions of experiential models lies in its explanation of the apparent category related organization of the cortex. If experiential features are the basis of concept representation, then one would expect a common set of experiential features to account for activation patterns in a way that generalizes across categories. Another key prediction is that accessing conceptual information should entail activation of relevant association cortices. Both of these hypotheses are operationalized and tested in the subsequent chapters of this dissertation.

## Operationalized models

### Concept representation as experiential attributes

As stated, the primary model under investigation throughout this dissertation is the experiential model introduced by Binder and colleagues (2016). This model, called concept representation as experiential attributes (CREA), contains 65 components across

14 domains. The experiential representations for words were constructed from human ratings. In brief, the experiential components were selected based on known neural processing systems such as color, shape, visual motion, touch, audition, motor control, and olfaction, as well as other fundamental aspects of experience whose neural substrates are less clearly understood, such as space, time, affect, reward, numerosity, and others. As described previously (Binder et al. 2016), ratings were collected using the crowd sourcing tool Amazon Mechanical Turk to present a series of 65 queries about a target concept. Volunteers rated the relevance of each experiential domain to a given concept on a 0–6 Likert scale. Throughout this dissertation, the final vectorized representations of concepts was found by averaging ratings across participants for each concept and each feature. This feature set was highly effective at clustering concepts into canonical taxonomic categories (e.g., animals, plants, vehicles, occupations, etc.; Binder et al. 2016) and has been used successfully to decode fMRI activation patterns during sentence reading (Anderson et al. 2016a; Anderson et al. 2019). Previous comparisons with other semantic models, including WordNet, Word2Vec, and GloVe, showed that this experiential model is better at predicting neural similarity structure in the general semantic network, and those analyses found that other models did not add predictive



**Figure 1: Illustration of CREA model. a)** An “example query screen” provided to the Amazon Mechanical Turk respondents. **b)** An illustration of a “mean attribute vector” for a concrete object noun category. For ease of interpretation, the “attributes are grouped and colour-coded by general domain and similarity” (Binder et. al., 2016). Both of the illustrations in this figure are taken directly from the text of Binder et al. (2016).

power beyond that provided by the experiential model (Fernandino et al. 2022; Tong et al. 2022).

## Distributional Models

The distributional family of models is one where word meaning is specified in terms of usage. Although not directly related to any of the models of semantics already put forward, because distributional models do not directly appeal to any neurobiological mechanisms but generate word representations based only on co-occurrences of abstract symbols (i.e., words), they are sometimes considered as the stand-in for an operationalized amodal representation.

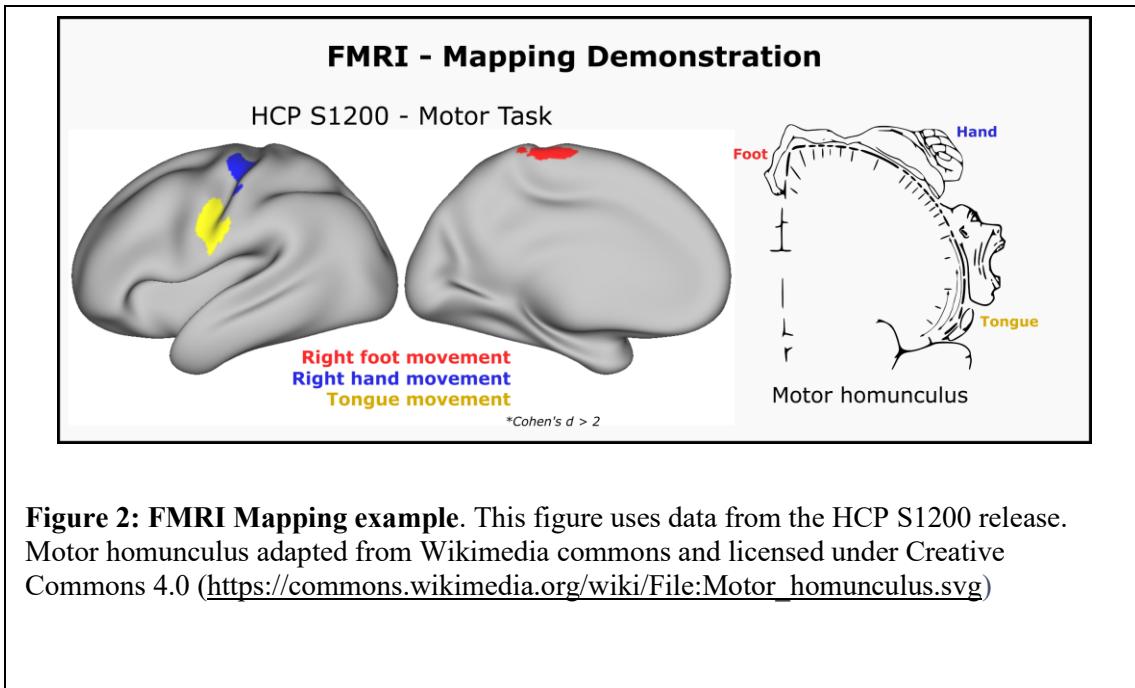
One way to define a distributional model is by the distributional hypothesis that “linguistic items with similar distributions have similar meanings” (Harris 1954; Liu et al. 2020)<sup>9</sup>. The underling idea is summarized in a popular quote that says “you shall know a word by the company it keeps” (Firth 1957). The basic heuristic at play is that words used in similar contexts (e.g., words that can be substituted in the sentence “the family purchased a new pet [blank]”), have similar meanings. One popular implementation of this idea is Word2Vec (Mikolov et al. 2013), where, in the “bag-of-words” version, a shallow neural network is used to predict the likelihood of a word given some context.

## Approach

This dissertation is structured to explicitly test several hypotheses implied by an experiential account of concept representation. To accomplish this, we first provide technical validation of the neuroimaging data analyzed throughout this dissertation and

---

<sup>9</sup> The exact quote comes from Liu et al. (2020), but it is commonly used definition. Harris (1954) is the standard citation provided for the distributional hypothesis, however, I was unable to find “distributional hypothesis” explicitly defined as it is commonly used in contemporary texts.



**Figure 2: FMRI Mapping example.** This figure uses data from the HCP S1200 release. Motor homunculus adapted from Wikimedia commons and licensed under Creative Commons 4.0 ([https://commons.wikimedia.org/wiki/File:Motor\\_homunculus.svg](https://commons.wikimedia.org/wiki/File:Motor_homunculus.svg))

discuss experimental design considerations for ways to optimize future data collection.

This is then followed by examination of the experiential predictions.

Functional Magnetic Resonance Imaging: The primary tool for investigating the localization of function in the brain is functional magnetic resonance imaging (fMRI). The general principle that underlies fMRI is what is known as T2\* relaxation. In short, a specific type of MRI is differentially sensitive to oxygenated and deoxygenated hemoglobin, leading to what is known as blood-oxygen level dependent (BOLD) imaging. This type of imaging allows for mapping regions of the brain where changes in blood flow are correlated with task demands. The efficacy of this imaging technique was famously demonstrated by researchers at the Medical College of Wisconsin using a “finger tapping” paradigm (Bandettini 2012). **Figure 2** shows the results of a modified version of this task demonstrating the ability of fMRI to map regions with independently known functions (e.g., the sensory motor homunculus). FMRI has subsequently become a

ubiquitous tool in research for relating brain structures to different functions, and is the tool used throughout this dissertation to map regions of cortex that reliably activate or deactivate in response to the experimental task.

While fMRI has proved to be a powerful tool, it is also known for being a tool that generates “noisy” data, which necessitates averaging signal from many measurements. In general, signal is averaged over multiple presentations of stimuli, and across participants. We carried out retrospective analyses on our data to quantify the benefit of different signal averaging approaches to guide future experiments, and in the process performed thorough validation of the quality of data used throughout subsequent chapters of this dissertation.

*Stimulus repetition and sample size considerations (Chapter 2):* The second chapter covers an understudied topic in newer fMRI studies based on multivoxel pattern analysis; experimental design. How to properly power a study is a notoriously difficult task in many fMRI experiments. While statistical power has a strong dependence on the reliability of the underlying measurement, little is known about how experimental design choices influence reliability. This chapter investigates the implicit trade-off that researchers commonly face between number of participants and amount of data collected from each participant. While it is ostensibly obvious that the more repetitions of stimuli, the better the resolving ability, little work has been done to quantify the improvement one gets for additional repetitions compared to additional number of participants.

Questions around data quality are discussed prior to testing claims that directly bear on experiential accounts of category organization to demonstrate data quality and to provide increased confidence in the data pipeline upon which subsequent chapters are

built. This chapter also provides practical insight for future experiments as up until our analysis, no studies had examined the improvement brought by stimulus repetitions compared to additional participants in the context of a multivariate analysis technique employed in later chapters of the dissertation.

*A shared representational basis for dissociable semantic categories (Chapter 3):* As extensively motivated throughout this introduction, a critical test for grounded theories of cognition is that the observed phenomena of category related organization can be explained by weighted combinations of feature sensitivity maps. This hypothesis is tested in this chapter for four of the most prominent categories that have been found to dissociate in the CRSD literature. The analysis in this chapter expands on previous results in the literature by explicitly focusing on the ability of feature maps to generalize across categories of concepts.

*The neural representation of body part concepts (Chapter 4):* This chapter builds on the previous chapter and does an in-depth examination on one category of concrete object nouns that, relative to the other categories, has received less attention: body parts. Although there is strong evidence for selective loss and preservation of body part knowledge, this category has often not been included in fMRI studies on semantic representation. A finding of the previous chapter was that in univariate pair-wise contrasts, body parts demonstrated a large and robust activation greater than other categories in a large left-lateralized network. This chapter examines the body-part category in more detail, with a focus on an experiential account of the category's representation.

*Selected experiential feature maps (Chapter 5):* The previous chapters examined the ability of an experiential model to account for fMRI activation patterns in a generalizable way. However, they did not examine a core prediction of the experiential account: that the spatial layout of the semantic system is constrained by perceptual processors. This prediction implies two results that are tested in the last chapter of this dissertation. The first is that the semantic system is widespread and characterized by hierarchical convergence zones of different input modalities. The second is that the spatial layout of experiential feature representation should overlap with and be constrained by known perceptual processors. Both of these predictions are tested in this chapter through the generation of a selected set of experiential feature maps. Given that the second prediction requires interpretation based on relatively well known perceptual processors, a subset of the 65 experiential features was selected based on the criteria that many cognitive neuroscientists would have prior expectations/intuitions about where the feature might localize. This chapter then discusses the convergence and interpretability of feature maps in light of extant literature.

# Chapter 2: Stimulus Repetition and Sample Size Considerations<sup>10</sup>

## Introduction

Representational similarity analysis (RSA) is an increasingly popular multi-voxel pattern analysis method in which measured neural activity patterns are used to construct a neural representational dissimilarity matrix (RDM), which is then compared to dissimilarity matrices predicted by one or more stimulus models or measured using an entirely different neural encoding technique (Kriegeskorte et al. 2008). This technique has been used several times by members in our own lab and throughout the rest of this dissertation. Common aims of RSA include comparing different stimulus models in terms of strength of correlation with the neural RDM as a means of adjudicating the validity of different models, and mapping the brain regions where neural activity patterns reflect a particular kind of information. Achieving these aims depends on the reliability of the estimated neural RDM, which like all physiological recordings is affected by noise and individual variability. This chapter examines the quality of our data. Here we consider two strategies for enhancing reliability of the neural RDM, namely, presenting the stimulus set multiple times (and combining the signal across presentations) and increasing the number of participants.

---

<sup>10</sup> The primary results of this have been published as: Mazurchuk, S., Conant, L. L., Tong, J.-Q., Binder, J. R. & Fernandino, L. Stimulus repetition and sample size considerations in item-level representational similarity analysis. *Lang., Cogn. Neurosci.*, 1–12 (2023)

Effects of sample size on reliability have been a focus of some univariate fMRI studies (e.g., Thirion et al. 2007; Turner et al. 2018). These studies demonstrate important trade-offs between the number of participants and amount of data collected in each participant. When limited amounts of individual-level data are acquired, results may still not be replicable even with large sample sizes (Nee 2019; Turner et al. 2018). Increasing the amount of data acquired for each participant may allow adequate reliability to be achieved with smaller sample sizes (Nee 2019). However, considerably less is known about how these factors impact the reliability of results in multivariate (e.g., multivoxel) pattern analysis.

Except for the choice of distance metric (Allefeld and Haynes 2014; Bobadilla-Suarez et al. 2020; Ritchie et al. 2021; Walther et al. 2016), little attention has been given to the factors that may contribute to the reliability of the neural RDM. In most RSA studies, stimuli are repeated so that the signal-to-noise ratio can be increased through averaging. This is particularly true for studies in which the RDM is constructed from neural responses to single items (rather than categories of items), which we refer to as “item-level RSA”. However, the benefit of averaging across repetitions may be offset by the possibility of response adaptation, also known as repetition suppression, in which subsequent neural responses to the same stimulus are weaker as measured by hemodynamic methods (Dobbins et al. 2004; Grill-Spector et al. 2006; Lee et al. 2020). Notably, it is unclear whether this decreased magnitude of responses is problematic for RSA, as Arbuckle et al. (2019) found that representational geometry stayed relatively stable across significant changes in average fMRI activity. While repetition suppression

| Title  | First Author           | Year | Journal                  | Number of Participants | Number of Stimuli | Number of presentations |
|--|------------------------|------|--------------------------|------------------------|-------------------|-------------------------|
| Neural similarities and differences between native and second languages in the bilateral fusiform cortex in Chinese–English bilinguals               | Liu, Xiaoyu            | 2023 | Neuropsychologia         | 27                     | 64                | 4                       |
| The depth of semantic processing modulates cross-language pattern similarity in Chinese–English bilinguals   | Li, Huijing            | 2022 | Hum. Brain Mapp.         | 26                     | 120               | 1                       |
| Functional Gradient of the Fusiform Cortex for Chinese Character Recognition   | Guo, Wanwan,           | 2022 | Eneuro                   | 51                     | 120               | 1                       |
| Context free and context-dependent conceptual representation in the brain  | Gao, Zhiyao            | 2022 | Cereb Cortex             | 28                     | 192               | 1                       |
| A Distributed Network for Multimodal Experiential Representation of Concepts   | Tong, Jiaqing          | 2022 | J Neurosci               | 39                     | 320               | 6                       |
| Orienting to Different Dimensions of Word Meaning Alters the Representation of Word Meaning in Early Processing Regions                              | Meersmans, Karen       | 2021 | Cereb Cortex             | 22                     | 120               | 8                       |
| Language distance in orthographic transparency affects cross-language pattern similarity between native and non-native languages                     | Dong, Jie              | 2021 | Hum. Brain Mapp.         | 23                     | 240               | 1                       |
| Representation of associative and affective semantic similarity of abstract words in the lateral temporal perisylvian language regions               | Meersmans, Karen       | 2020 | Neuroimage               | 26                     | 64                | 4                       |
| Neural Components of Reading Revealed by Distributed and Symbolic Computational Models   | Staples, Ryan          | 2020 | Neurobiology Lang        | 18                     | 464               | 1                       |
| Distinct Fronto-temporal substrates of distributional and taxonomic similarity among words: evidence from RSA of BOLD signals                        | Carota, Francesca      | 2020 | Neuroimage               | 23                     | 96                | 6                       |
| Lexical learning in a new language leads to neural pattern similarity with word reading in native language   | Li, Huijing            | 2019 | Hum. Brain Mapp.         | 24                     | 90                | 6                       |
| Cross-Language Pattern Similarity in the Bilateral Fusiform Cortex Is Associated with Reading Proficiency in Second Language                         | Qu, Jing               | 2019 | Neuroscience             | 39                     | 120               | 2                       |
| Representational Similarity Mapping of Distributional Semantics in Left Inferior Frontal, Middle Temporal, and Motor Cortex                          | Carota, Francesca      | 2017 | Cereb Cortex New York Ny | 23                     | 96                | 6                       |
| Representational similarity analysis reveals task-dependent semantic influence of the visual word form area.   | Wang, Xiaosha          | 2017 | Sci Rep-uk               | 20                     | 45                | 10                      |
| Decoding levels of representation in reading: A representational similarity approach   | Fischer-Baum, Simon    | 2017 | Cortex                   | 20                     | 35                | 12                      |
| Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding  | Borghesani, Valentina  | 2016 | Neuroimage               | 16                     | 24                | 4                       |
| Representational similarity encoding for fMRI: Pattern-based synthesis to predict brain activity using stimulus-model-similarities                   | Anderson, Andrew James | 2016 | Neuroimage               | 9                      | 60                | 6                       |
| Reading visually embodied meaning from the brain: Visually grounded computational models decode visual-object mental imagery induced by written text | Anderson, Andrew James | 2015 | Neuroimage               | 11                     | 51                | 6                       |
| Similarity of fMRI Activity Patterns in Left Perirhinal Cortex Reflects Semantic Similarity between Words  | Bruffaerts, Rose       | 2013 | J Neurosci               | 19                     | 24                | 4                       |
| Representational Similarity Analysis Reveals Commonalities and Differences in the Semantic Processing of Words and Objects                           | Devereux, B.J          | 2013 | J Neurosci               | 14                     | 60                | 6                       |

**Table 2: Characteristics of previous RSA studies.** Number of participants, stimuli, and presentations used in previous RSA studies.

has been extensively investigated in univariate analysis of fMRI data, its impact on RSA results is largely unknown.

Exemplifying the need for a better understanding of these trade-offs, a search of the RSA literature restricted to item-level RSA studies using visually presented single words as stimuli reveals a large variation in participant numbers and stimulus presentations (**Table 2**). Many studies have around 20 participants (Carota et al. 2020;

Dong et al. 2021; Li et al. 2019; Meersmans et al. 2021; Staples and Graves 2020; Wang et al. 2017), with a range from 9 (Anderson et al. 2016b) to 51 (Guo et al. 2022). The median number of stimulus presentations was 4 (Borghesani et al. 2016; Bruffaerts et al. 2013; Liu et al. 2023; Meersmans et al. 2021), with a range from 1 (i.e., no repetitions) (Dong et al. 2021; Gao et al. 2022; Guo et al. 2022; Li et al. 2022; Staples and Graves 2020) to 12 (Fischer-Baum et al. 2017).

While the effects of stimulus repetition and participants sample size on neural RDM reliability are largely unexamined, the importance of between-subject reliability for the interpretation of RSA results is widely recognized (Ritchie et al. 2021). Reliability of the neural RDM is an essential prerequisite for making valid inferences regarding relationships with model RDMs and discriminating among models. In the RSA literature, the reliability metrics typically used are the upper and lower bound estimates of the noise ceiling, which are measures of the highest potential performance of a model for a given data set. Noise ceiling estimates allow the obtained RSA correlation to be interpreted in the context of the limitations of the specific experiment (e.g., level of measurement noise, sample size, amount of individual-level data). In the current study, to further assess interindividual consistency, or internal consistency of the group-averaged neural RDM, and to relate our analysis to the broader context of classical measurement theory, we also examined another measure of reliability, Cronbach's alpha (hereafter referred to as "alpha"). This choice was motivated by the close connections between alpha and other reliability measures such as split-half correlation and intra-class correlation (ICC), which allow it to also be interpreted as a measure of reproducibility (Bravo and Potvin 1991), along with its quick and efficient calculation, allowing for resampling analysis across

varying numbers of participants and stimulus presentations. In addition, alpha allows researchers to estimate reliabilities for sample sizes larger than the obtained sample (de Vet et al. 2017) .

Here we examined sample size and stimulus repetition effects in two large data sets using item-level RSA with lexical stimuli. In both studies, a large stimulus set was presented 6 times over three scanning sessions on separate days. The analyses focused on a region of interest (ROI) encompassing primarily multimodal cortical areas in the frontal, temporal, and parietal lobes, derived from a previous ALE meta-analysis of semantic cognition studies (Binder et al. 2009). This large ROI provides a global estimate of RDM reliability across the high-level association areas of the cortex, averaging over local differences in reliability due to spatial variation in MRI signal quality. To ensure that the pattern of results was not dependent on the particular choice of ROI, we also replicated some of the analyses in smaller ROIs defined within a commonly used cortical parcellation (Desikan et al. 2006), but these results are not presented here. The goal of the analyses is to validate our own data collection and processing technique and to provide researchers with useful information on the potential trade-offs between the number and schedule of stimulus presentations and the number of participants to help optimize study design and use of limited resources, such as scanner time. Importantly, the intention is not to make strong recommendations regarding optimal values for these parameters, as these can vary based on the experimental design, but rather to examine how specific design choices may affect reliability in condition-rich, item-level RSA studies in terms of their general trends.

## Materials and Methods

### ***Participants***

The chapters of this dissertation have slightly different numbers of participants as analysis was performed at different time points and used different inclusion/exclusion criteria. A listing of which participants were analyzed in each chapter, along with some demographic information is shown in **Table 3**. For the purpose of this chapter, Study 1 included 40 adult, native speakers of English (25 women) with a mean age of 28 years. While Study 2 ultimately enrolled 45 participants the data analyzed for Study 2 in this in this chapter only included 39 adult, native speakers of English (27 women) with a mean age of 27 years. Four individuals participated in both studies. All participants were right-handed according to the Edinburgh Handedness Scale (Oldfield 1971), had at least a high school education, and had no history of neurologic or psychiatric conditions. All participants gave written informed consent and were compensated for their time. The study was approved by the Medical College of Wisconsin Institutional Review Board.

### ***Stimuli and task***

The stimuli in Study 1, described in detail in a previous publication (Fernandino et al. 2022), consisted of 320 English nouns, half of which were names of objects and half names of events. Object nouns were animals, tools, plants/foods, and vehicles; event nouns were sounds, negative events, social events, and communication events. The stimuli in Study 2, described in Tong et al. (2022), were 300 English nouns from 6 categories: animals, artifacts, plants/foods, body parts, human traits, and quantities.

| Subject ID      | Gender | Age | Education (Years) | Number of Sessions | Button-Press ICC | Chapter 2 | Chapter 3 | Chapter 4 | Chapter 5 |
|-----------------|--------|-----|-------------------|--------------------|------------------|-----------|-----------|-----------|-----------|
| SOE101          | Male   | 22  | 15.5              | 3                  | 0.66             | Study 1   |           |           | Included  |
| SOE102          | Female | 21  | 13                | 2                  | 0.68             |           |           |           | Included  |
| SOE103          | Male   | 24  | 12                | 3                  | 0.71             | Study 1   |           |           | Included  |
| SOE104          | Male   | 27  | 20                | 3                  | 0.58             | Study 1   |           |           | Included  |
| SOE105 / CAT135 | Female | 34  | 16                | 3                  | 0.71             | Study 1   |           |           | Included  |
| SOE106          | Female | 32  | 18                | 3                  | 0.46             | Study 1   |           |           | Included  |
| SOE107          | Female | 36  | 16                | 2                  | 0.09             |           |           |           |           |
| SOE108          | Female | 33  | 18                | 3                  | 0.67             | Study 1   |           |           | Included  |
| SOE109          | Female | 33  | 18                | 3                  | 0.83             | Study 1   |           |           | Included  |
| SOE110          | Male   | 38  | 20                | 3                  | 0.70             | Study 1   |           |           | Included  |
| SOE111          | Male   | 41  | 16                | 3                  | 0.52             | Study 1   |           |           | Included  |
| SOE112          | Female | 33  | 22                | 3                  | 0.58             | Study 1   |           |           | Included  |
| SOE113          | Female | 34  | 12.5              | 3                  | 0.73             | Study 1   |           |           | Included  |
| SOE114          | Female | 32  | 21                | 3                  | 0.78             | Study 1   |           |           | Included  |
| SOE115          | Male   | 40  | 16                | 3                  | 0.69             | Study 1   |           |           | Included  |
| SOE116          | Female | 33  | 18                | 3                  | 0.60             | Study 1   |           |           | Included  |
| SOE117          | Male   | 24  | 16                | 3                  | 0.73             | Study 1   |           |           | Included  |
| SOE118          | Male   | 20  | 14                | 3                  | 0.75             | Study 1   |           |           | Included  |
| SOE120          | Female | 21  | 15                | 3                  | 0.74             | Study 1   |           |           | Included  |
| SOE121          | Male   | 23  | 16                | 3                  | 0.64             | Study 1   |           |           | Included  |
| SOE122          | Male   | 21  | 15                | 3                  | 0.44             | Study 1   |           |           | Included  |
| SOE123          | Male   | 24  | 18                | 3                  | 0.72             | Study 1   |           |           | Included  |
| SOE124          | Female | 32  | 17                | 3                  | 0.61             | Study 1   |           |           | Included  |
| SOE125          | Female | 22  | 17                | 3                  | 0.86             | Study 1   |           |           | Included  |
| SOE126 / CAT130 | Female | 38  | 17                | 3                  | 0.70             | Study 1   |           |           | Included  |
| SOE128          | Male   | 30  | >12               | 3                  | 0.23             | Study 1   |           |           | Included  |
| SOE129          | Female | 22  | 16                | 3                  | 0.74             | Study 1   |           |           | Included  |
| SOE130          | Female | 33  | 20                | 3                  | 0.67             | Study 1   |           |           | Included  |
| SOE131          | Female | 25  | 17.5              | 3                  | 0.77             | Study 1   |           |           | Included  |
| SOE133          | Female | 27  | 18                | 3                  | 0.60             | Study 1   |           |           | Included  |
| SOE134          | Female | 22  | 15                | 3                  | 0.81             | Study 1   |           |           | Included  |
| SOE135          | Female | 20  | 14                | 3                  | 0.60             | Study 1   |           |           | Included  |
| SOE136          | Male   | 34  | 16                | 3                  | 0.75             | Study 1   |           |           | Included  |
| SOE137          | Male   | 24  | 17                | 3                  | 0.65             | Study 1   |           |           | Included  |
| SOE138          | Male   | 37  | 16                | 3                  | 0.58             | Study 1   |           |           | Included  |
| SOE139          | Female | 26  | 20                | 3                  | 0.45             | Study 1   |           |           | Included  |
| SOE140          | Female | 34  | 19                | 3                  | 0.48             | Study 1   |           |           | Included  |
| SOE141          | Female | 27  | 17                | 3                  | 0.73             | Study 1   |           |           | Included  |
| SOE142          | Female | 25  | 20                | 3                  | 0.78             | Study 1   |           |           | Included  |
| SOF143 / CAT108 | Female | 26  | 16                | 3                  | 0.61             | Study 1   |           |           | Included  |
| SOE144 / CAT113 | Female | 20  | 14.5              | 3                  | 0.57             | Study 1   |           |           | Included  |
| SOE145          | Female | 38  | 25                | 3                  | 0.55             | Study 1   |           |           | Included  |
| CAT101          | Female | 21  | 13.5              | 3                  | 0.74             | Study 2   | Included  | Included  | Included  |
| CAT102          | Female | 28  | 23                | 3                  | 0.66             | Study 2   | Included  | Included  | Included  |
| CAT103          | Female | 24  | 18.5              | 3                  | 0.62             | Study 2   | Included  | Included  | Included  |
| CAT104          | Female | 28  | 20                | 3                  | 0.75             | Study 2   | Included  | Included  | Included  |
| CAT105          | Female | 20  | 14.5              | 3                  | 0.61             | Study 2   | Included  | Included  | Included  |
| CAT106          | Female | 18  | 12.5              | 3                  | 0.57             | Study 2   | Included  | Included  | Included  |
| CAT107          | Female | 20  | 15                | 3                  | 0.64             | Study 2   | Included  | Included  | Included  |
| CAT108 / SOE143 | Female | 26  | 16                | 3                  | 0.61             | Study 2   | Included  | Included  | Included  |
| CAT109          | Female | 35  | >12               | 3                  | 0.46             | Study 2   | Included  | Included  | Included  |
| CAT110          | Female | 23  | 16.5              | 3                  | 0.84             | Study 2   | Included  | Included  | Included  |
| CAT111          | Female | 25  | 12                | 2                  | 0.67             |           |           |           | Included  |
| CAT113 / SOE144 | Female | 21  | 14.5              | 3                  | 0.67             | Study 2   | Included  | Included  | Included  |
| CAT114          | Female | 19  | 13.5              | 3                  | 0.78             | Study 2   | Included  | Included  | Included  |
| CAT116          | Female | 23  | 17                | 3                  | 0.59             | Study 2   | Included  | Included  | Included  |
| CAT117          | Female | 21  | 14                | 3                  | 0.59             | Study 2   | Included  | Included  | Included  |
| CAT118          | Male   | 40  | 12                | 3                  | 0.54             | Study 2   | Included  | Included  | Included  |
| CAT119          | Female | 29  | 19                | 3                  | 0.73             | Study 2   | Included  | Included  | Included  |
| CAT120          | Female | 33  | 22                | 3                  | 0.78             | Study 2   | Included  | Included  | Included  |
| CAT121          | Female | 25  | >12               | 3                  | 0.61             | Study 2   | Included  | Included  | Included  |
| CAT122          | Male   | 37  | >12               | 3                  | 0.59             | Study 2   | Included  | Included  | Included  |
| CAT125          | Female | 26  | 16                | 3                  | 0.63             | Study 2   | Included  | Included  | Included  |
| CAT126          | Female | 26  | 12                | 3                  | 0.70             | Study 2   | Included  | Included  | Included  |
| CAT127          | Male   | 23  | 14                | 3                  | 0.66             | Study 2   | Included  | Included  | Included  |
| CAT129          | Male   | 38  | 16                | 3                  | 0.64             | Study 2   | Included  | Included  | Included  |
| CAT130 / SOE126 | Female | 40  | 16                | 3                  | 0.69             | Study 2   | Included  | Included  | Included  |
| CAT131          | Male   | 36  | 14                | 3                  | 0.12             | Study 2   |           |           |           |
| CAT132          | Male   | 27  | 16                | 3                  | 0.72             | Study 2   | Included  | Included  | Included  |
| CAT133          | Male   | 20  | 14.5              | 3                  | 0.75             | Study 2   | Included  | Included  | Included  |
| CAT134          | Female | 22  | 15                | 3                  | 0.62             | Study 2   | Included  | Included  | Included  |
| CAT135 / SOE105 | Female | 36  | 16                | 3                  | 0.68             | Study 2   | Included  | Included  | Included  |
| CAT136          | Female | 29  | 16                | 3                  | 0.60             | Study 2   | Included  | Included  | Included  |
| CAT137          | Female | 27  | 16                | 3                  | 0.80             |           | Included  | Included  | Included  |
| CAT138          | Male   | 19  | 13                | 3                  | 0.39             | Study 2   | Included  | Included  | Included  |
| CAT140          | Male   | 28  | 16                | 3                  | 0.49             | Study 2   | Included  | Included  | Included  |
| CAT141          | Male   | 24  | 19                | 3                  | 0.37             | Study 2   | Included  | Included  | Included  |
| CAT142          | Female | 28  | 21                | 3                  | 0.81             | Study 2   | Included  | Included  | Included  |
| CAT143          | Male   | 25  | 17                | 3                  | 0.12             | Study 2   |           |           |           |
| CAT144          | Female | 38  | 20                | 3                  | 0.54             | Study 2   | Included  | Included  | Included  |
| CAT145          | Female | 30  | 18                | 2                  | 0.74             |           |           |           | Included  |
| CAT146          | Female | 24  | 16                | 3                  | 0.76             | Study 2   | Included  | Included  | Included  |
| CAT147          | Male   | 29  | 17                | 3                  | 0.79             | Study 2   | Included  | Included  | Included  |
| CAT149          | Male   | 22  | 16                | 3                  | 0.86             | Study 2   | Included  | Included  | Included  |

**Table 3: Participant demographics and study assignment.** Button-Press ICC refers to the consistency, assessed by a consistency-based intraclass correlation coefficient, of the participant's rating across different presentations of the stimuli. A minimum ICC greater than 0.2 was used as inclusion criteria for chapters 3, 4, and 5 of this dissertation.

| Objects     |              |           |            |            |             |             |            |               |            | Events        |              |              |              |  |  |
|-------------|--------------|-----------|------------|------------|-------------|-------------|------------|---------------|------------|---------------|--------------|--------------|--------------|--|--|
| Animal      |              | Food      |            | Tool       |             | Vehicle     |            | Communication |            | Natural Event |              | Social Event |              |  |  |
| alligator   | hamster      | asparagus | eggplant   | anchor     | hoe         | ambulance   | plane      | advice        | lecture    | avalanche     | landslide    | banquet      | gathering    |  |  |
| ant         | hawk         | banana    | flower     | axe        | key         | automobile  | rocket     | apology       | lesson     | battle        | lightning    | bash         | housewarming |  |  |
| baboon      | hippopotamus | bean      | ham        | baseball   | keyboard    | barge       | rowboat    | class         | meeting    | blizzard      | monsoon      | carnival     | jubilee      |  |  |
| bison       | horse        | beer      | honey      | binoculars | ladle       | bicycle     | sailboat   | commemoration | plea       | bombing       | murder       | celebration  | luncheon     |  |  |
| butterfly   | jackal       | blueberry | jam        | book       | magazine    | boat        | scooter    | comment       | praise     | brawl         | outbreak     | christening  | march        |  |  |
| cardinal    | lion         | bread     | ketchup    | calculator | microscope  | bobsled     | skateboard | commentary    | protest    | cyclone       | plague       | circus       | musical      |  |  |
| caterpillar | monkey       | broccoli  | lemonade   | camera     | newspaper   | bus         | sled       | complaint     | quarrel    | downpour      | raid         | cocktail     | outing       |  |  |
| chameleon   | moose        | carrot    | milk       | candle     | pencil      | canoe       | sleigh     | compliment    | rant       | drought       | riot         | concert      | pageant      |  |  |
| cheetah     | mosquito     | champagne | mushroom   | cash       | rake        | car         | steamer    | debate        | rebuke     | earthquake    | shooting     | conference   | parade       |  |  |
| chicken     | mouse        | cheese    | mustard    | comb       | sandpaper   | carriage    | streetcar  | denial        | rebuttal   | epidemic      | squall       | contest      | party        |  |  |
| chimpanzee  | octopus      | cherry    | nectarine  | corkscrew  | scissors    | convertible | submarine  | deposition    | recitation | explosion     | stampede     | convention   | picnic       |  |  |
| chipmunk    | penguin      | chestnut  | pineapple  | crutch     | skillet     | elevator    | subway     | dictation     | sermon     | famine        | storm        | cookout      | prom         |  |  |
| cricket     | rhinoceros   | chocolate | plant      | dime       | spatula     | escalator   | taxi       | discourse     | showdown   | flood         | tempest      | cruise       | rally        |  |  |
| crow        | salmon       | cider     | pudding    | faucet     | stapler     | ferry       | tractor    | dispute       | squabble   | gunshot       | thunderstorm | dance        | reception    |  |  |
| dog         | snake        | coffee    | pumpkin    | football   | stethoscope | glider      | train      | eulogy        | testimony  | gust          | tornado      | expedition   | reunion      |  |  |
| dolphin     | tiger        | cranberry | raspberry  | fork       | straw       | helicopter  | tricycle   | greeting      | thanks     | hail          | twister      | expo         | safari       |  |  |
| duck        | trout        | cucumber  | sauerkraut | glass      | thermometer | jeep        | trolley    | grievance     | threat     | hailstorm     | volcano      | fair         | symphony     |  |  |
| elephant    | turkey       | custard   | spaghetti  | hairbrush  | ticket      | limousine   | truck      | huddle        | trial      | hurricane     | war          | feast        | tour         |  |  |
| fish        | turtle       | dandelion | tobacco    | hammer     | tongs       | locomotive  | van        | interrogation | tribute    | inferno       | whirlwind    | festival     | tournament   |  |  |
| goldfish    | whale        | egg       | tomato     | handsaw    | umbrella    | motorcycle  | wagon      | joke          | wisecrack  | invasion      | wildfire     | fiesta       | wedding      |  |  |

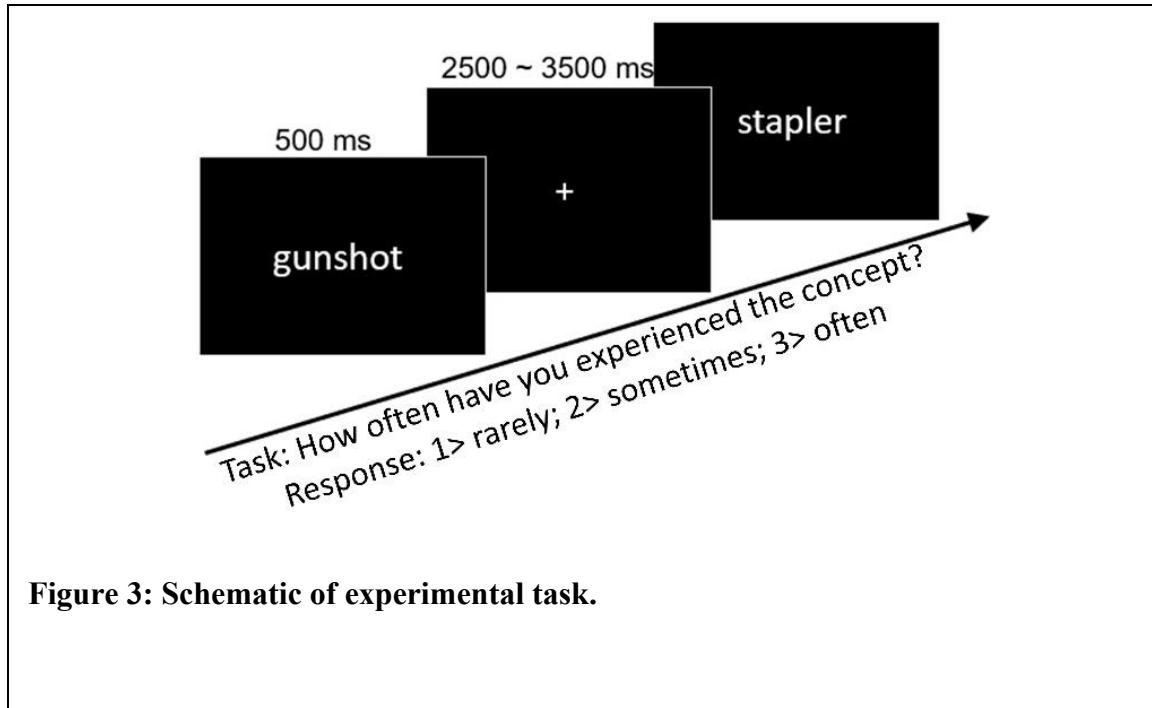
**Table 4: Stimuli for Study 1**

| Animal      |              | Artifact   |             | Body Part  |          | Human Trait |           | Plant-Food |            | Quantity   |            |
|-------------|--------------|------------|-------------|------------|----------|-------------|-----------|------------|------------|------------|------------|
| alligator   | hippopotamus | anchor     | ladle       | abdomen    | lip      | apathy      | humor     | apple      | eggplant   | acre       | lifetime   |
| baboon      | horse        | baseball   | magazine    | ankle      | liver    | arrogance   | kindness  | asparagus  | grapes     | adulthood  | loaf       |
| bird        | jackal       | binoculars | microscope  | armpit     | muscle   | charisma    | lust      | banana     | honey      | atom       | meter      |
| buffalo     | leopard      | bottle     | nail        | beard      | mustache | charm       | malice    | beer       | lemon      | billion    | molecule   |
| butterfly   | lion         | bowl       | pan         | belly      | navel    | civility    | maturity  | bread      | lemonade   | centimeter | month      |
| camel       | monkey       | broom      | pencil      | bladder    | nipple   | conceit     | mercy     | broccoli   | milk       | clan       | morsel     |
| cardinal    | moose        | calculator | pipe        | cartilage  | nose     | cynicism    | modesty   | butter     | mushroom   | colony     | nickel     |
| cat         | mosquito     | camera     | rake        | cheek      | nostril  | deceit      | paranoia  | cabbage    | mustard    | crumb      | ounce      |
| caterpillar | mouse        | candle     | rope        | clavicle   | pancreas | decency     | patience  | cake       | pear       | decade     | penny      |
| chameleon   | octopus      | cane       | ruler       | diaphragm  | pelvis   | devotion    | prejudice | carrot     | pineapple  | district   | pint       |
| cheetah     | penguin      | cash       | saddle      | earlobe    | retina   | disdain     | pride     | celery     | plant      | dollar     | platoon    |
| chicken     | pig          | comb       | sandpaper   | elbow      | shoulder | distrust    | regret    | cereal     | potato     | dozen      | realm      |
| chimpanzee  | rabbit       | corkscrew  | scale       | eyebrow    | skeleton | ego         | respect   | champagne  | pudding    | eighty     | semester   |
| chipmunk    | rhinoceros   | crutches   | scissors    | eyelid     | skull    | empathy     | sadism    | cheese     | pumpkin    | era        | speck      |
| cow         | salmon       | faucet     | screwdriver | finger     | spine    | envy        | scorn     | cherry     | raspberry  | flock      | splinter   |
| cricket     | seal         | flashlight | skillet     | fingernail | stomach  | ethics      | shame     | chestnut   | rice       | gallon     | stride     |
| deer        | snail        | football   | spatula     | forearm    | testicle | fame        | spite     | chocolate  | rose       | generation | swarm      |
| dog         | snake        | fork       | spoon       | forehead   | thigh    | glamor      | tact      | cider      | sandwich   | gram       | tablespoon |
| dolphin     | spider       | glass      | stapler     | heel       | thumb    | grace       | talent    | coffee     | sauerkraut | handful    | territory  |
| dragon      | squirrel     | hairbrush  | stethoscope | instep     | toenail  | greed       | tolerance | corn       | strawberry | heartbeat  | thousand   |
| elephant    | tiger        | hammer     | thermometer | intestines | tooth    | guile       | treachery | cucumber   | sugar      | herd       | ton        |
| fish        | trout        | handsaw    | tongs       | kidney     | torso    | guilt       | trust     | custard    | tangerine  | inch       | trio       |
| goat        | turkey       | iron       | towel       | knuckle    | trachea  | heroism     | vanity    | daisy      | tea        | infancy    | triplet    |
| goldfish    | turtle       | keyboard   | typewriter  | leg        | waist    | honesty     | wisdom    | dandelion  | tobacco    | instant    | weekend    |
| hamster     | whale        | knife      | umbrella    | ligament   | wrist    | honor       | wit       | egg        | tomato     | kilometer  | zone       |

**Table 5: Stimuli for Study 2**

Both experiments used the same stimulus presentation procedure and task. On each trial, a noun was displayed in white font on a black background at the center of the screen for 500 ms, followed by a 2.5-s blank screen. Participants were instructed to rate each word on how often they encountered the corresponding entity or event in their daily life, using a scale from 1 (“rarely or never”) to 3 (“often”). Responses were indicated using three keys operated with the right hand. Each trial was followed by a central fixation cross with variable duration between 1 and 3 s (mean 1.5s; **Figure 3**). Each run started and ended with an 8-s fixation cross. Stimulus presentation and response recording were performed with Psychopy 3 software (Peirce et al. 2019) running on a Windows desktop computer and a Celeritas fiber optic response system (Psychology Software Tools, Inc.). Stimuli were displayed on an MRI-compatible LCD screen positioned behind the scanner bore and viewed through a mirror attached to the head coil.

In both studies the entire stimulus set was presented six times with a different pseudo-randomized order for each presentation and each participant. The presentation of the complete set of stimuli was split into 4 runs each containing 80 trials for Study 1, and 3 runs each containing 100 trials in Study 2. Data collection was performed over the course of three scanning sessions on separate days, with two complete stimulus set presentations per session. For Study 1 the interval between sessions 1 and 2 averaged 18 days ( $sd = 24.8$ ), and the interval between sessions 2 and 3 averaged 25 days ( $sd = 37.1$ ). For Study 2, the respective intervals between the sessions were 23 days ( $sd = 29.1$ ) and



23.5 days ( $sd = 26.0$ ). These intervals did not significantly differ in either study

(Wilcoxon  $p = .98$  and  $p = .52$  in studies 1 and 2, respectively)

### ***MRI***

Scanning was performed on a GE Healthcare Premier 3T MRI scanner with a 32-channel Nova head coil at the Medical College of Wisconsin's Center for Imaging Research. Each session consisted of a structural T1-weighted MPRAGE scan, a structural T2-weighted CUBE scan, 3 pairs of T2-weighted spin echo echo-planar scans (5 volumes each) acquired with opposing phase-encoding directions for later image unwarping, and either 8 (Study 1) or 6 (Study 2) gradient-echo echo-planar imaging (EPI) functional scans (multiband factor = 4, TR = 1500 ms, TE = 33 ms, flip angle = 50, in-plane matrix = 104  $\times$  104, slice thickness = 2.0 mm, axial acquisition, 68 slices, field-of-view = 208 mm,

voxel size =  $2 \times 2 \times 2$  mm). Studies 1 and 2 had 251 and 311 volumes per run, respectively.

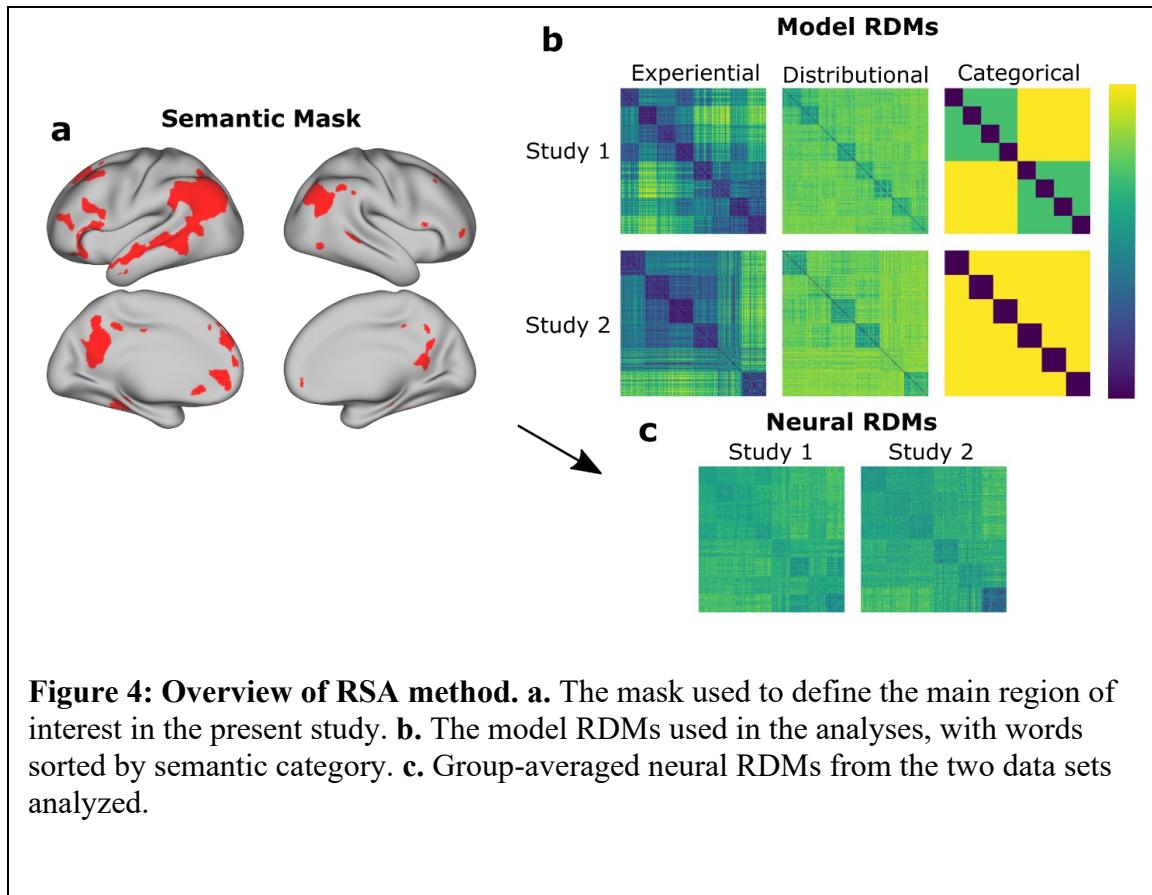
### ***FMRI processing***

MR images were preprocessed via a containerized version of the fMRIprep 22.0.1 pipeline (Esteban et al. 2019). Three field maps were used to estimate a B0-nonuniformity map using ‘topup’ (FSL 6.0.5). The three T1-weighted (T1w) images were corrected for intensity non-uniformity (INU). A T1w reference map was computed by averaging these T1w images after registration using ‘mri\_robust\_template’ (FreeSurfer 7.2.0). The T1w reference was then skull-stripped with a Nipype implementation of the ‘antsBrainExtraction.sh’ workflow. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter, and gray matter was performed on the brain-extracted T1w image using FAST (FSL 6.0.5). Brain surfaces were reconstructed using ‘recon-all’ (FreeSurfer 7.2.0). For each of the EPI runs for each participant, a reference volume and its skull-stripped version were generated using fMRIprep. Head-motion parameters with respect to the EPI reference were estimated before any spatiotemporal filtering using MCFLIRT (FSL 6.0.5). EPI runs were slice-time corrected, and the EPI reference was co-registered to the T1w reference using ‘bbregister’ (FreeSurfer), a boundary-based registration algorithm with six degrees of freedom. Grayordinate files using the fsLR-32k\_midthickness cortical surface model (Glasser et al. 2013) containing 91k samples were also generated using the highest resolution fsaverage as an intermediate standardized surface space. No additional spatial smoothing was applied.

Beta and t values were calculated directly on the surface time-series data using 3dREMLFIT in AFNI (Cox 1996), incorporating 6 motion estimates, their derivatives, CSF and white matter regressors, and 4<sup>th</sup> order baseline polynomials for detrending. We separately z-scored response time estimates across each set of stimulus presentations and included them as nuisance regressors in regression models having at least 2 presentations of the stimuli. In addition, we censored volumes that had a framewise displacement greater than .9mm. Each word, including those with multiple presentations, was its own regressor in every model (i.e., except for reaction time, all models had the same number of regressors). We ensured that no functional data from omitted runs contributed to the estimated beta values calculated from different presentation combinations by running separate regressions for every set of presentation combinations.

### ***ROI selection***

The primary ROI used to generate the neural RDMs was a functionally defined mask based on a previous activation likelihood estimate (ALE) meta-analysis of 120 functional neuroimaging studies of semantic language processing (Binder et al. 2009). The ALE map was projected on the HCP fsLR-32k\_midthickness cortical surface model and binarized at a significance threshold of  $p < .01$ . The ROI included substantial portions of association cortex in frontal, parietal, and temporal cortex (**Figure 4**). We chose to use a pre-defined semantic ROI to minimize bias that could be introduced in the analysis by selecting an ROI through a reliability measure and subsequently performing reliability analysis. However, to ensure the generality of our results, we repeated the analyses in 5 anatomically defined ROIs associated with language processing (not shown here). This



**Figure 4: Overview of RSA method.** **a.** The mask used to define the main region of interest in the present study. **b.** The model RDMs used in the analyses, with words sorted by semantic category. **c.** Group-averaged neural RDMs from the two data sets analyzed.

was done by projecting the Desikan-Killiany atlas to the HCP fsLR-32k\_midthickness cortical surface. The named left hemisphere ROIs we analyzed were ‘superior temporal’, ‘banks superior temporal’, ‘supramarginal’, ‘inferior parietal’, and ‘pars opercularis’.

### Semantic models

To assess variation in the correlation of subsets of the data to a model, we used as references semantic models based on three qualitatively different types of information to verify that the pattern of results did not depend on the choice of semantic model. The models consisted of an experiential model of concept representation (CREA; Binder et al. 2016), a popular distributional model (word2vec; Mikolov et al. 2013), and a categorical model based on taxonomic category membership. Both the experiential and distributional models are described in the introduction to this dissertation. The categorical model was

created to encode the taxonomic structure of the concept sets in both studies. For Study 1 this consisted of two super-categories (events and objects) with 4 sub-categories in each, and for Study 2 the model consisted of 6 categories of concepts (see Figure 4 for visualization).

### ***RSA and Cronbach's alpha***

Cronbach's alpha is a reliability coefficient that quantifies the internal consistency of a measure. Values typically range from zero to one, with higher values indicating a more reliable test. Neural RDMs were generated using Pearson correlation distances of vertex-wise z-scored beta-estimates calculated from the regression. As applied to RSA, each participant can be thought of as a test item, and alpha as capturing the overall internal consistency of the group-averaged neural RDM. This fits well with the RSA approach described in the original paper by Kriegeskorte et al. (2008), where single subject neural RDMs are averaged together, and the resulting average RDM is compared to different models. Using group-averaged RDMs when comparing models has the advantage of testing a model's ability to explain only the variance that is common across all participants. Importantly, Cronbach's alpha is also equivalent both to an ICC for the average of measurements across participants, using the two-way random/mixed effect model ANOVA consistency definition (Bravo and Potvin 1991; McGraw and Wong 1996), and to the average of all possible split-half tau-equivalent reliabilities for a dataset (Warrens 2015). It thus provides an index of consistency, or reproducibility, of the results across independent subsets of participants.

Our analysis used the off-diagonal elements of participants' neural RDMs. Following the steps used by the python RSA toolbox when pooling RDMs, we z-scored each participant's RDM values separately prior to calculating alpha (Nili et al. 2014). We used the python Pinguion toolbox (Vallat 2018) to calculate alpha as a function of the number of stimulus presentations and the number of participants. A random resampling approach was used, in which alpha was calculated 1,000 times for each sample size ranging from 5 to 37 participants, and for 1 to 6 stimulus presentations at each sample size. We also show the relation between average reliability and average RSA correlation for RDMs derived from a particular combination presentation. Specifically, using RDMs derived from the full set of stimuli presentations, we used 1,000 resamples at each sample size ranging from 5 to 37 participants and calculated the average alpha and correlation to three different semantic models. This relation was then plotted and fitted with a simple square-root function having one degree of freedom for scaling.

We also demonstrated how RDM reliability for a given participant sample size can be estimated based on a smaller sample. To accomplish this, using the method proposed by de Vet and colleagues (2017) to estimate the effect on inter-rater reliability of averaging additional raters, we applied the Spearman-Brown (SB) prophecy formula to the ICC two-way random effect model ANOVA single measurement value calculated from all participants in each study.

### ***Repetition effects***

To examine potential repetition suppression effects, the reliability analyses were repeated for each of the six single presentations as well as selected combinations of presentations

(e.g., the first presentations of each of the three scanning sessions). In addition, RSA Spearman correlation values between the neural RDMs and the semantic model RDMs were calculated separately for each stimulus set presentation to estimate differences in data quality between the presentations. The RSA correlation was used instead of reliability because the latter is a group-level point estimate, whereas the former allowed us to test for significant differences through a 2-way repeated measures ANOVA using the individual correlation values for each participant. Further, since the correlation between the neural RDM and a model RDM is usually the value of interest in an RSA study, we felt that an understanding of what presentations contained the most signal of interest would be most informative for researchers using RSA. Post-hoc pairwise t-tests were adjusted for multiple comparisons using the Benjamini-Hochberg false-discovery rate (Benjamini and Hochberg 1995). Correlations with the model were also calculated separately using: 1) only the first presentation from each session (i.e., combination of presentations 1, 3, and 5); 2) only the second presentation from each session; and 3) both presentations from the first session combined with the first presentations from the remaining sessions (i.e., presentations 1, 2, 3, and 5).

### ***Noise ceiling estimates***

A common definition of the noise ceiling is the one popularized by the RSA toolbox (Nili et al. 2014). In this formulation, the lower-bound estimate of the noise ceiling is the average correlation between a single participant's RDM and the average RDM from the remaining participants. This metric is an underestimate of the shared variance and can be viewed in contrast to the upper-bound estimate of the noise ceiling, in which the RDM of

each participant is correlated with the average RDM of all participants including themselves. This provides an overestimate of the shared variance. As the number of participants increases, both the upper and lower noise ceiling bounds asymptotically approach each other. Intuitively, these values must converge because, as the number of participants approaches infinity, the contribution of an individual participant to an average tends toward zero. When the noise ceiling is calculated using the Pearson correlation coefficient, and we use the assumptions of classical measurement theory, we can calculate explicit equations governing the expected values of the upper- and lower-bound noise ceiling estimates. We determined the relationship between a point estimate at one sample size and the expected value of the noise ceiling estimate at all other sample sizes. A more detailed motivation for the formula can be found in the supplemental material of Mazurchuk et. al. (2023). Given N time series each denoted by  $R_n$ , the lower-bound noise ceiling estimate as a function of n subjects follows the form:

$$A = \frac{1}{N} \sum_{n=1}^N Var(R_n)$$

$$B = Var\left(\frac{1}{N} \sum_{n=1}^N R_n\right)$$

$$C = \frac{A - NB}{1 - N}$$

$$D = \frac{A - B}{1 - \frac{1}{N}}$$

$$LowerNoise(n) = \sqrt{\frac{C}{\sqrt{A} * \sqrt{C + \frac{D}{n-1}}}}$$

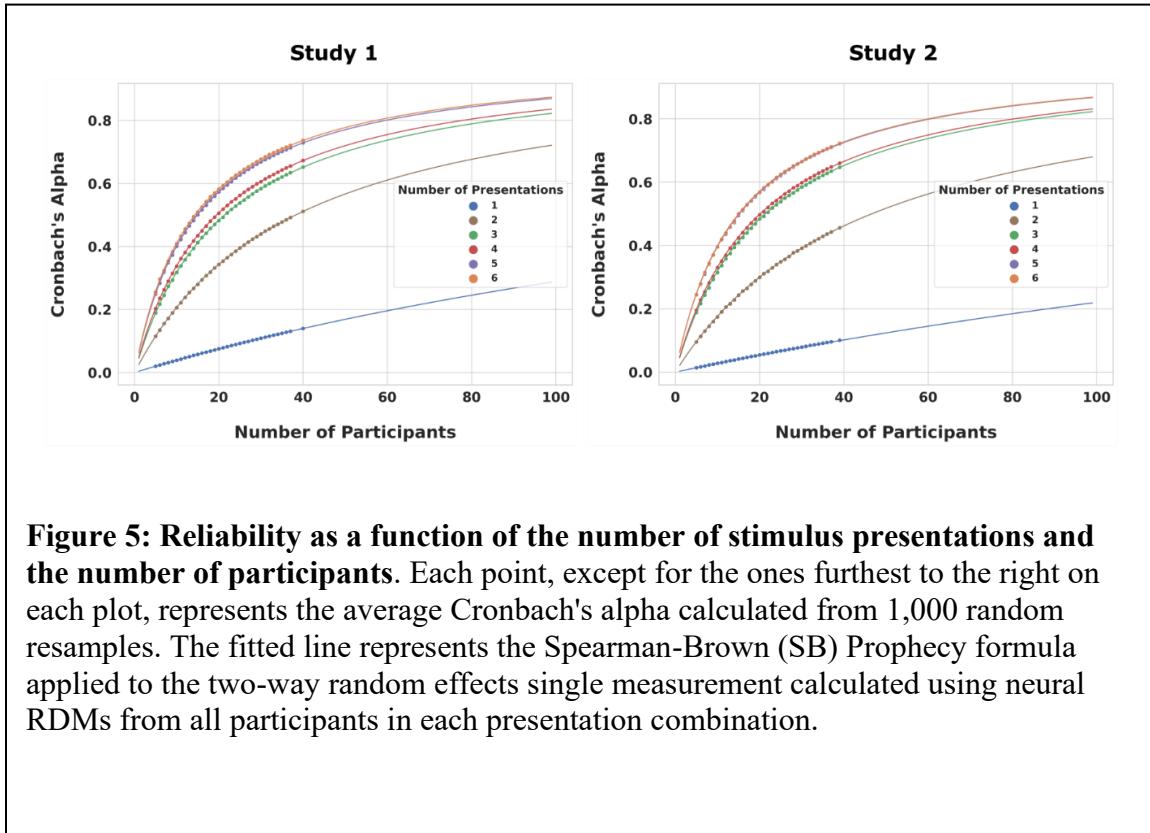
Example code demonstrating how to calculate the expected value of the noise ceiling at different sample sizes can be found in an online GitHub repository ([https://smazurchuk.github.io/rsa\\_reliability/lab/index.html](https://smazurchuk.github.io/rsa_reliability/lab/index.html)). To validate the decomposition of variance implicit in the above derivation, we plotted the predicted curve against point estimates of the noise ceilings at different sample sizes derived from 1,000 random resamples at sample sizes ranging from 5 to 37.

## Results

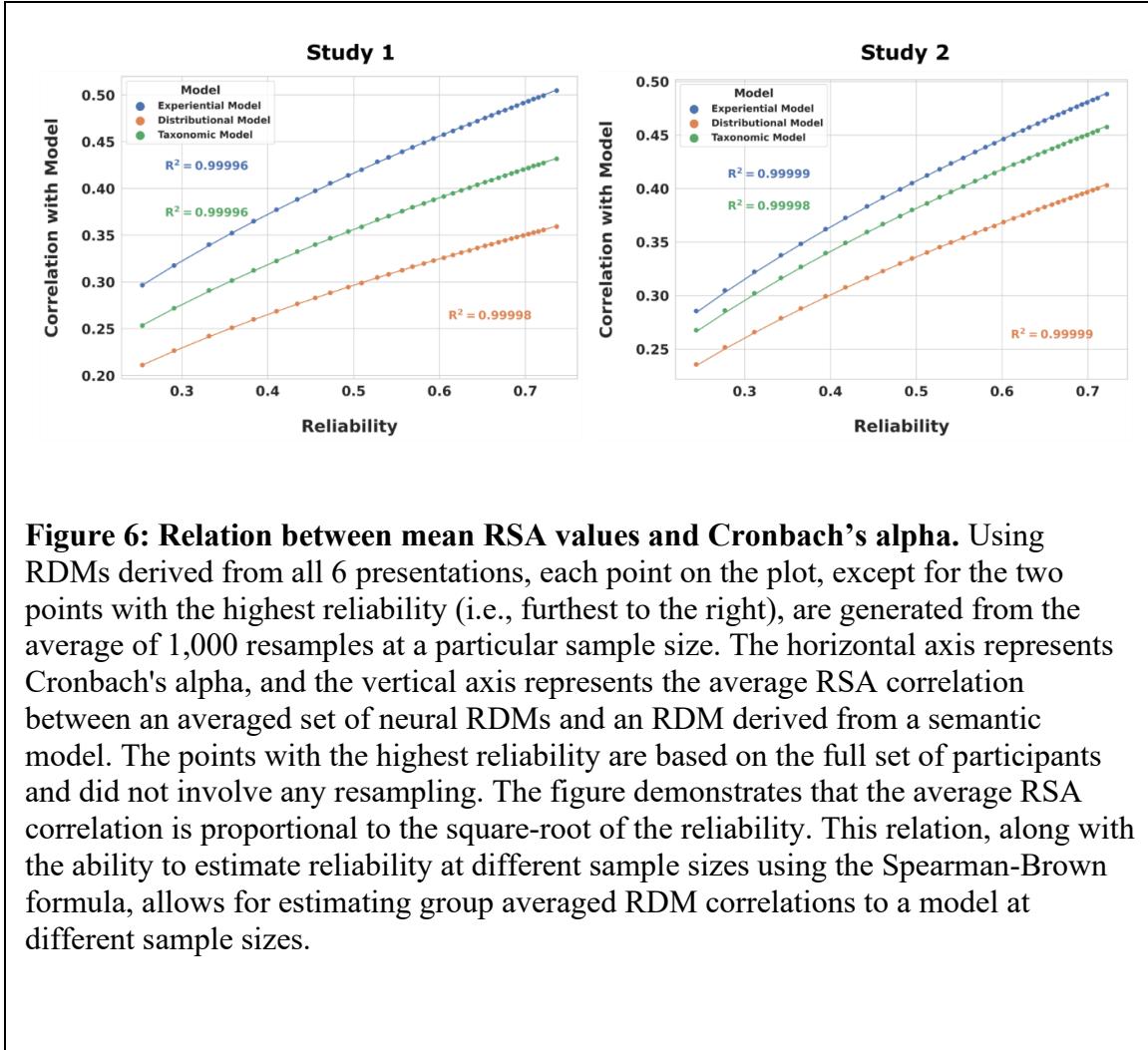
### ***Cronbach's alpha by number of participants***

As shown in **Figure 5**, in both studies, reliability as measured by Cronbach's alpha increased monotonically with participant sample size. Observed values of alpha closely matched those computed using the Spearman-Brown (SB) Prophecy formula applied to ICC values. Visual inspection of the trend indicates that reliability does not begin to plateau until at least 30 participants are present in the group average. Using all 6 presentations in Studies 1 and 2, a sample size of 20 participants yielded alphas of .58 and .56, and using the complete sample yielded alphas of .74 and .72, respectively. To reach alpha = .8 would require around 58 and 60 participants in Studies 1 and 2 with the current task design. When all 6 presentations were included, for each resample where alpha was calculated we also calculated the RSA correlation values between (group averaged) neural RDMs and three semantic models. Alpha and RSA correlations were averaged at each sample size and plotted against each other demonstrating a square-root relation (**Figure 6**).

### Cronbach's alpha by number of stimulus repetitions

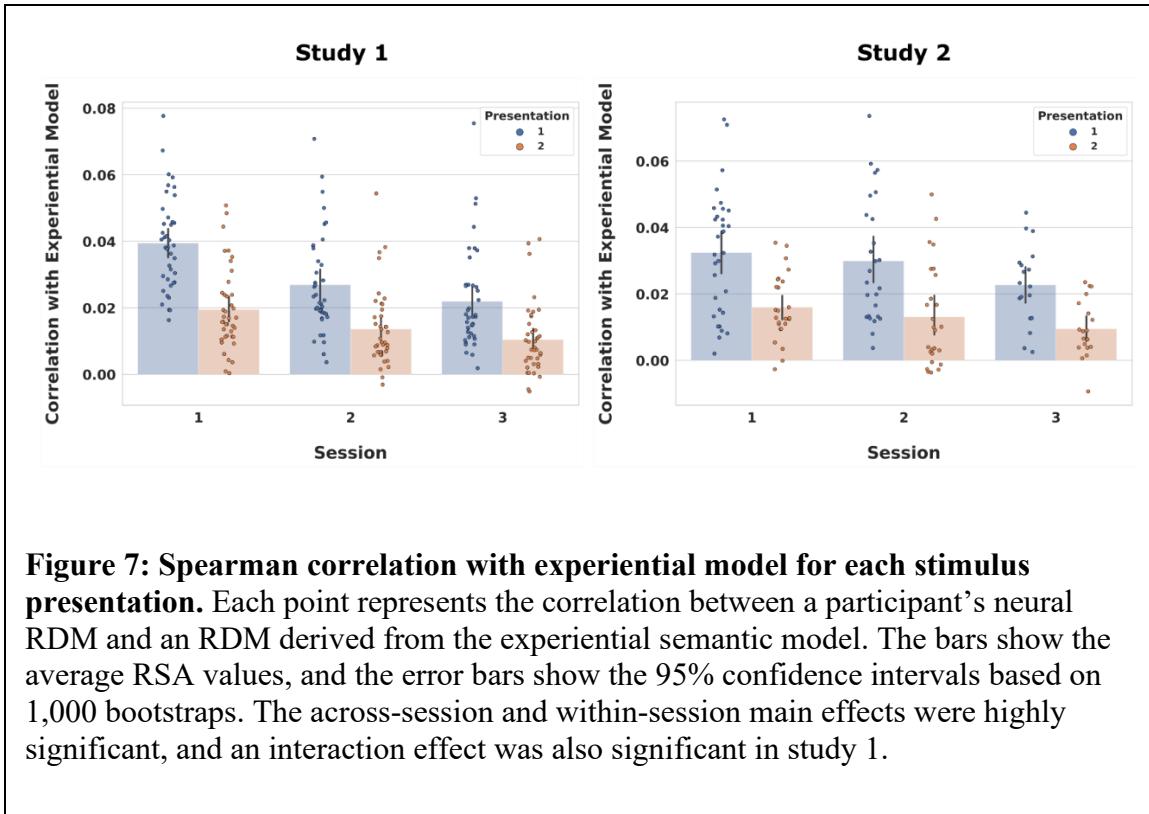


**Figure 5** also illustrates the degree of improvement in reliability with an increasing number of stimulus presentations. Alpha remained low when only 1 presentation was included, even with large participant samples. In both studies, the second presentation substantially improved reliability, with an approximately 5-fold increase in alpha values when using two stimulus presentations compared to just one. Smaller but still substantial improvements were produced by the third presentation and the fifth presentation, with hardly any additional gain from the fourth and sixth presentations. This pattern suggests a gradual decline in the added value of repetitions, particularly repetitions occurring in the same session. This overall pattern of findings was also replicated in 5 other ROIs (see supplemental material of Mazurchuk et. al, (2023) for detail).



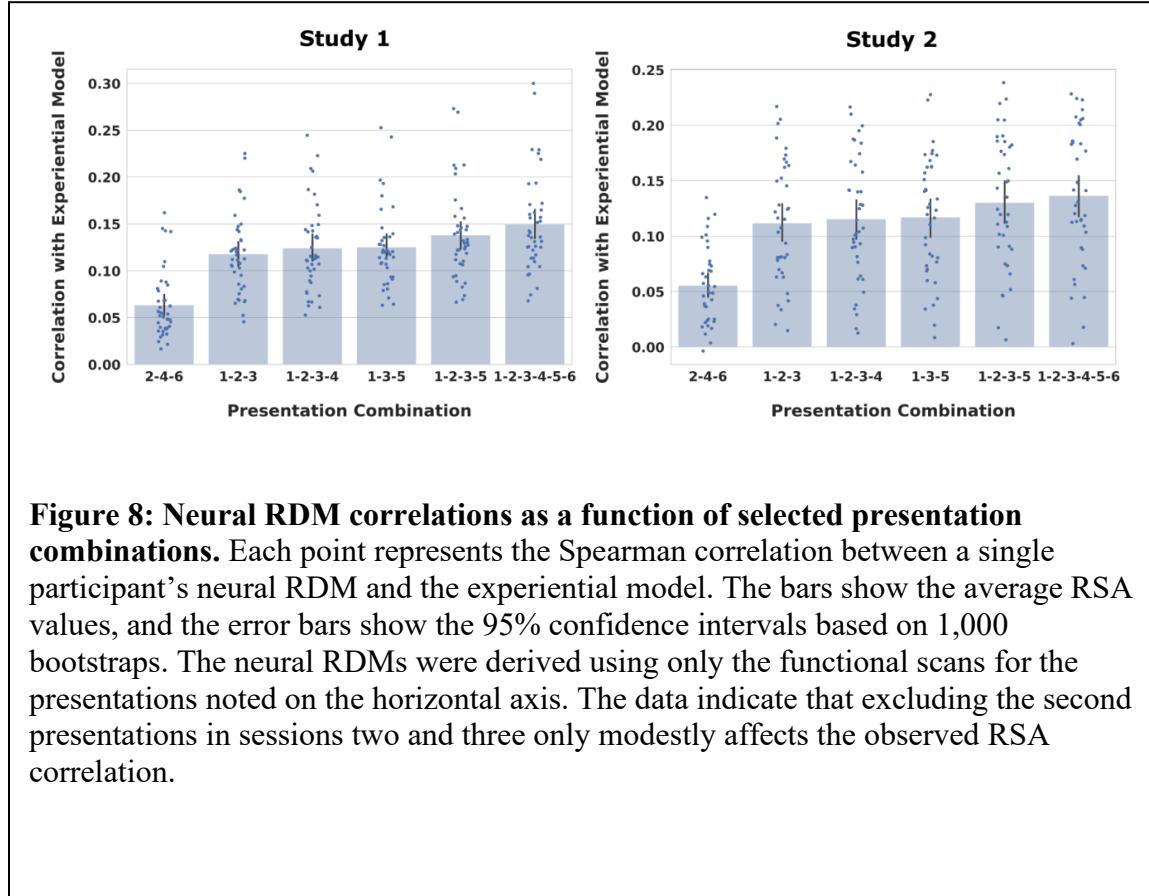
**Figure 6: Relation between mean RSA values and Cronbach's alpha.** Using RDMs derived from all 6 presentations, each point on the plot, except for the two points with the highest reliability (i.e., furthest to the right), are generated from the average of 1,000 resamples at a particular sample size. The horizontal axis represents Cronbach's alpha, and the vertical axis represents the average RSA correlation between an averaged set of neural RDMs and an RDM derived from a semantic model. The points with the highest reliability are based on the full set of participants and did not involve any resampling. The figure demonstrates that the average RSA correlation is proportional to the square-root of the reliability. This relation, along with the ability to estimate reliability at different sample sizes using the Spearman-Brown formula, allows for estimating group averaged RDM correlations to a model at different sample sizes.

The between- and within-session repetition effects are seen more clearly in **Figure 7**, which shows the mean RSA correlation value between the neural RDM and the experiential model RDM for each presentation in each study. In Study 1, both the within-session and across-session main effects were highly significant ( $F(2,39) = 240.82$ ,  $p < .001$ ; and  $F(2,78) = 31.29$ ,  $p < .001$ , respectively). An interaction effect was also significant ( $F(2,78) = 5.56$ ,  $p < .01$ ). An analysis of the simple effects revealed highly significant within-session differences for all three days, with significantly stronger correlations seen with the first presentation relative to the second (corrected  $p < .001$ ). Across sessions, the RSA correlation for the first presentation on the first day was



significantly stronger than the first presentations on each subsequent day (corrected  $p < .001$ ). Similarly, the correlation for the second presentation on the first day was stronger than that seen on the subsequent days (corrected  $p < .001$ ). The declines observed between the second and third days were smaller in magnitude. The decline was still significant for the second presentation across those days (corrected  $p < .05$ ), but only approached significance for the first presentation ( $p < .06$ ). In Study 2, both within and across session main effects were highly significant ( $F(2,38) = 67.70, p < .001$ ; and  $F(2,76) = 13.90, p < .001$ , respectively) with no significant interaction. Follow-up analyses with different models and different ROIs demonstrated the same qualitative effects (see Mazurchuk et. al. (2023) for details). We also analyzed the participant response times and found main effects of stimulus repetition for both within- and across-session repetition in Study 1 ( $F(2,78) = 110.0, p < .001$ ; and  $F(2,39) = 3.58, p < .05$ ) and a main effect of within-

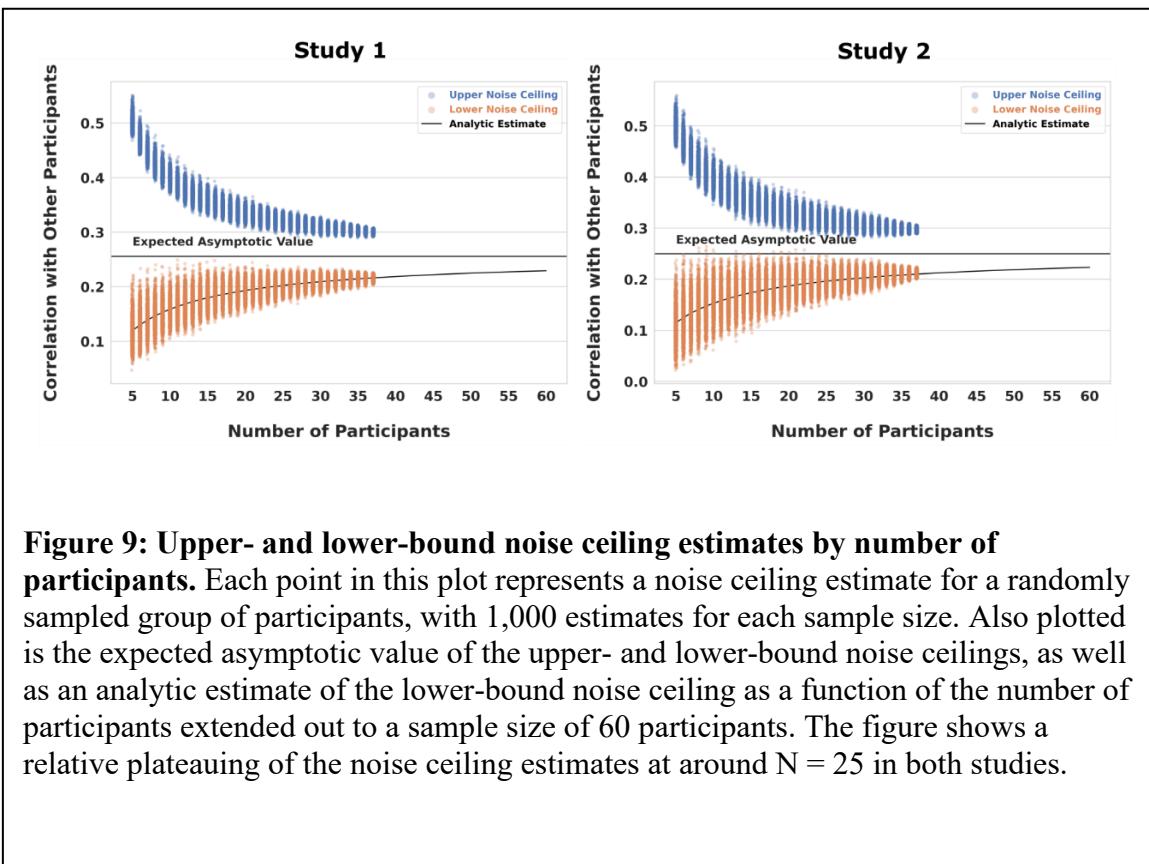
session repetition in Study 2 ( $F(2,38)=130.7$ ,  $p<.001$ ), which mirrored the repetition suppression patterns observed in the fMRI data.



**Figure 8: Neural RDM correlations as a function of selected presentation combinations.** Each point represents the Spearman correlation between a single participant's neural RDM and the experiential model. The bars show the average RSA values, and the error bars show the 95% confidence intervals based on 1,000 bootstraps. The neural RDMs were derived using only the functional scans for the presentations noted on the horizontal axis. The data indicate that excluding the second presentations in sessions two and three only modestly affects the observed RSA correlation.

These effects are illustrated in a different way in **Figure 8**, which shows mean RSA neural-model correlation values for different combinations of presentations. The large effect of within-session repetition is illustrated by the much greater correlations using presentations 1, 3, and 5 (first presentation in each session) compared to presentations 2, 4, and 6. The correlation using the first presentations in each session (1, 3, 5) was slightly but significantly higher than the correlation obtained from presentations 1-3 in Study 1 (mean difference = .010,  $p < .005$ ), with no statistically significant difference in Study 2. Leaving out data from later within-session repetitions (i.e., presentations 4 and 6) resulted in correlation values ( $r = 0.138$  and  $r = .130$ ) that were

nearly as high as with all 6 presentations ( $r = 0.15$  and  $r = 0.136$ ), though this small difference was statistically reliable in both studies (mean Study 1 difference = .01,  $p < 0.00001$ ; mean Study 2 difference = .006,  $p < .05$ ). Similar patterns of results for presentation combinations, with some small variation in relative order of combinations (1-2-3-4) and (1-3-5), were found in the analysis of 5 other ROIs. The overall pattern of results seen in Figure 5 were also observed when using reliability point estimates instead of RSA correlations.



**Figure 9: Upper- and lower-bound noise ceiling estimates by number of participants.** Each point in this plot represents a noise ceiling estimate for a randomly sampled group of participants, with 1,000 estimates for each sample size. Also plotted is the expected asymptotic value of the upper- and lower-bound noise ceilings, as well as an analytic estimate of the lower-bound noise ceiling as a function of the number of participants extended out to a sample size of 60 participants. The figure shows a relative plateauing of the noise ceiling estimates at around  $N = 25$  in both studies.

### *Noise ceiling by number of participants*

Analysis of the upper and lower bound estimates of the noise ceiling showed asymptotic behavior (Figure 9). For Study 1, with 10 participants included in the analysis, the mean lower noise ceiling had a value of 0.16; with 20 participants, the value increased to 0.19;

and with all participants, the value was 0.22. For Study 2, with 10 participants included in the analysis, the mean lower noise ceiling had a value of 0.15; with 20 participants, the value increased to 0.18; and with all participants, the value was 0.21. With all participants included in the analyses, the differences between the upper and lower noise ceilings were .08 in both studies. The resampled lower noise ceiling was well fit by our model equation. Extrapolation of these curves shows that they converge to 95% of their asymptotic value at sample sizes of 134 and 140 for Studies 1 and 2, respectively.

## Discussion

Our findings demonstrate significant within- and across-session repetition suppression effects on the RSA correlation between neural and model RDMs. Thus, these results highlight the importance of another design consideration beyond simply the number of repetitions, namely, whether the stimulus repetitions occur within the same fMRI scanning session or across separate days. While increasing the number of stimulus presentations generally improved the reliability of the group-averaged RDM, there were substantial adaptation, or repetition suppression, effects, which were most pronounced for within-session repetitions, with some release from adaptation observed across days. Additional presentations of the stimulus within the same fMRI scanning session appear to produce minimal benefit after the first day.

The magnitude and duration of the repetition suppression effect are notable. Repetition suppression is rarely mentioned as a potential problem in RSA studies that employ stimulus repetition, and it has likely been underestimated in MVPA studies more broadly. For example, Zhang et al. (2022) stated that “repetition suppression effects in

fMRI are short-lived, dissipating on the order of seconds, and are strongest when few other stimuli are presented between repetitions [...]." Their study used 32 stimuli presented 6 times on the same day; we found a large within-session repetition suppression effect even with 320 stimuli presented only twice on each day. Further, diminishing returns across scanning sessions were notable even with an average interval between the first and last session of approximately 45 days. Similar repetition suppression effects were seen across two separate data sets using different stimuli and participants as well as across multiple ROIs. Importantly, such effects are likely to be much larger in studies that use smaller stimulus sets or more than two repetitions in a single day.

In our condition-rich, item-level RSA design, we found limits and tradeoffs regarding sample size and number of stimulus presentations analogous to those previously reported in univariate fMRI (Nee 2019). Specifically, very large sample sizes would be required to achieve adequate reliability with fewer than three stimulus presentations. In fact, increasing sample size was of minimal benefit with only one presentation, as reliability remained low even with sample size projections over 100. Stimulus repetition, therefore, is a useful strategy for enhancing reliability and can be "traded" for a smaller sample size. In our data, for example, the alpha obtained with 40 participants and 2 presentations (~.5) could have been reached with half as many participants and a third presentation on a different day. However, beyond the third presentation, there were diminishing returns to adding more repetitions, particularly within a session. While the highest alphas and RSA correlations were seen with the full 6 presentations, the results suggest that for these two data sets, it would have been possible

to obtain nearly indistinguishable results with only 4 of the 6 presentations by eliminating a second within-session presentation of the stimuli after the first session, thus substantially reducing total scanning time.

To be clear, the sample size curves presented here are not intended to provide hard recommendations for sample size. The absolute values of alpha obtained in any given study likely depend on several factors, including the number and type of stimuli used, the task performed, and the image analysis methods. The overall pattern of sample size and repetition effects, however, was observed to generalize across two studies using different stimuli and participants, and across ROIs differing in size and location, thus we believe that these patterns are likely to generalize to other studies using an item-level RSA approach. As mentioned earlier, the repetition suppression effects we observed are likely to be even larger in studies using smaller stimulus sets and more closely spaced repetitions. Furthermore, it seems likely that the advantages of having at least 3 stimulus presentations and the general pattern of trade-offs between sample size and stimulus presentations would apply even to studies using other tasks and stimulus types.

Because it is not possible to make universal recommendations for the amount of individual-level data and sample size that would be appropriate to all experimental designs, Nee (2019) suggested that researchers examine these factors in their own pilot data and estimate the optimal amount of data and sample size from those data. The current findings provide support for the application of ICCs and the Spearman-Brown prophecy formula to the group-averaged neural RDM at intermediate points in data collection in order to estimate the potential gains in reliability to be expected by increasing sample size given the obtained level of interindividual consistency. This usage

was proposed by de Vet et al. (2017) as a method to estimate the number of raters needed to attain an adequate level of interrater reliability. The current findings extend that usage to estimates of sample size for studies using neural RDMs. We also derived a formula relating how the upper- and lower-bound estimates of the noise ceiling change as a function of the number of participants.

We chose to assess the reliability of the group-averaged neural RDM because it is a function of both between-participant reliability and the number of participants. While individual participant neural RDMs are typically used for statistical inference in RSA, the power to detect a significant correlation depends on both the reliability of the individual RDMs, captured in the ICC two-way random effects, single measurement value, and the number of participants, captured in the Spearman-Brown prophecy formula applied to that ICC value. Beyond this, the correlation between a model RDM and a group averaged neural RDM is more closely related to how well a model accounts for ‘shared’ or ‘explainable’ variance.

The calculation of alpha or ICCs may provide additional insights regarding reliability and can be interpreted within a broader context. However, we note that the upper and lower noise ceilings have the benefit of being applicable to many distance metrics (e.g., Pearson, Spearman, Kendal-tau-a, Jaccard), whereas classical reliability measures are often not applicable when considering measures other than Pearson correlations.

There was relatively slow convergence of upper and lower noise ceilings within the range of sample sizes typically used in fMRI. We do not believe that the persistence of a sizable difference between the upper and lower bound estimates of the noise ceiling

is particular to our data, and it is likely the case that additional participants would only marginally reduce this difference at common sample sizes. Although not explicitly reported in the results (because resampling variability is difficult to assess when subsampling from a finite pool of participants), we note that the variability in the lower noise ceiling estimate was high when fewer than 10 participants were used in its calculation. However, this large variability is also compounded by the fact that the non-independence of subsets means that the variance observed at any given sample size is an underestimate of the variance that would be found if a larger dataset was used.

One tacit assumption throughout this paper has been that the observed changes in reliability are driven by the information of interest (in our case, semantic information), and that higher reliability is desirable. However, as sometimes occurs in fMRI, very reliable effects could simply reflect nuisance factors that are consistent across participants, such as the blocking structure of trials (Cai et al. 2019). While there are almost certainly sources of shared variance in our study that are not related to the (semantic) information of interest, the similarity between the repetition suppression patterns found in the reliability curves and in the analyses using semantic models indicates that the observed changes in reliability likely reflect changes in the amount of semantic information. This is likely to be the case only when nuisance variables (such as the blocking structure of trials) are reasonably well controlled.

In closing, while much work remains to be done to thoroughly characterize the factors that affect the reliability of neural RDMs, the present analyses provide some initial steps in this process, and we hope they will encourage the reporting of intermediate reliability statistics to aid in comparisons across studies. In the case of RSA, this amounts

to reporting a standard measure of reliability for neural RDMs along with the correlation measured.

## Chapter 3: A Shared Representational Basis Across Four Dissociable Semantic Categories

### Introduction

As reviewed in the introduction to this dissertation, reports of category-related semantic deficits (CRSDs), where knowledge for one category of items is impaired over and above impairment for other categories, has been taken as evidence that the brain's conceptual system might be organized by semantic domain. This finding of category-related organization has also been extensively corroborated by many fMRI studies investigating semantic representations (Mahon and Caramazza 2009). Elaborated on in the introduction, analysis of CRSDs has led to several explanations, such as innate domain-specific constraints for certain categories of concepts (Mahon and Caramazza 2011), to grounded theories of cognition, where semantic knowledge is represented in a distributed fashion across different association cortices. However, while CRSDs initially provided influential motivation for grounded theories of cognition, the ability of any particular grounded model to account for the organization of the prominent categories in the CRSD literature has hitherto not been explicitly tested. We address this knowledge gap in an fMRI study that focuses on the four dominant categories reported in the CRSD literature, and specifically test the ability of different semantic models to encode information that generalizes across categorical boundaries.

FMRI studies examining the categorical organization of the semantic system have varied in the categories included, with some categories being highly represented (e.g., animals and artifacts; Chouinard and Goodale 2010) and other categories being less

represented, despite considerable evidence of categorical dissociation (e.g., body parts; Laiacona et al. 2006; Sacchett and Humphreys 1992; Schwoebel and Coslett 2005). Further, although recent emphasis in fMRI experiments has been on the use of naturalistic stimuli (e.g., Huth et al. 2016), this also presents a problem for analysis specifically aimed at examining category related organization as natural stimuli are difficult to match on both number of exemplars from each category, as well as lexical properties across compared categories. Although some category dissociations are well-evidenced (e.g., animals and artifacts), differences in the other categories included in studies makes performing non-standard pairwise comparisons (e.g., animals and body parts) difficult. We address these limitations in our study by including matched stimuli from the four categories (animals, artifacts, plants/foods, and body parts) that were found to be “the categories of category specific deficits” in a seminal review by Capitani et. al. (2003)<sup>11</sup> (see **Table 4**, Chapter 2). Comparing all categories in a similar fashion allows us to draw attention to each of the 6 unique pair-wise contrasts implied by the CRSD literature, and quantitatively assess the magnitude of categorical organization across the cortex. We examine the degree of category organization across the cortex through univariate contrasts as well as multivariate cluster analysis.

Another contribution of our study is the choice to use visually presented word stimuli to study category organization. Many papers on category representation use picture stimuli, which have the potential to introduce perceptual confounds such as differences in spatial frequency, and contrast differences, among others. Of three ALE

---

<sup>11</sup> Although Capitani and colleagues only refer to the categories of “animate objects, inanimate biological objects, and artefacts” they also state “the category of body parts dissociates from animals and fruits and vegetables, and is most often impaired together with artefacts”.

meta-analyses that examined category-related activation for living and nonliving concepts, one used only picture-stimuli (Chouinard and Goodale 2010), one did not specify the stimuli (Chen et al. 2017), and one that stratified analysis based on stimulus modality (visual vs auditory) found that activation patterns “were heavily influenced by methodological factors including [...] stimuli modality” (Derderian et al. 2021). While Derderian et al. (2021) did not stratify analysis within the vision modality (i.e., both pictures and text were considered together), one earlier meta-analysis that included 10 studies that used only word stimuli found “‘Living’ foci showed no significant overlap” (Binder et al. 2009). Given that many fMRI studies have found category effects in visual processing regions, it is unknown to what extent perceptual differences could confound findings of apparent categorical organization. Using word stimuli reduces the likelihood that any observed activation differences are the result of physical confounds present in the stimuli indicating that activation differences are more likely to be truly semantic in nature.

Another contribution is not just in the focused choice of categories, but also in the stratification of analysis. Although several papers have shown the ability of different semantic models to decode stimuli from fMRI activation patterns, the analysis has not been typically stratified by category to focus on across category generalization. For example, in a study where stimuli come from four categories, a taxonomic model might decode stimuli far above chance if trained and tested using the four categories, but such models have no across-category generalization. That is, the model would be expected to perform at chance accuracy if the training only consisted of stimuli from 2 categories, and testing was performed on the 2 left-out categories. Stratifying analysis is essential for

testing hypotheses related to a shared representational basis across semantic domains as it could be that entirely different feature sets represent items in different domains. To specifically test for a *shared* representational basis, our analysis stratifies all training and test sets by the relevant categories reported in the CSRD literature.

In principle, a shared representational basis could be in accord with several models. For example, a shared representational basis does not entail that the basis is composed of experiential or interpretable features. As discussed in the introduction, distributional models are language models derived from a large corpus of text and are capable of creating vectors that capture the location of a concept in high-dimensional semantic space. While the dimensions of these models are, in general, not directly interpretable, they have repeatedly demonstrated the ability to predict fMRI activation patterns. Since we are interested in whether apparent category selectivity derives from different “loadings” on a shared representational basis, both distributional and experiential models can bear on this hypothesis. The primary class of models that do not evidence a shared representational basis are categorical models.

As reported by Fernandino et. al., and further corroborated in our lab, we find that popular distributional models perform similarly in RSA of fMRI data, so we used Word2Vec (W2V; Mikolov et al. 2013) as a representative model from the distributional class. We tested not only whether the W2V model generalized across categories, but through variance partitioning, whether it captured unique information that generalized across categories.

The ability to predict category difference maps, as well as individual stimuli, would be supporting evidence for the hypothesis that category related organization can be

explained using linearly weighted feature maps with a shared neural code across all categories.

## Materials and Methods

Participants: Participants are largely overlapping with those in study the same as those listed in “Study 2” from in Chapter 2 of this dissertation except that one additional participant was removed from analysis, resulting in the analyzed dataset the 38 participants indicated in Chapter 2 of this dissertation (see Table 2 for details).

Stimuli and Task: The complete stimulus set used in the fMRI study is the same as described in Chapter 2 (Table 4) and described in Tong et al. (2022). For review, it consisted of 300 English nouns from 6 categories: animals, manipulable artifacts (hereafter “tools”), plants/foods, body parts, human traits, and quantities, with 50 words in each category. The scanning protocol followed that outlined in Chapter 2 of this dissertation.

Semantic Models: The model-based analyses used experiential (CREA) and distributional (Word2Vec) models of semantic content. These models are discussed in the introduction to this dissertation.

Acquisition Parameters: Acquisition parameters were the same as outlined in Chapter 2 of this dissertation.

fMRI First Level Model: Beta and t values were calculated directly on the surface time-series as described in Chapter 2.

Analysis: Analysis is divided into model independent and model dependent sections.

Model independent analysis examined the 6 univariate contrasts and the intrinsic degree of clustering present in the multivariate patterns. Model dependent analysis used vertex-wise encoding models to predict group contrasts and activation patterns for individual stimuli at the participant level.

Model-Free Univariate: Univariate pairwise comparison maps were generated from a vertex-wise contrast where average activation (beta-statistic) for a category was contrasted between two categories for each participant. Vertex differences were then tested against zero using a one-sided Wilcoxon signed-rank test and resulting p-values were FDR corrected and thresholded at  $p < .01$ .

Model-Free Multivariate: While dissociations have been reported for all 4 of the target categories, the patterns for which categories are lost vs. preserved have varied. For example, Capitani et al. (2003) writes that body parts are “most often impaired together with artefacts.” For this reason, it is not known whether categories are equally separable. The degree of intrinsic clustering of the neural data was examined both at the individual and group level. For all analyses described here, neural data were derived from t-statistic maps for individual word-stimuli taken from within a large semantic mask. The semantic mask was generated by taking the default, frontoparietal, limbic, ventral attention and dorsal attention networks derived from the Yeo 2011 resting state parcellation (**Figure 11b**). This mask was chosen as it provides a liberal estimation of the areas involved in semantic representation and is broad enough to capture cortex-wide patterns.

The degree of multivariate pattern clustering within each participant was quantified using the pseudo F-statistic calculated from permutational multivariate

analysis of variance (PERMANOVA; Anderson 2001). The pseudo F-statistic is a ratio of between-cluster dispersion to within-cluster dispersion, and significance is determined by permuting the labels. We used Pearson correlation distances between activation patterns, instead of Euclidean distances, so that dissimilarity matrices could be averaged across participants, and because correlation coefficients are not sensitive to mean differences between patterns. Using correlations emphasizes pattern similarity, without considering any average activation difference between categories of stimuli. Pseudo F-score significance was determined by permuting the labels 10,000 times to estimate the null distribution in each participant.

To qualitatively visualize the degree of intrinsic clustering present in the data, the intrinsic correlational structure of the data was visualized using a non-linear dimensionality reduction technique called uniform manifold approximation (UMAP; McInnes et al. 2018). Some non-linear dimensionality reduction techniques (e.g., UMAP, tSNE, MDS) are commonly used to examine whether there are “natural” clusters in a dataset based on the pairwise distances between observations. Along with the neural data (~38,000 dimensional), two dimensional representations of the experiential ratings (CREA, 65 dimensional) and distributional vectors (Word2Vec, 300 dimensional) were created. For each participant, pairwise distances between stimuli were calculated using the Pearson correlation distance, and these distances were then averaged across participants prior to visualization with UMAP.

*Group-level encoding:* We explicitly tested the hypothesis that group-level category difference maps could be explained as the result of weighted combinations of feature sensitivity maps. As this analysis was focused on explaining the group-level effects, and

not examining individual-level variability, to keep permutation testing of the results computationally tractable, beta-statistic maps for each stimulus were first averaged across participants, so that training and testing used only the 250 group-averaged beta-statistic maps. Feature sensitivity maps represent the estimated sensitivity (i.e., responsivity) of each vertex to a specific feature. Feature sensitivity was estimated using vertex-wise encoding models on group-averaged word beta maps. Specifically, we used 10-fold cross-validated ridge regression, as implemented in the Himalaya toolbox (Tour et al. 2022) to assign feature sensitivity weights to each vertex using stimuli that were not members of either of the two target categories. Once fitted, the model was used to predict the activation pattern for the stimuli from the left-out target categories. These predicted maps were then averaged and contrasted. The resulting predicted difference map was correlated with the observed difference map using Spearman's Rho. For example, the 'Animal - Artifact' map was predicted by estimating feature weights using the 150 group-averaged word maps in the 'Plant/Food', 'Body Part', and 'Human Trait' categories. Significance was determined by permuting the labels of the training t-statistic maps and refitting the encoding model and calculating the Spearman correlation 10,000 times.

While Spearman rho is a measure of effect size, the null distributions for each contrast had a unique distribution. A standardized t-distribution was fit to each null distribution to account for the non-normality of the null correlation values. To provide a more intuitive measure of effect-size, we estimated the corresponding Z-values for each Spearman rho through the fitted t-distribution. Specifically, we analytically converted the observed t-value to a probability and then converted that probability to a corresponding Z-value.

*Individual-stimuli encoding:* We also tested the ability of the encoding models to predict participant-specific word activation maps. For each participant and for each of the 4 target categories, vertex-wise encoding models were fit using stimuli outside the target category. Weighted combinations of the feature maps were then used to predict *individual* stimuli from the left-out category in each participant. The predicted maps were correlated with the observed maps, and the average rank of the true target map was calculated, with this average rank then again averaged across participants. An average rank of the true map above the 50<sup>th</sup> percentile indicates that in a forced-choice paradigm between the true map and a random map from the same category, the true map is selected more than 50% of the time. In each participant, word-map labels were shuffled 1,000 times to create a null distribution of percentiles. Within each shuffle, the percentiles were then averaged together across participants to generate the group-level null distribution.

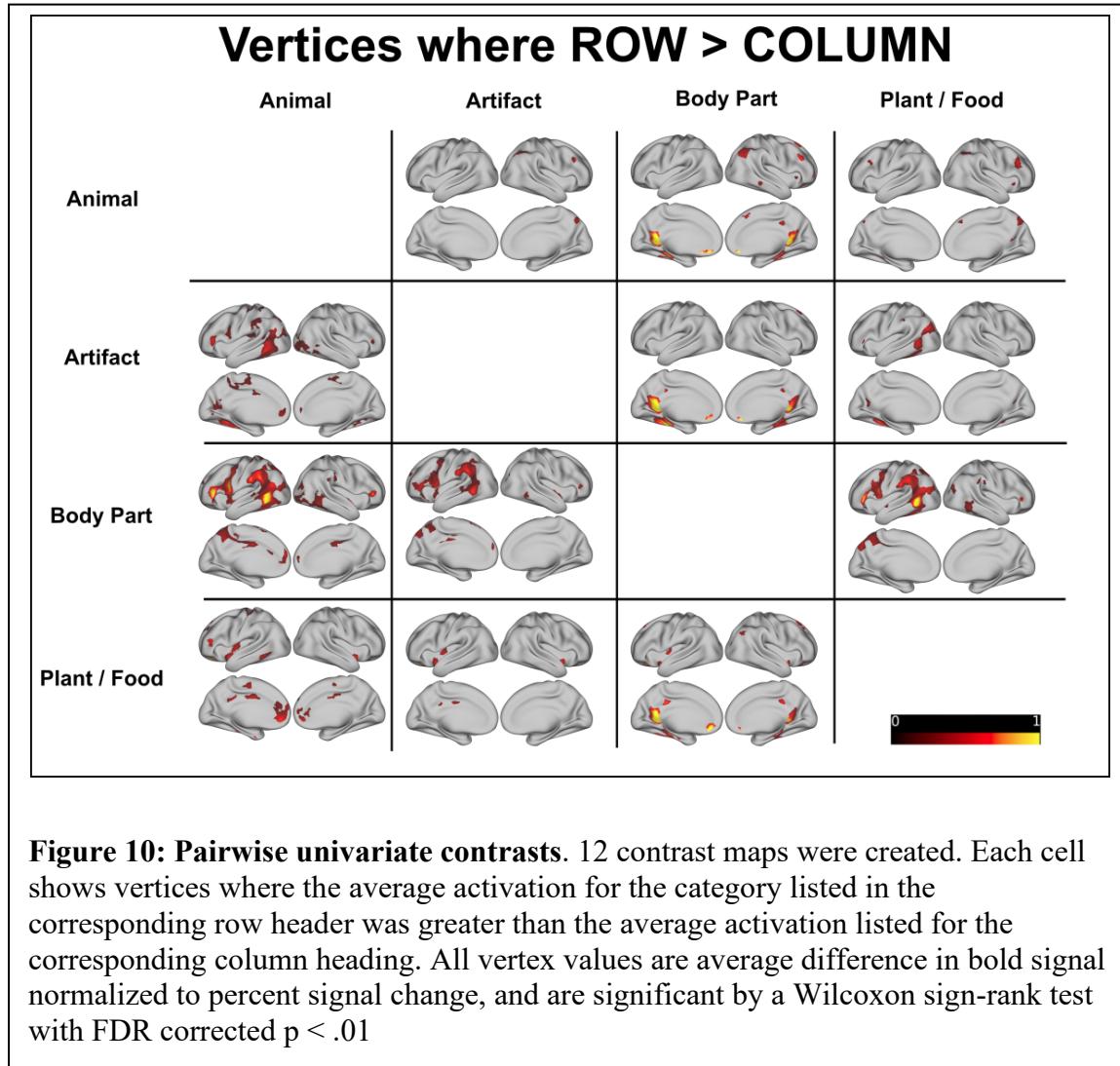
*Variance Partitioning:* Lastly, we were interested in whether the semantic models captured unique information that generalized across categories. For example, it could be that both models predict stimuli above chance to a similar level, but they could be doing so using different information. One way to assess a model's unique contribution is through the use of variance partitioning (LeBel et al. 2021; Lescroart et al. 2015). In this approach, a model is trained on a training set, and an out-of-sample (OOS)  $r^2$  is estimated on held-out data. We trained three models, (just CREA, just W2V, and both) on each of the folds and computed a signed square of the correlation coefficient between predicted and observed values on the left-out category. The unique contribution of each model was determined by set theoretic relations. If A =  $r^2$  of a regression model with CREA, B =  $r^2$

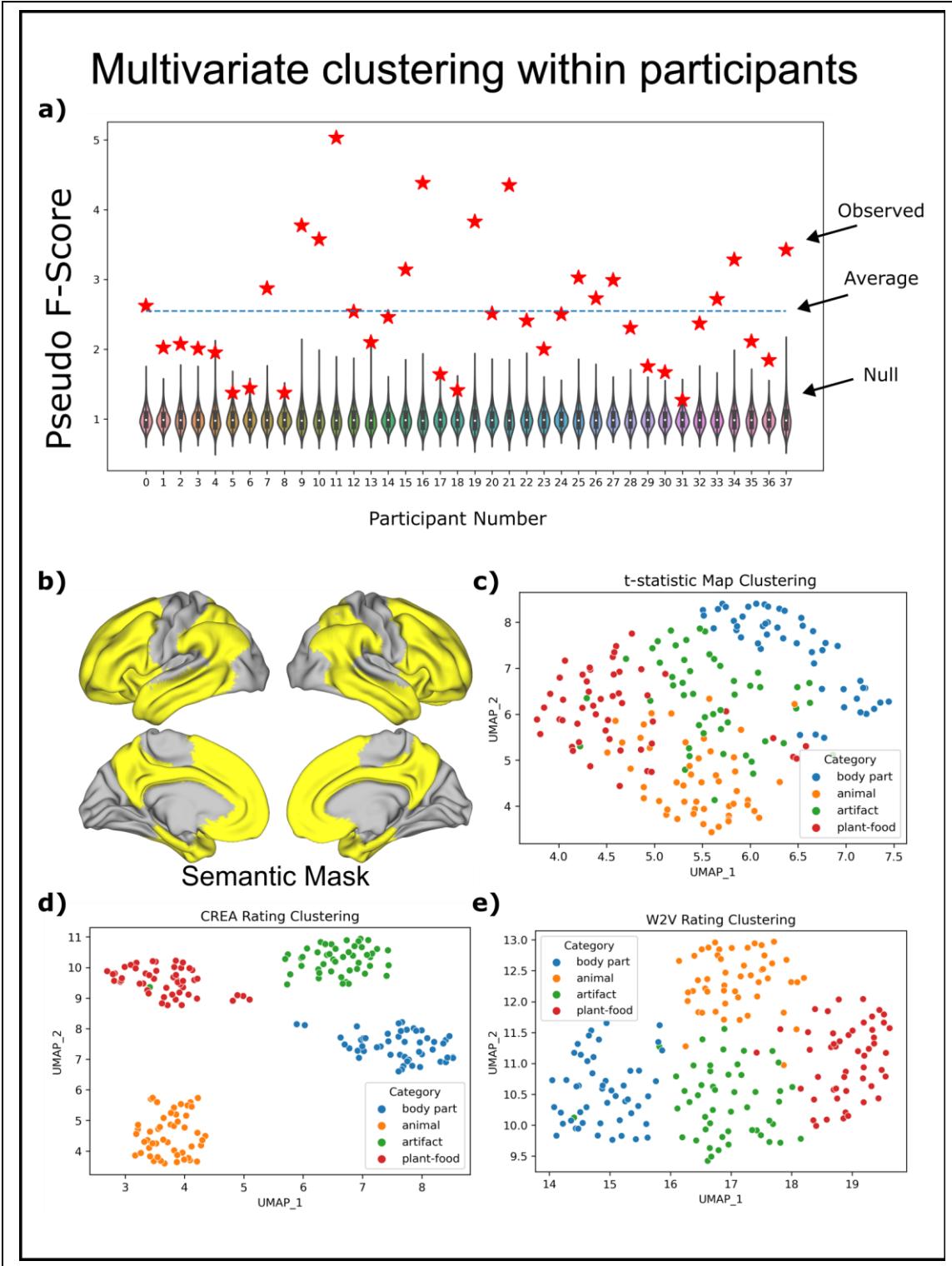
of a regression model with W2V, and  $C = r^2$  of a model with both CREA and W2V, then the unique CREA = (C-B), unique W2V = (C-A), and Shared = (A + B) – C.

## Results

Pairwise contrasts indicate that all 12 one-sided category comparisons have vertices that are significantly more activated for words of one category compared to words from a different category (**Figure 10**). The animal category consistently showed greater activation in the right intraparietal sulcus (IPS) across all three category comparisons. The animal category also had a smaller focus in the right middle frontal gyrus (MFG) across comparisons. Artifacts showed greater activation in the left collateral sulcus into the fusiform gyrus for all three contrasts. The body part category showed greater activation in several large regions across all three category comparisons. Namely, activation was greater for body parts in the left posterior middle temporal gyrus (pMTG), supramarginal gyrus (SMG), inferior frontal gyrus (IFG), precentral sulcus, and dorsal precuneus. Lastly plants/food consistently activated the left insula and orbital frontal

region across all three contrasts.





**Figure 11: Multivariate clustering of data.** Top: A pseudo F-score was calculated for each participant's neural data based on the 4 categories of the stimuli. The pseudo F-score is a ratio of between cluster dispersion over within cluster dispersion, and a null distribution of F-scores was calculated by shuffling the category labels 10,000

times for each participant. The range of each violin plot indicates the largest and smallest F-value observed in any of the permutations. Pearson correlation distances were used to quantify distances between activation patterns. **Bottom:** A dimensionality reduction technique (UMAP) is applied to semantic-network derived activation-patterns, experiential ratings, and distributional representations. To visualize the neural data, neural RDMs were created for each participant using Pearson correlation distances. These neural RDMs were then averaged prior to visualizing the pairwise distances using UMAP. The two-dimensional projections highlight the intrinsic correlational structure of the data.

As shown in **Figure 11a** and **11c**, evidence for category-related organization was also found through PERMANOVA and was visible through UMAP. Both semantic models also demonstrated clustering along categorical lines (**Figure 11d** and **11e**).

Shown in **Figure 12 (top)**, all 6 unique category-difference maps were significantly predicted by the experiential encoding model. Four of the six category difference maps were significantly predicted by the distributional model. Although both models had significant predictions of the held-out difference, the pattern of generalizability was different between the models. For example, the plant/food minus animal contrast had the second highest equivalent Z-value of the predicted experiential maps ( $Z = 4.01$ ), but the lowest equivalent Z-value of the predicted distributional maps ( $Z = .51$ ). For each of the contrasts, variation was also observed in dispersion of the null distributions across models and categories. For example, while the Spearman rho value of 0.165 corresponded to a Z-value of 2.48 for the CREA artifact minus body part comparison, the Spearman rho of .176 for the W2V animal minus body part comparison only corresponded to a Z-value of 2.27.

Semantic models performed similarly well in predicting individual stimuli in individual participants (**Figure 13**). The largest difference in performance between the

two models occurred in the plant/food category, where the experiential model outperformed the distributional model. For all left out categories, in contrast to the pairwise category comparison, the null distributions were visibly similar for both semantic models tested.

### Vertex-wise encoding prediction of group contrast

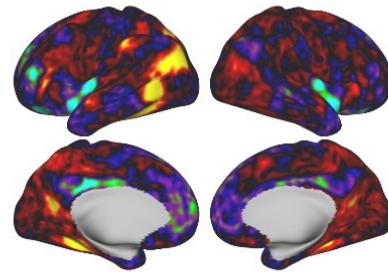
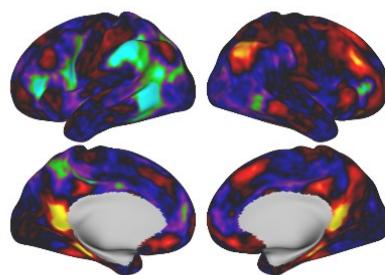
| CREA         |          |               |                 |                 | Word2Vec |        |                |                 |                                 |
|--------------|----------|---------------|-----------------|-----------------|----------|--------|----------------|-----------------|---------------------------------|
|              | Animal   | Artifact      | Body Part       | Plant / Food    |          | Animal | Artifact       | Body Part       | Plant / Food                    |
| Animal       |          | .201 (p=.002) | .428 (p < .001) | .291 (p < .001) |          |        | .220 (p =.004) | .176 (p = .013) | .034 (p = .304)                 |
| Artifact     | Z = 2.83 |               | .165 (p = .007) | .214 (p < .001) |          |        | Z = 2.74       |                 | .217 (p = .005) .265 (p < .001) |
| Body Part    | Z = 4.38 | Z = 2.48      |                 | .184 (p = .002) |          |        | Z = 2.27       | Z = 2.57        |                                 |
| Plant / Food | Z = 4.01 | Z = 3.67      | Z = 2.78        |                 |          |        | Z = 0.51       | Z = 4.11        | Z = 4.49                        |

### Observed and predicted category contrasts

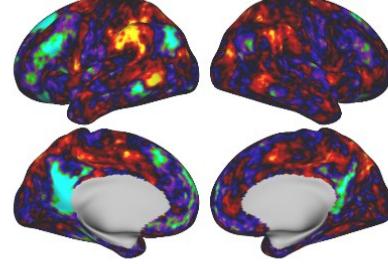
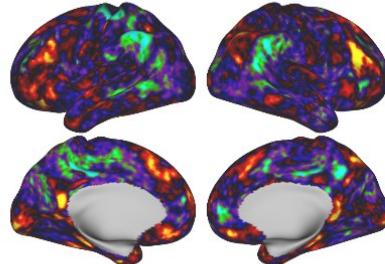
Animal &gt; Body part

Artifact &gt; Plant/Food

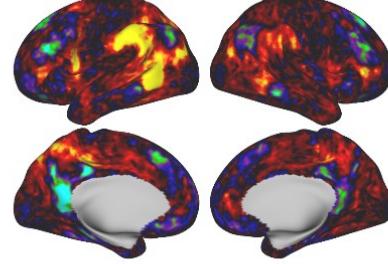
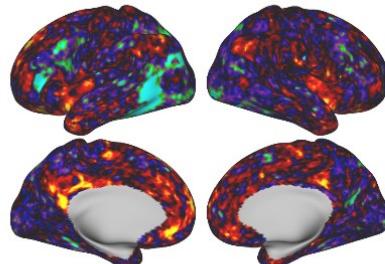
Observed



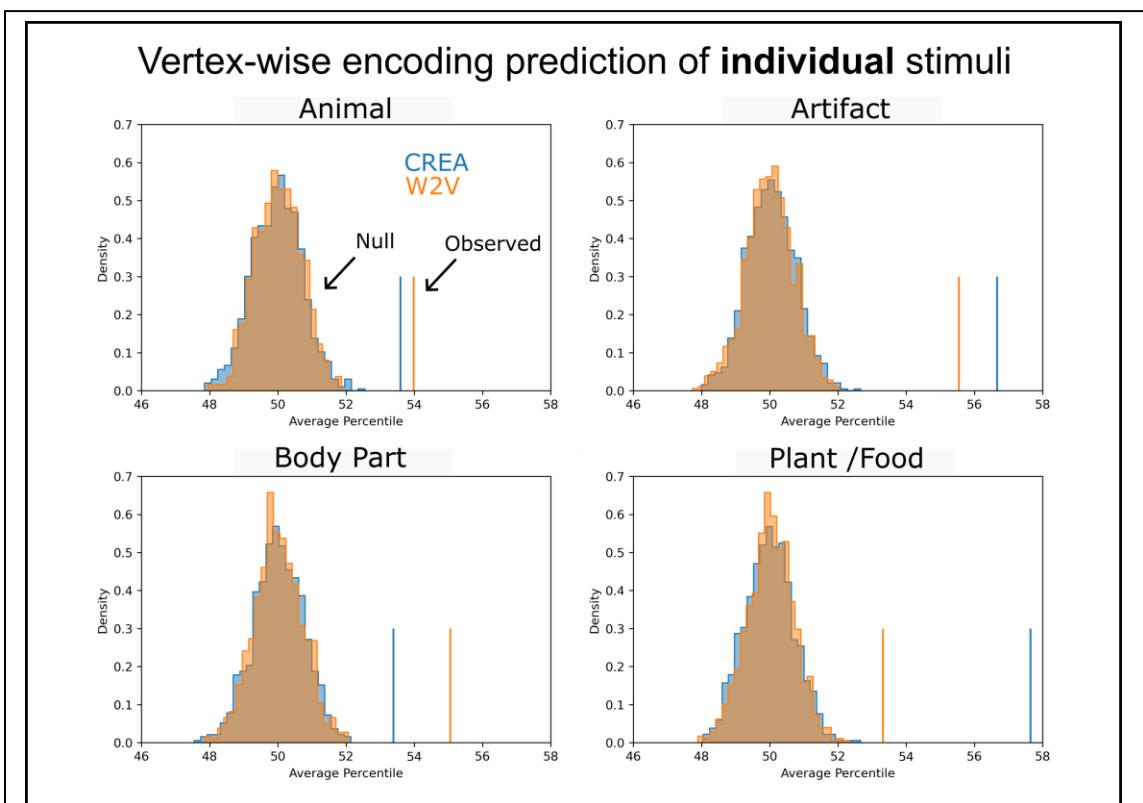
CREA



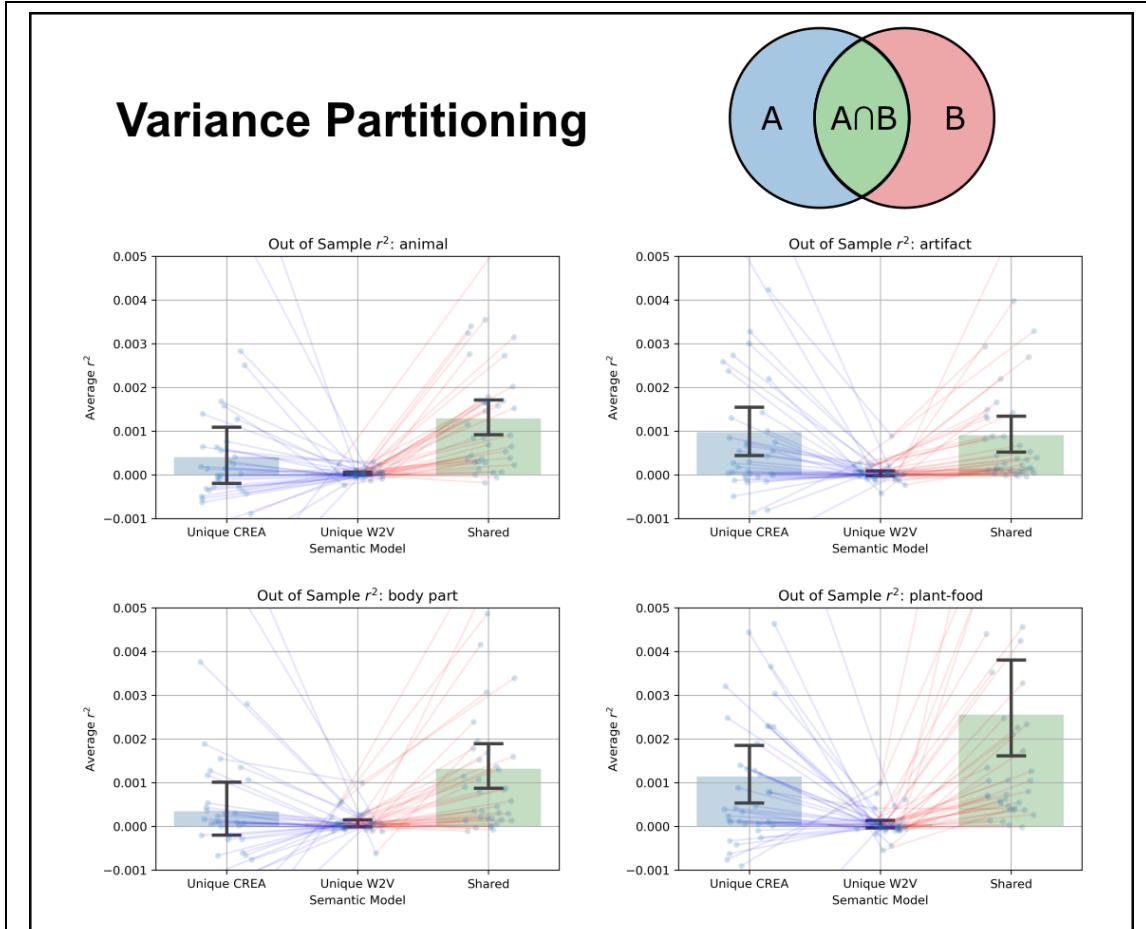
W2V



**Figure 12: Vertex-wise encoding – Category Difference.** **Top:** Vertex-wise encoding models were used to predict category contrasts. Each box above the diagonal contains the Spearman correlation coefficient between the predicted difference map and the observed difference map. Significance was determined by permuting the training labels 10,000 times. Values below the diagonal are equivalent Z-scores for the Spearman correlations that were estimated through the null distribution. **Bottom:** Visualization of observed and predicted average category differences for two category comparisons.



**Figure 13: Vertex-wise encoding – Individual Words.** Encoding models were used to predict individual stimuli in the labelled left out category. All left out patterns were correlated with the predicted pattern, and the average percentile of the target map was calculated for each participant, and these percentiles were then averaged across participants. Significance was determined by shuffling the training data 1,000 times.



**Figure 14: Variance partitioning.** Each ‘dot’ represents the unique variance attributed to a given feature space for a given participant. In all plots, error bars represent the 95% confidence interval of the mean calculated from 1,000 bootstraps.

Variance partitioning was used to estimate each model's unique explanatory power. As shown in **Figure 14**, in all cases, the average out of sample  $r^2$  is highest for the shared model, and lowest for the unique W2V component.

## Discussion

Our results show that there is a strong degree of clustering of concepts across the cortex and that weighted combinations of feature maps can account for a significant degree of this organization.

Univariate contrasts: Each of the 12 pair-wise univariate contrasts had vertices that were more significant for one category than for another. This result is in line with much prior work. One anticipated finding from the pair-wise comparisons is that sufficiently powered studies are likely to find *some* regions of cortex that are activated more for one category than another. This finding is implicit evidence for distributed representations because if conceptual information was represented as amodal symbols in a dedicated language network, one might not expect all pair-wise contrasts to have significant vertices.

The specific contrasts observed between categories were broadly in keeping with the literature. For example, the left lateralized network demonstrating greater activation for artifacts over animals is similar to meta-analyses contrasting animals and tools (Binder et al. 2009; Chouinard and Goodale 2010). However, there were some differences as the right IPS and right MFG activation we observed in the reverse contrast was not found in the Goodale meta-analysis. Two prior studies of CSRD in patients with neurodegenerative disease (mainly SD) also linked animal > artifact deficits with pathology in the right hemisphere (Brambati et al. 2006; Henderson et al. 2021). This preferential representation of animal concepts in the right hemisphere is interesting and of unknown significance. Further follow-up is required to better understand this phenomenon.

**Multivariate clustering:** As revealed by both pseudo F-scores and UMAP visualization, there was a striking degree of category organization across the cortex. While this result could be partially expected (e.g., by the performance of taxonomic models in RSA; Tong et al. 2022), the magnitude demonstrates that each of the categories appear similarly distinct, with no two categories strongly overlapping. The exhibited degree of category organization corroborates the findings of Capitani et al. (2003) and supports the significance of finding a shared neural substrate.

One benefit of using a dimensionality reduction technique to visualize clusters in data is that it can address the problem Gessell and colleagues (2021) refer to as cross-cut categories. The problem is that, given a set of stimuli and corresponding brain activation patterns, there might be several ways of grouping the stimuli into categories such that the categories could be ‘decoded’ from the brain activation patterns. The “cross-cut categories” problem is that there appears no principled way to determine what categorical distinctions are ‘actually’ represented. Dimensionality reduction techniques provide an elegant alternative to simple decoding approaches. Because UMAP, like MDS, tries only to visualize the pairwise distances between the observations, it has no built-in priors about the data’s structure. As a result, any ‘natural’ categories that appear are those driven by the data. Figure 2c demonstrates a strong degree of separation between all 4 semantic categories such that any criteria used to group stimuli together would still need to account for the minimal overlap between any two categories. That is, while the data could be accounted for by *more specific* models (e.g., ones that further fractionate the categories), it is difficult to imagine different criteria that divide stimuli at the same level of specificity as the four chosen categories, and still account for the observed structure. That

said, figure 2c is also in accord with *less specific* models, as the animate (animal and plant/food categories) are also (necessarily) separable from the inanimate categories (body part and artifact categories). This result accords well with the prior literature on CRSDs.

Encoding analysis: Taken together, the results of the encoding analyses, both at the group and individual levels, provide evidence of a shared neural representation. In the case of predicting the group contrast, we note that significant prediction requires the challenging task of modelling both left-out target categories by estimating feature loading using stimuli from unrelated categories. The ability to do this task means that relevant features must be shared across all categories. One limitation of the analysis strategy taken regards the applicability of averaging word activation maps across participants. An alternative approach is to model the contrast differences in each participant, and then averaging across participants. However, this approach would require roughly 40 times the amount of time it already takes to generate the null distribution with 10,000 permutations. The permutation of the training data to generate empiric null distributions provides rigorous support that the results depend on properly creating feature maps and would not succeed with arbitrary features.

One interesting observation of the encoding analysis was that the semantic models exhibit distinct patterns of generalization ability. While both semantic models had many significant predictions, the models varied in performance. Further, the pattern of generalization was different in the encoding (Figure 13), and variance partitioning analysis (Figure 14). In Figure 13, the W2V model had a higher average percentile than CREA for two categories, whereas CREA had a higher average correlation than W2V in

all cases in Figure 4. Interestingly these are the same two categories where there was no significant unique explanatory information of the CREA model, meaning that the mean  $r^2$ 's are not statistically different from each other.

Differences in how the models perform could be the result of how well the training or testing stimuli are modelled, and further analysis is required to tease these effects apart. However, separating these effects could also amplify the artifacts that can occur with cross-validation in the setting of finite sample sizes (especially when there are far more predictors than targets) and is outside this paper's scope.

*Variance partitioning:* The variance partitioning analysis indicates that much of the across category generalization is shared by the experiential and semantic models. There was little unique information represented in W2V that generalized across categories. This finding, that most of the relevant semantic information in distributional semantic models is also present in experiential models fits with prior results from Fernandino et al. (2022).

*Future Directions:* Work remains to be done in determining the relevant experiential content for each category. Further, understanding the features that best generalize across categories could provide more insight into the neurobiologically relevant features. Other models and other categories should also be considered.

## Conclusions

Our results support a shared representational basis across the categories of CRSDs. The experiential model generally outperformed the distributional model in encoding information that generalized across categories. Variance partitioning analysis indicated

that while both models were mostly capturing similar information, where there was unique information that generalized, it was mostly in the experiential model. Taken together, there is compelling evidence that observed category related organization of semantic content can be accounted for by an experiential account of concept representation.

## Chapter 4: The neural representation of body part

### concepts<sup>12</sup>

#### Introduction

The previous chapter examined the degree of category organization in the cortex. One finding of the previous chapter was the larger activation in many areas for body parts compared to other categories. Reiterating some of the background of the previous chapter, although the category of body part concepts has shown both selective impairment (Suzuki et al. 1997) and sparing (Shelton et al. 1998) after focal brain damage, the extant fMRI literature on body part representation almost exclusively focuses on the perception of pictorial stimuli (Downing et al. 2001) without a clear distinction between high-level visual perceptual and post-perceptual concept knowledge. The dearth of fMRI studies investigating the neural substrate of lexicalized body part concept representations is a limiting factor in our ability to explain CRSDs because this neuropsychological phenomenon is often assessed through semantic tasks and is considered an impairment of conceptual knowledge rather than visual perception. Filling this knowledge gap could help resolve ongoing debates regarding the dissociable components of body-part knowledge (Longo et al. 2010; Vignemont 2007; Vignemont 2010) and the involvement of particular cortical areas in its representation (Boccia et al. 2020; Kemmerer and Tranel 2008; Schwoebel and Coslett 2005).

---

<sup>12</sup> A version of this chapter has been accepted and is in press at *Cerebral Cortex* with an approximate citation of: Mazurchuk, S., Fernandino, L., Tong, J.-Q., Conant, L. L. & Binder, J. R. The Neural Representation of Body Part Concepts. *Cereb. Cortex* (2024)

One proposed model of body representation in the brain is the dyadic taxonomy (Paillard 1999), which distinguishes *body schema*, a dynamic representation of current body part positions that can be thought of as a low-level sensory representation of the body, and *body image*, a representation that is semantic in nature and includes long-term knowledge about “what” body parts are. Another proposed model is the triadic taxonomy (Schwoebel and Coslett 2005), which splits the body image representation further into the *body structural description*, a fixed topological map of body locations and boundaries, and *body semantics*, which includes knowledge of body part names, functions, and relations with artifacts (Boccia et al. 2020). Kemmerer and Tranel (2008) argue that the body structural description is essentially semantic in nature, and they divide body semantics (body image in the dyadic taxonomy) into four components representing knowledge of body part shapes, locations, functions, and “cultural associations,” each proposed to have somewhat distinct neural correlates. In the current study, we define body semantic knowledge operationally as any information activated by body part words that cannot be attributed to word-form (orthographic and phonological) properties of the stimuli. Such information could include any of the types of knowledge described by Kemmerer and Tranel.

At present, there is debate about whether body part concept knowledge is functionally and anatomically dissociable from other categories of lexical concepts. Although most reviews of data from brain damaged individuals concluded that body-part knowledge is typically spared (Capitani et al. 2003; Gainotti 2004; Goodglass et al. 1966), there is clear evidence that selective impairment of this category can occur (Kemmerer and Tranel 2008; Laiacona et al. 2006; Sacchett and Humphreys 1992;

Schwoebel and Coslett 2005; Suzuki et al. 1997; Warrington and McCarthy 1987).

Although most of these patients had large lesions, the three reported by Schwoebel & Coslett (2005) with selective impairment of body-related functional and associative knowledge all had damage confined to the left temporal lobe, with maximal overlap in the posterior middle temporal gyrus (pMTG, part of Brodmann area 37).

Previous fMRI work on lexical-semantic body part representation is limited to two studies using voxel-wise mass univariate analysis, both of which had relatively few participants and few stimuli (Goldberg et al. 2006a; Le Clec'H et al. 2000). Goldberg and colleagues (2006a) found that a task requiring similarity judgments on items from the “functionally biased” categories of body parts and clothing activated the left pMTG in comparison to items from two “visually biased” categories (fruits and birds). Le Clec'H and colleagues (2000) reported greater activation during a location judgment task on body part names (“Is it above or below the shoulders?”) compared to a judgment task on numbers (“Is it less than 12?”) in the left intraparietal sulcus and several left frontal regions. Our aim is to expand on this work and clarify the neural correlates of body part concept representations, both on their own and relative to other categories of concrete object nouns. To minimize the possibility of any observed category activation differences being the result of unaccounted for confounds, we also examined the preferential representation of body part concepts using explicit models of semantic content. We used two different kinds of models: the 65-dimensional model of concept representation as experiential (i.e., embodied) attributes (hereafter, the CREA model; Binder et al. 2016) and a popular distributional semantic model (Word2Vec; Mikolov et al. 2013).

We hypothesized that body part concepts would be preferentially represented in the left pMTG region, based on the lesion analysis of Schwoebel and Coslett (2005), the proximity of this region to lateral occipitotemporal areas specialized for visual perception of body parts (Bracci et al. 2015; Orlov et al. 2010; Spiridon et al. 2006), and on strong fMRI evidence linking the pMTG with representation of object-directed actions, where body parts are effectors (Jastorff et al. 2010; Lingnau and Downing 2015; Oosterhof et al. 2010). Complementary analyses that address different aspects of representation were conducted to identify convergent results across univariate and multivariate analysis strategies.

Our analysis strategies are divided into those that are independent of an explicit model of semantic content (e.g., standard univariate contrasts between categories) and those that depend on a model of semantic content. Similar to the previous chapter, we first performed a univariate analysis contrasting activation elicited by body part words to activation elicited by each of three other concrete noun semantic categories (animals, plants/foods, and manipulable artifacts). These pairwise comparisons were then entered into a conjunction analysis to determine if there are regions where, in the context of a general semantic judgment task, body part words elicit greater activation than words from other categories. Complementing this univariate analysis, multivariate pattern analysis (MVPA) was performed to identify regions of the cortex where category identity can be decoded from activation patterns. This analysis assumes that regions of cortex where category identity can be decoded from activation patterns carry semantic information that is relevant to that category. Specifically, we used searchlight MVPA to compare the

ability of a pattern classifier to discriminate body parts from other categories relative to the classifier's ability to discriminate other categories from each other.

A final model-free analysis assessed the reliability of the neural representational geometry of body part concepts across participants. Our interest in performing reliability analysis stems from its relation to a common definition of semantics. At a minimum, a 'semantic' representation is a set of relationships among items that is shared across participants. For example, many semantic models encode the relation that 'birds' and 'airplanes' are more similar than 'birds' and 'computers'. The 'neural similarity' between pairs of words can be computed by comparing the multi-voxel activation patterns they evoke, and the global representational geometry of the neural space can be computed as a matrix of all such pairings, referred to as the neural representational dissimilarity matrix (neural RDM; Kriegeskorte et al. 2008). Since neural RDMs quantify relationships between experimental conditions, operationalizing the previous definition of semantics implies a requirement of similarity in neural RDMs across participants to claim that semantic content is represented. This between-participant similarity can be quantified in reliability metrics such as noise ceilings and intraclass correlation coefficients (ICCs). We emphasize, however, that while a shared representational geometry is a prerequisite for many definitions of a semantic space, it is not sufficient to establish the presence of semantic content. For example, visual word stimuli likely evoke similar neural representational geometries in the primary visual cortex across participants, even though this shared representation is not canonically considered semantic. Performing searchlight RDM reliability analysis allows us to identify all the possible regions of cortex where a model could account for shared variance, and the regions of cortex without shared

variance. To investigate regions that preferentially represent body part knowledge, we searched for regions that had greater shared variance across body part words than across items from other categories of words.

The model-based analyses included a cortical surface-based searchlight RSA and a vertex-wise encoding analysis. To test whether the semantic models capture the representational geometry among body part concepts, we performed searchlight RSA restricted to body part concepts. To assess whether the semantic models predicted any body part selective regions, we calculated RSA correlations for each category and searched for regions that had higher RSA scores for the body part category compared to any of the other categories. These analyses aim to identify regions where activation patterns capture differences between body part concepts more than differences between items within other categories. This is relevant to defining a ‘category-preferring’ region because CRSDs are typically characterized by an inability to discriminate between items *within* a category, rather than by an inability to determine if an item belongs to a particular category.

Lastly, one prediction of embodied theories of semantics is that CRSDs arise from spatial constraints imposed by perceptual processing areas relevant to the affected category. We used encoding models to test whether apparent category localization can be explained by feature content that is shared across categories but simply more important for some categories. Specifically, we tested whether any univariate results showing increased activation to body part concepts relative to other categories of concepts are predicted by semantic encoding models trained using concepts outside the body part category.

## Materials and Methods

Relative to many fMRI studies of lexical semantics, the present study included a larger number of participants (38), stimuli (50 per category), and image volumes (5,600) to increase statistical power. We also improved on previous imaging studies of body part concept representation (Goldberg et al. 2006a; Le Clec'H et al. 2000) by using fMRI methods with higher spatial and temporal resolution.

*Participants:* Participants are the same as those described in Table 3 in chapter 2 of this dissertation.

*Stimuli and Task:* The complete stimulus set used in the fMRI study is the same as described in chapter 2 (Table 4) and described in Tong et al. (2022). For review, it consisted of 300 English nouns from 6 categories: animals, manipulable artifacts (hereafter “tools”), plants/foods, body parts, human traits, and quantities, with 50 words in each category. We focus here on the four concrete object categories (animals, tools, plants/foods, and body parts). The scanning protocol followed that outlined in chapter 2 of this dissertation.

*Semantic Models:* The model-based analyses used experiential (CREA) and distributional (Word2Vec) models of semantic content. These models are discussed in the introduction to this dissertation.

*Acquisition Parameters:* Acquisition parameters were the same as outlined in chapter 2 of this dissertation.

*Between-Category Comparisons:* The analysis of category effects in the neural data is divided into two parts based on whether the analysis depends on an explicit model of semantic content.

As done in chapter 3 of this dissertation, standard univariate analyses were performed to generate beta-maps for each of four categories (animals, tools, plants/foods, and body parts). Participant category beta-maps were then contrasted at the group level, and these pair-wise differences were tested for significance at  $p < .01$  using a Wilcoxon signed-rank test and FDR correction (Benjamini and Hochberg 1995) implemented in the python package ‘Statsmodels’ (Seabold and Perktold 2010). To identify putative body part selective regions, we created a conjunction map of regions where body part activation was significantly greater than each of the other three categories. To aid in visualization, the conjunction analysis was restricted to clusters of at least 40 mm<sup>2</sup>.

To assess whether the main results were affected by inclusion of a broad range of items in the body part category, univariate conjunction analyses were also performed using subsets of words in the body part category (**Table 5**). One possibly important factor in our stimulus set is whether the body parts are internal (e.g., liver) or external (e.g., nose). Another major division is between body parts that can be used as effectors and those that cannot be. To operationalize this latter distinction, we considered a body part to be a possible effector if it can be voluntarily moved in isolation. Lastly, the items were divided by general location into head, upper limb, torso, and lower limb. The body part words were placed into each group based on discussion and consensus among several authors who prepared this work for publication. We recognize, however, that these categorizations are in some cases ambiguous and open to debate. Similar to the

| Word       | Internal / External | Movable | Location  |
|------------|---------------------|---------|-----------|
| armpit     | external            | no      | UpperLimb |
| beard      | external            | no      | Head      |
| earlobe    | external            | no      | Head      |
| fingernail | external            | no      | UpperLimb |
| instep     | external            | no      | LowerLimb |
| mustache   | external            | no      | Head      |
| navel      | external            | no      | Torso     |
| nipple     | external            | no      | Torso     |
| toenail    | external            | no      | LowerLimb |
| waist      | external            | no      | Torso     |
| tooth      | external            | no      | Head      |
| testicle   | external            | no      | Torso     |
| retina     | internal            | no      | Head      |
| bladder    | internal            | no      | Torso     |
| cartilage  | internal            | no      | NA        |
| diaphragm  | internal            | no      | Torso     |
| intestines | internal            | no      | Torso     |
| kidney     | internal            | no      | Torso     |
| ligament   | internal            | no      | NA        |
| liver      | internal            | no      | Torso     |
| pancreas   | internal            | no      | Torso     |
| stomach    | internal            | no      | Torso     |
| trachea    | internal            | no      | Torso     |
| abdomen    | external            | yes     | Torso     |
| belly      | external            | yes     | Torso     |
| cheek      | external            | yes     | Head      |
| eyebrow    | external            | yes     | Head      |
| eyelid     | external            | yes     | Head      |
| forehead   | external            | yes     | Head      |
| nose       | external            | yes     | Head      |
| nostril    | external            | yes     | Head      |
| torso      | external            | yes     | Torso     |
| ankle      | external            | yes     | LowerLimb |
| elbow      | external            | yes     | UpperLimb |
| finger     | external            | yes     | UpperLimb |
| forearm    | external            | yes     | UpperLimb |
| heel       | external            | yes     | LowerLimb |
| leg        | external            | yes     | LowerLimb |
| lip        | external            | yes     | Head      |
| pelvis     | external            | yes     | Torso     |
| shoulder   | external            | yes     | UpperLimb |
| thigh      | external            | yes     | LowerLimb |
| thumb      | external            | yes     | UpperLimb |
| wrist      | external            | yes     | UpperLimb |
| knuckle    | external            | yes     | UpperLimb |
| clavicle   | internal            | yes     | Torso     |
| skeleton   | internal            | yes     | NA        |
| skull      | internal            | yes     | Head      |
| spine      | internal            | yes     | Torso     |
| muscle     | internal            | yes     | NA        |

**Table 6: Body part stimuli groupings.** Body parts were divided into subcategories based on consensus agreement amongst study authors.

univariate analysis with all body parts, for each subcategory, conjunction analysis was performed relative to the three other concrete object noun categories. In addition, words within each subcategory (e.g., internal and external) were directly contrasted.

MVPA analyses used surface-based ‘searchlight’ regions of interest covering the entire cortical surface as well as vertex-wise t-values instead of beta-values. Searchlights were identical for the support vector machine (SVM) classifier, neural reliability, and RSA analyses. Searchlight patches were created using the Connectome Workbench toolbox (Marcus et al. 2011) as 10mm radius circular surface areas defined on the HCP-fsLR-32k midthickness surface.

The first MVPA consisted of training cross-validated SVM classifiers using all 6 pairs of the 4 categories. To create an index of regions more selective for body parts, the 3 pair-wise comparisons that included the body part category were averaged and compared to the average accuracy when body parts were not a target category. The rationale for this analysis is that information that is particularly relevant to body parts should be more useful for discriminating body part concepts from other object concepts than for discriminating between non-body-part categories. We note that the SVM results are qualitative in nature, as we opted to not estimate p-values using a binomial distribution because of evidence that the null distribution for cross-validated classifiers does not follow the binomial distribution (Noirhomme et al. 2014). A peak-finding algorithm was applied to the smoothed results (4 mm FWHM), in which extrema were required to be separated by at least 15 mm and have a minimum value of 0.07.

The second MVPA examined within-category reliability of the neural RDMs across participants for each category. Neural RDMs were calculated at the individual

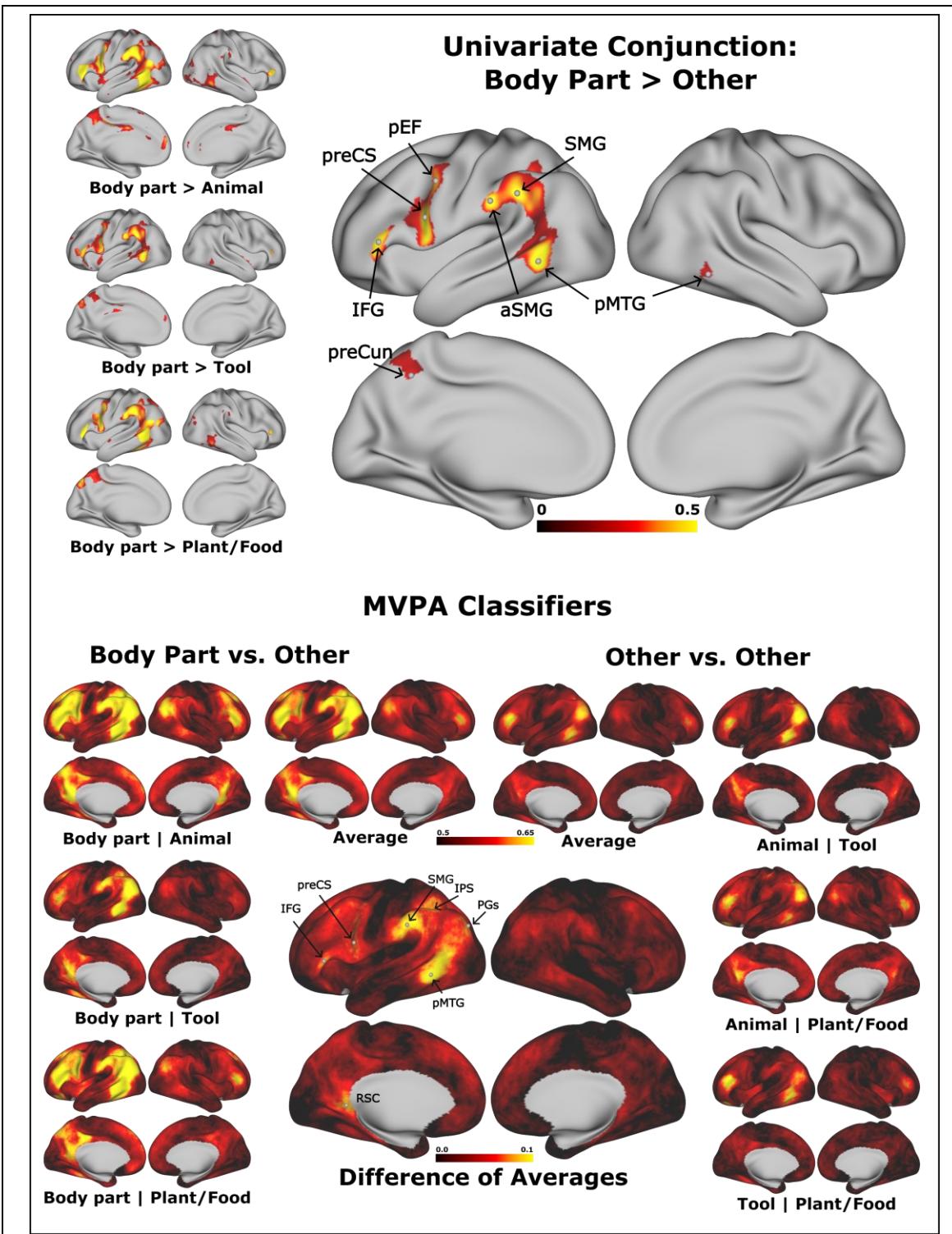
level using the Pearson correlation distance between the set of t-values within each surface patch. The reliability of these RDMs was calculated using the consistency-based ICC value. The ICC metric is related to standard noise ceiling measures but has the advantage of an associated framework for statistical interpretation. We note that ICC values are often interpreted based on ranges of the statistic (e.g., Cicchetti (1994) suggests that values between 0.60 and 0.74 be considered ‘good’ agreement) rather than using p-values, and we therefore show the unthresholded statistic map so that patterns can be qualitatively compared to RSA maps. To determine whether there were regions that displayed body part *selectivity*, the reliability maps were contrasted between categories. Because there is no standard method for comparing ICC values with a repeated measures paradigm, we instead searched for vertices where the body part ICC value was above the range of the 99% confidence interval for each of the other three categories.

The third MVPA used model-dependent RSA to compare neural RDMs for each category with semantic model RDMs derived from either the 65-feature CREA model or Word2Vec. For RSA, similar to the within-category ICC analysis, we were interested in the qualitative pattern and therefore show the unthresholded RSA map. RSA using items restricted to one category indicates the regions involved in semantic representation of that category but does not speak to the preference of a region for a particular category. To parallel the approach taken in the univariate analysis, we used a Wilcoxon signed rank test on the paired differences between within-category RSA maps. The resulting p-values were FDR corrected and thresholded at  $p < .01$ . The p-value maps generated from this analysis were then eroded and dilated by 2mm to smooth the results for visualization.

Encoding Analysis. While a significant RSA correlation for a category of nouns indicates that a semantic model captures the relevant relations *within* the category, it does not necessarily mean that the model accounts for differences between categories. That is, the hitherto described RSA does not address whether the semantic models explain activation differences *between* body parts and non-body-part items. To test whether the semantic models can account for the mass univariate results, we performed a vertex-wise encoding analysis to generate predicted activation differences between body part words and other categories at each vertex, which were then compared to the observed differences. This analysis used 150 individual word beta maps from the non-body-part conditions to predict beta maps for the left out 50 body part concepts. The encoding model was a ridge regression model that used leave-one-out cross-validation to determine the optimal penalty term on the training data. The 50 predicted maps were averaged across words and compared to the average of the 150 training words to determine the predicted contrast between body part and non-body-part words for each participant. These differences were smoothed using a 6mm FWHM gaussian and tested against zero at each vertex using a Wilcoxon signed rank test with significance set as FDR corrected p-values < 0.01.

## Results

We group the analyses of neural data into semantic model independent and model dependent analyses. Performance on the familiarity judgment task during fMRI showed an excellent response rate of 97.5% and good intraindividual consistency across repeated presentations of the same word (mean ICC = 0.655).



**Figure 15: Mass univariate and multi-voxel pattern classifier analyses.** Top: The top portion of the figure shows where BOLD signal was greater for body part concepts than concepts from other categories based on standard mass univariate contrasts. All individual contrasts are significant at FDR corrected  $p < .01$  using a 1-tailed Wilcoxon

signed-rank test. The conjunction map shows where body parts produced greater activation in all comparisons. Colors in the individual comparisons represent average difference in beta-values. **Bottom:** The lower portion shows average classification accuracies for held out data in cross-validated searchlight SVM category classification. Classes were balanced, and 50% indicates chance accuracy. Shown in the center is the difference in average accuracy when one of the categories was body parts compared with when body parts are not a target category. Abbreviations are identified in the caption for table 2.

The univariate conjunction analysis results in **Figure 15** show regions of cortex that are activated more for body part nouns compared to other concrete object noun categories. These regions included the left pMTG, left supramarginal gyrus (SMG), left precentral sulcus (including the premotor eye field), left IFG, and a small portion of the intraparietal sulcus (IPS) (**Table 6**). In addition, there were small peaks in the left precuneus, left parainsular cortex, and right pMTG.

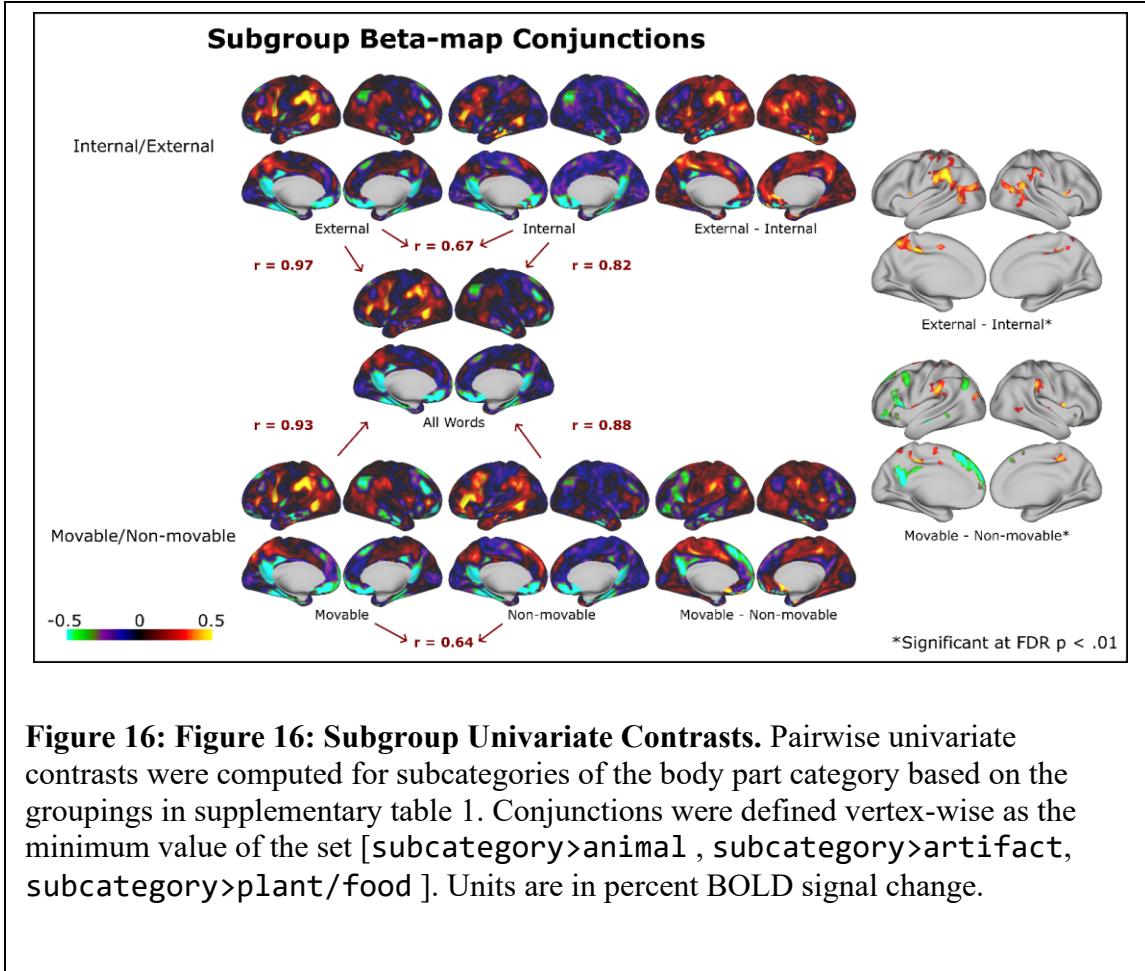
Unthresholded maps that used subsets of the body part category showed similar results to the full set of stimuli (**Figure 16**) and are discussed below.

*Internal/External:* There were more external than internal body parts (n = 36 and 14, respectively). The external body part conjunction map looked similar to the full set (Pearson  $r = .97$ ). The internal body part map, while being qualitatively similar to the full set ( $r = .82$ ), had noticeably less activation in the left SMG (Figure 16). While the external body part map had significant vertices in the conjunction analysis, the internal body part map had no significant vertices. A direct comparison of internal and external body parts showed that the left SMG was significantly more activated by external relative to internal body parts. Taken together, these results suggest that external body parts are

likely responsible for much of the observed findings in the analysis containing all body part stimuli.

| <b>Analysis</b> | <b>Region</b> | <b>Coordinates (MNI)</b> |     |     | <b>Value at Vertex</b> | <b>HCP Parcel</b> |
|-----------------|---------------|--------------------------|-----|-----|------------------------|-------------------|
| Univariate      | pMTG          | -55                      | -56 | -3  | 0.668                  | L_PHT             |
|                 | preCS         | -50                      | 5   | 22  | 0.646                  | L_6r              |
|                 | SMG           | -60                      | -43 | 36  | 0.557                  | L_PF              |
|                 | aSMG          | -63                      | -27 | 30  | 0.538                  | L_Pfop            |
|                 | IFG           | -50                      | 33  | 8   | 0.525                  | L_IFSa            |
|                 | paraIns       | -41                      | -6  | -15 | 0.455                  | L_PI              |
|                 | pEF           | -46                      | -1  | 40  | 0.435                  | L_PEF             |
|                 | IPS           | -36                      | -50 | 38  | 0.404                  | L_IP2             |
|                 | preCun        | -7                       | -53 | 46  | 0.288                  | L_PCV             |
|                 | pMTG          | 59                       | -53 | -12 | 0.274                  | R_PHT             |
| SVM             | pMTG          | -55                      | -55 | -6  | 0.111                  | L_PHT             |
|                 | SMG           | -60                      | -35 | 35  | 0.107                  | L_PF              |
|                 | IPS           | -41                      | -46 | 41  | 0.101                  | L_IP2             |
|                 | RSC           | -9                       | -48 | 3   | 0.093                  | L_RSC             |
|                 | IFG           | -53                      | 29  | 4   | 0.084                  | L_45              |
|                 | PGs           | -36                      | -82 | 34  | 0.081                  | L_PGs             |
|                 | pMFG          | -28                      | 16  | 48  | 0.081                  | L_8Av             |
|                 | preCS         | -52                      | 6   | 20  | 0.079                  | L_6r              |

**Table 7: Peak coordinates corresponding to labels corresponding to the univariate analysis shown in Fig. 15.** Coordinates refer to MNI 152 template space and parcel names are taken from the surface-based atlas in Glasser et al (2016). For the univariate analysis, values correspond to the smallest bold difference between body parts and any other category. For the SVM analysis, values correspond to the average difference in cross-validated accuracy between category comparisons containing body parts, and category comparisons that do not contain body parts. IFG = inferior frontal gyrus, IPS = intraparietal sulcus, paraIns = parainsular cortex, pEF = premotor eye field, PGs = superior part of Von Economo and Koskinas area PG, pMFG = posterior middle frontal gyrus, pMTG = posterior middle temporal gyrus, preCS = precentral sulcus, preCun = precuneus, RSC = retrosplenial cortex, SMG = supramarginal gyrus.

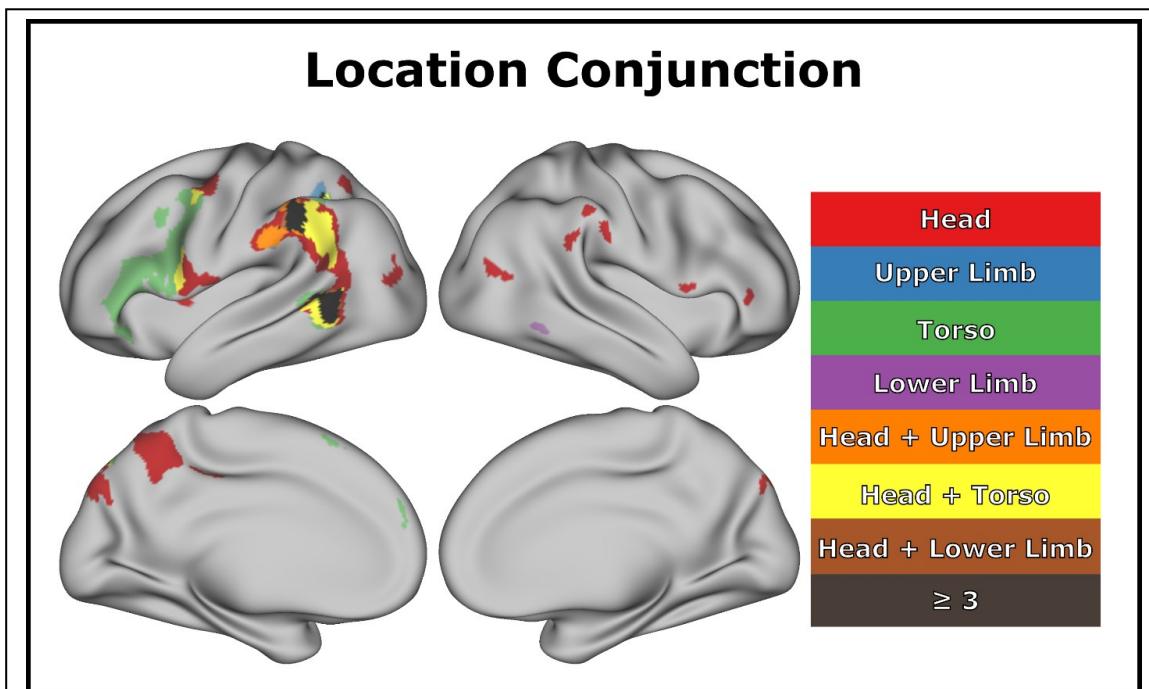


**Figure 16: Subgroup Univariate Contrasts.** Pairwise univariate contrasts were computed for subcategories of the body part category based on the groupings in supplementary table 1. Conjunctions were defined vertex-wise as the minimum value of the set [subcategory>animal , subcategory>artifact, subcategory>plant/food ]. Units are in percent BOLD signal change.

Movable/Non-movable: There were a similar number of movable and non-movable body parts included in the study ( $n = 27$  and 23, respectively). Movable body parts had high overlap with external body parts (22 of 27 moveable body parts were external), and correspondingly, the conjunction maps for movable and non-movable body parts were similar to the maps for external and internal body parts, respectively. Both movable and non-movable conjunction maps had high similarity to the conjunction map generated from the full set ( $r = 0.93$  and 0.88, respectively). Direct contrast between movable and non-movable showed increased activation for movable relative to immobile body parts in the bilateral SMGs. Conversely, non-movable body parts showed increased

activation relative to movable body parts in several left hemisphere areas including the posterior cingulate gyrus, dorsomedial prefrontal cortex, inferior frontal cortex, angular gyrus, and posterior middle frontal gyrus. However, except for the IFG, these regions are all less activated for body parts relative to other categories of words. We therefore do not believe that the factor(s) underlying the non-movable > movable activation in these regions is related to their membership in the body part category.

Location: Body parts were grouped into 4 somatotopic categories: Head (n=13), Upper Limb (n=9), Torso (n=18), and Lower Limb (n=6). As with previous subset analyses, a conjunction map was generated for each of these categories relative to the 3 other concrete object noun categories. Each of the unthresholded conjunction maps had



**Figure 17: Location contrasts.** Conjunction analysis with body part groups and other concrete object noun categories. All colored vertices are significant at FDR corrected  $p < .01$  and a minimum cluster size of 40 mm<sup>2</sup> was used to aid in visualization.

high similarity to the full set (all Pearson  $r > .8$ ). The thresholded results of these conjunction analyses are shown in **Figure 17**. The two largest subcategories of Torso and Head had the largest regions of activation, with Torso showing activation in the left IFG, and Head showing activation in the precuneus. Overall, the main regions of activation (pMTG, SMG, and preCS) are the same regions found using the full set of stimuli.

Direct contrasts between the four anatomical subgroups were also computed, as well as conjunctions. No subgroup had any significant vertices in conjunction analysis (e.g., a region where Head was significantly greater than all 3 other categories).

Subsets of external, internal, movable, and non-movable were all found to be highly correlated with a conjunction created from the full set of body part words (Pearson  $r$ 's = 0.97, 0.82, 0.93, 0.88, respectively; **Table 7**). While contrasts involving the other groups are discussed in the supplemental material, given the overall high similarity between subsets, all further results presented below are collapsed to the whole set to maximize power and generalizability.

The second analysis searched for regions of cortex where neural activation patterns discriminated body parts from other categories more than they did non-body-part categories from each other (Figure 1, bottom). Although qualitative in nature, the pattern is similar to that observed in the univariate analysis, with the largest peak differences located in the left SMG, left pMTG, left IFG, and left IPS (Table 2). In addition, there were small peaks in the left retrosplenial cortex (RSC), dorsal angular gyrus, precentral sulcus, and posterior middle frontal gyrus.

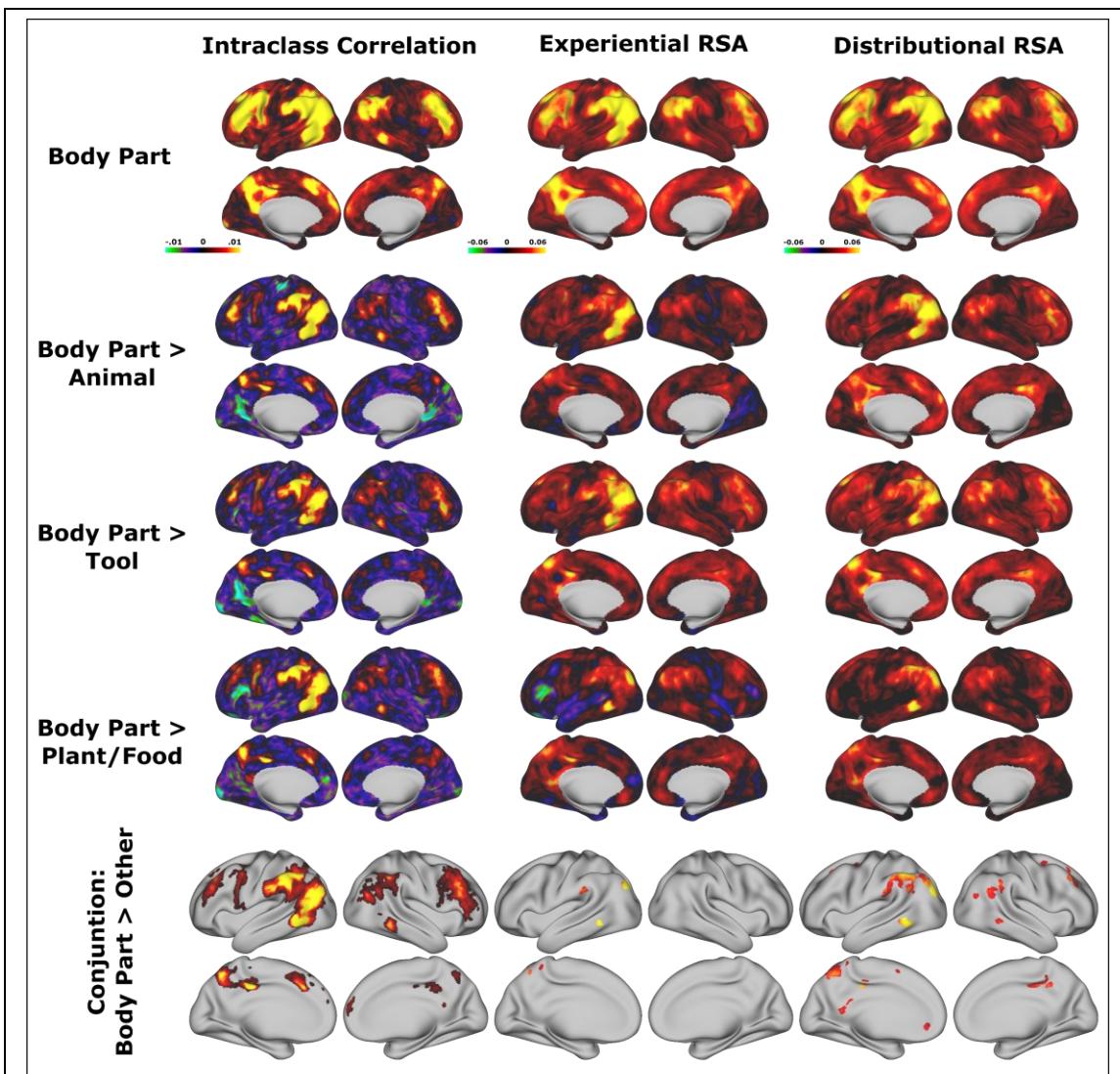
|              | External    | Internal    | Movable     | Non-movable | Head        | UpperLimb   | Torso       | LowerLimb   |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| All_BP_Words | <b>0.97</b> | <b>0.82</b> | <b>0.93</b> | <b>0.88</b> | <b>0.86</b> | <b>0.85</b> | <b>0.81</b> | <b>0.82</b> |
| External     |             | <b>0.67</b> | <b>0.95</b> | <b>0.80</b> | <b>0.89</b> | <b>0.89</b> | <b>0.71</b> | <b>0.86</b> |
| Internal     |             |             | <b>0.65</b> | <b>0.86</b> | <b>0.58</b> | <b>0.54</b> | <b>0.88</b> | <b>0.52</b> |
| Movable      |             |             |             | <b>0.64</b> | <b>0.87</b> | <b>0.89</b> | <b>0.61</b> | <b>0.85</b> |
| Non-movable  |             |             |             |             | <b>0.67</b> | <b>0.60</b> | <b>0.92</b> | <b>0.61</b> |
| Head         |             |             |             |             |             | <b>0.74</b> | <b>0.51</b> | <b>0.72</b> |
| UpperLimb    |             |             |             |             |             |             | <b>0.51</b> | <b>0.73</b> |
| Torso        |             |             |             |             |             |             |             | <b>0.51</b> |

**Table 8: Pair-wise Pearson correlations between different unthresholded conjunction maps.**

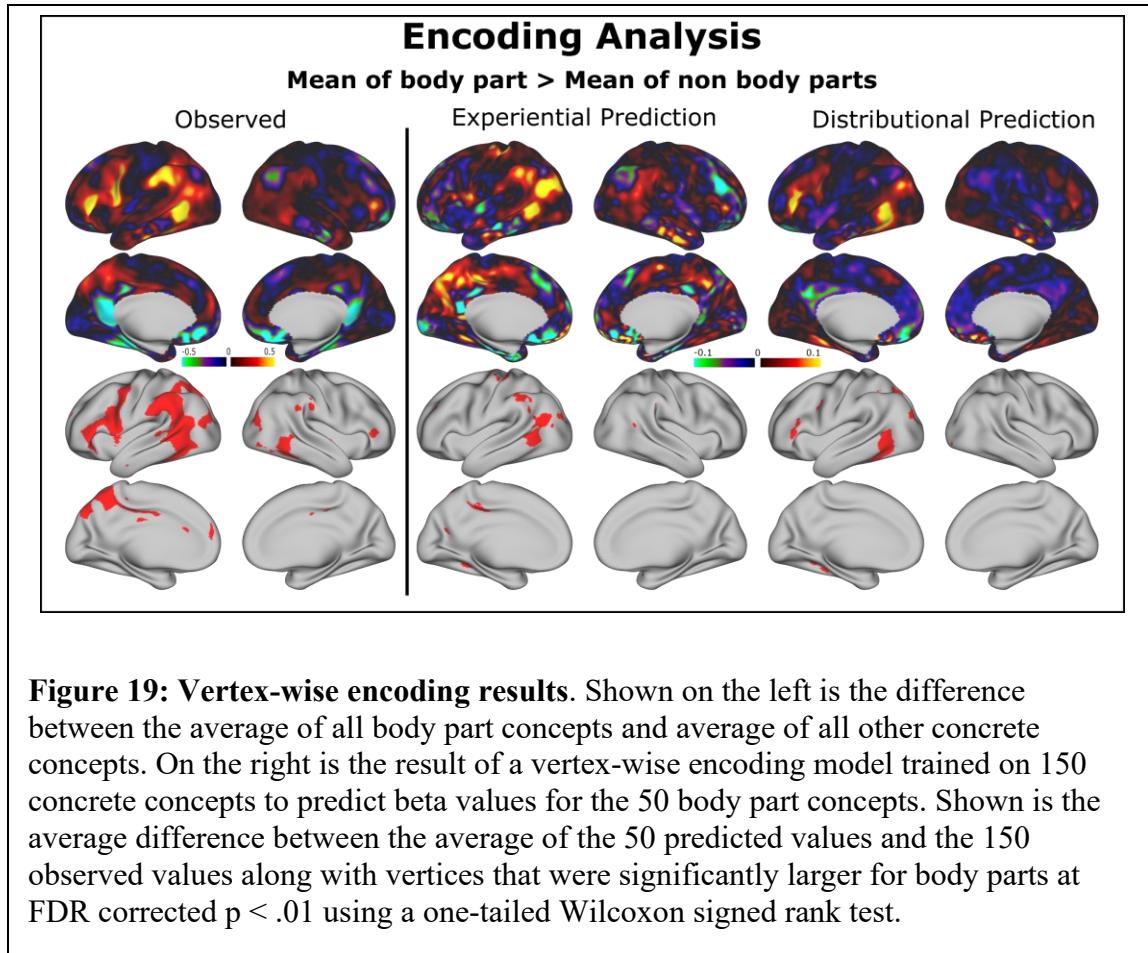
Searchlight neural RDM reliability and RSA results for the body part category (**Figure 18, top row**) showed a high degree of correspondence (Spearman rho across cortex: 0.801 for experiential model, 0.807 for distributional model, 0.963 between models). Peaks were found in the left pMTG, left SMG, left angular gyrus, left precentral sulcus, left posterior cingulate gyrus and precuneus, and right IFG.

To examine the neural substrate of body part *specific* representations, we contrasted body part ICC and RSA maps to those generated for other categories of nouns. In the case of the ICC analysis, body part selectivity was defined as regions where the ICC value for body parts was above the range of the ICC confidence interval for any other category. Peaks were found in the left pMTG, SMG, angular gyrus, precentral sulcus, and IFG, and several homologous regions in the right hemisphere. For the RSA contrasts, paired difference contrast maps were generated for body parts relative to the

other three categories. A conjunction map was then generated indicating the vertices where RSA values were significantly higher for body parts than for all other categories as assessed by an FDR corrected p-value  $< 0.01$  derived from a Wilcoxon signed-rank test. RSA contrasts using the two semantic models produced similar results. Both models had significant vertices in the left pMTG, SMG, posterior angular gyrus, and precuneus.



**Figure 18: ICC and RSA comparisons.** ICC and RSA maps were generated for each of four categories of nouns. Results for the body part category are shown in the top row, and contrasts between body part and other categories are shown in subsequent rows. The bottom row shows conjunction maps of the categorical comparisons. Colors in the ICC conjunction map represent the difference between the body part ICC value and the highest upper limit of the confidence interval for any of the other three categories. The RSA conjunction maps were generated by testing the mean of paired differences against zero using a Wilcoxon signed-rank test. The conjunction maps show vertices where body part RSA values were larger than for any other category using an FDR corrected  $p < 0.01$  threshold. The conjunction map color shows the mean paired difference between body part RSA values and whichever other category had the highest RSA value for that vertex.



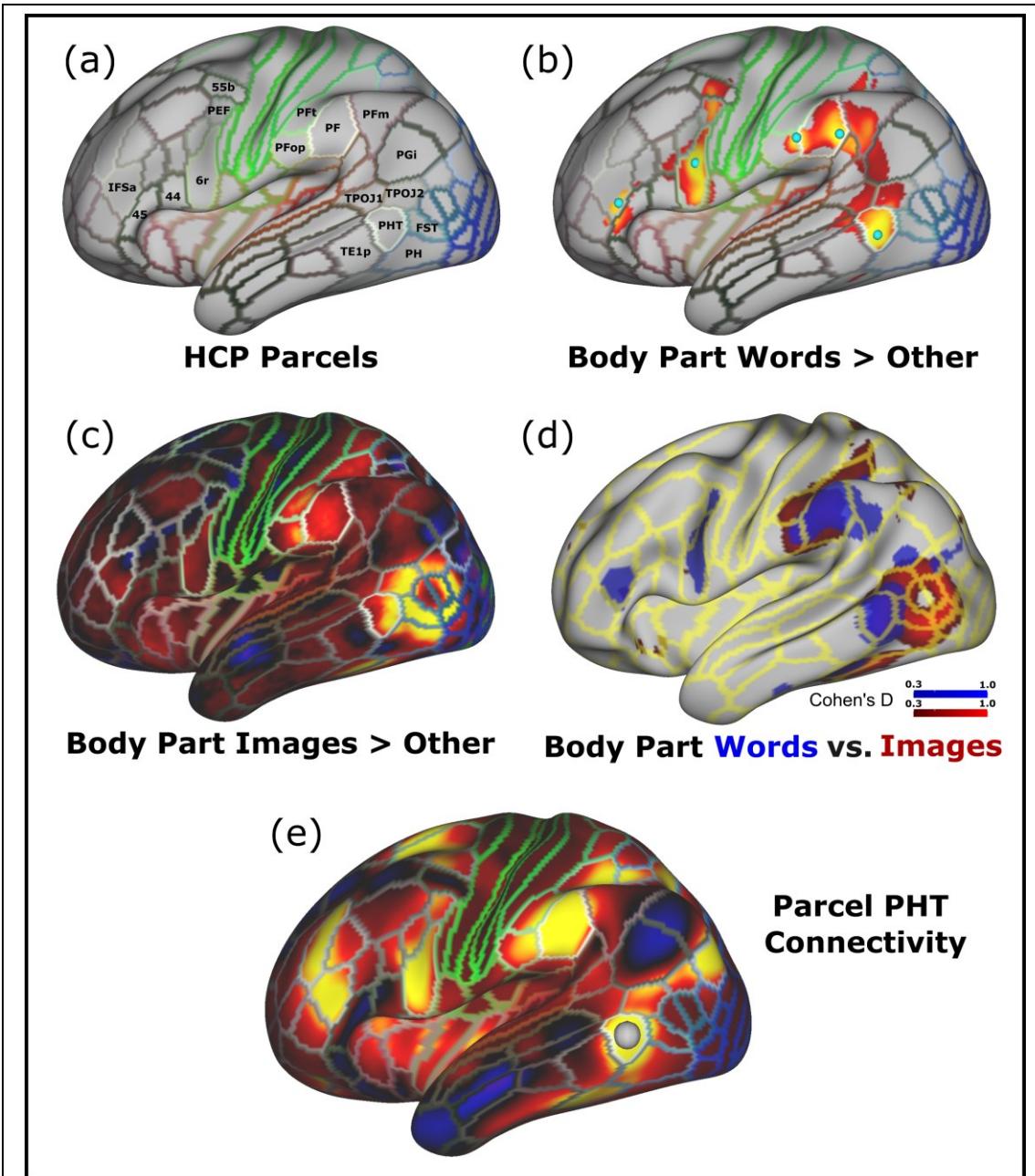
Finally, we tested whether semantic models could predict the body part selective regions found in univariate analysis by way of a vertex-wise encoding analysis. Unthresholded predicted maps resembled the observed maps using both the experiential and the distributional model (**Figure 19**), though the vertex-wise correlation between predicted and observed maps was higher for the experiential model (Spearman rho = 0.400 for experiential, 0.203 for distributional). The experiential model captured most of the features of the observed map, except for the higher activation for body parts in left inferior frontal cortex and the higher activation for non-body-part concepts in posterior left ventral temporal regions. Statistical significance was assessed at each vertex using a Wilcoxon signed rank test across all participants and using a threshold of FDR corrected

p-values < 0.01. Both models replicated the observed significant differences between body parts and other words in left pMTG and left angular gyrus. The experiential model better replicated clusters in the SMG and precuneus, whereas the distributional model better replicated clusters in the left IFG, precentral sulcus, and posterior inferior temporal gyrus.

## Discussion

Several complementary analyses were conducted to identify the neural substrates of body part word processing relative to other object categories. One region, the left pMTG, appeared in all analyses, indicating an important role for this region in preferential body part semantic representation. Broadly, the univariate analysis highlighted three regions: left pMTG, left SMG, and left precentral sulcus. These three regions also appeared in the SVM classifier analysis, the body part neural RDM reliability analysis, and the RSA results, although only the left pMTG and SMG showed higher RSA values for body parts compared to other categories. The model-based encoding analyses also predicted body part vs. other category differences in these three regions, though with some differences between the models. The close agreement between the reliability map and RSA results in Figure 2 indicates that both semantic models successfully capture shared variance across much of the cortex and gives support to the validity of the underlying semantic models. The reliability and RSA contrasts between categories identified similar regions to those that appeared in the univariate analysis, indicating converging evidence for the involvement of the relevant regions in body part representation.

The pMTG has been implicated in a variety of functions (Hodgson et al. 2022), and this general region appears to be composed of several functionally and anatomically distinct subregions (Glasser et al. 2016). As shown in **Figure 20a**, the multi-modal cortical parcellation created by the Human Connectome Project (HCP) divides the pMTG into distinct regions, including, from more anterior to posterior: TE1p, PHT, PH (which extends into posterior inferior temporal gyrus), and FST (Glasser et al. 2016). Notably, the pMTG region most strongly activated by body part words relative to other categories in our univariate analyses aligns quite closely with area PHT (**Figure 20b**). The HCP dataset also provides an opportunity to compare our results with a previously described “extrastriate body area” (EBA) that shows stronger activation to body part images than to other categories of objects (Downing et al. 2001; Downing et al. 2006; Orlov et al. 2010), as this dataset includes activation maps produced by contrasting object *pictures* from different categories, including body part images, during an n-back working memory task (Barch et al. 2013). As shown in **Figure 20c**, the region of strongest selectivity for body part images is largely posterior to PHT, involving areas FST, TPOJ2, and several extrastriate parcels surrounding the MT complex (V4t, LO3, TPOJ3), although there is also involvement of the posterior aspect of PHT. This proximity between areas showing category preferences for images and category preferences for words is consistent with a longstanding view of cortical organization in which sensory processing streams represent information in increasingly conjunctive and abstract form with greater distance from primary cortex (Damasio 1989; Mesulam 1998; Simmons and Barsalou 2003; Tanaka 1996). Recent work by Popham et al. (2021) provides similar evidence that visually- and verbally-elicited representations of object categories are aligned spatially near the



**Figure 20: HCP body part activations.** **a)** Select HCP parcels are labelled. The coloring corresponds to that used in the original report by Glasser et al. (2016). **b)** Body part word > other category univariate results overlaid with HCP parcel boundaries. **c)** Body part pictures > other category during a working memory task, univariate results from the HCP dataset. **d)** Overlap of the maps generated by body part words relative to other categories in the current study and body part images relative to other categories in the HCP dataset. The current results are displayed in Cohen's D units, the same units as the HCP data, and thresholded so that only values greater than 0.3 are displayed for both maps. **e)** Resting-state fMRI connectivity map using seed-based correlation, with a seed placed in left PHT applied to data from 1003 participants in the HCP S1200 release.

occipitotemporal border, with verbally-elicited representations immediately anterior to representations elicited by object pictures (Orlov et al. 2010).

Similarly, the SMG has been divided into several subregions based on anatomical and functional characteristics (Caspers et al. 2008; Caspers et al. 2006; Glasser et al. 2016). Our univariate contrasts of body part words compared to other categories produced activation peaks in areas PF and PFop of the HCP parcellation, though activation also involved portions of PFm, PSL, and PGi. Interestingly, the contrast of body part pictures vs. other categories in the HCP dataset also produces activation in this region (Figure 4c), but the center of mass of this activation is somewhat anterior to the activation cluster produced by body part words in our data. The two studies share a peak in the posterior aspect of PFop, but the highest peak for body part words is in posterior PF, with activation spreading more posteriorly into PFm, PSL, and PGi, none of which are activated by body part images. Body part images also activate the anterior parietal area PFt, a region not selectively activated by body part words. This partially overlapping anterior-posterior gradient is reminiscent of the posterior-anterior gradient observed in pMTG, in that body part words selectively activate cortical regions more distant from primary sensory cortex, in this case somatosensory cortex (Iwamura 1998), compared to body part images (**Figure 20d**).

The third region engaged selectively by body part words was the precentral sulcus, which has also been divided into several subregions. The largest cluster aligns very closely with HCP region 6r, which is in the rostral bank of the lower precentral sulcus. Other precentral areas activated were area PEF (the “premotor eye field”) just superior to 6r, and area 55b superior to PEF. These regions comprise rostral premotor

cortex; they lie adjacent to but “upstream” from primary motor cortex. Apart from weak activation of area 6r, these regions were not activated by body part images compared to other categories in the HCP dataset. A final activation cluster related to body part words straddled HCP areas IFSa and 45 in the pars triangularis of the IFG. The significance of this region is corroborated by the findings of Kemmerer and Tranel (2008), who reported that, among the 9 patients with left hemisphere damage and impaired naming of body parts, 8 displayed the highest degree of lesion overlap in the frontal opercular cortex.

The network of regions activated by body part words relative to other categories is remarkably recapitulated by a seed-based connectivity analysis of the HCP “resting state” fMRI dataset using area PHT as seed (**Figure 20e**). This suggests that these regions form a cohesive functional network with PHT, PF, and 6r among its principal nodes. A similar or identical network has been extensively implicated in action performance and action observation (Caspers et al. 2010; Grosbras et al. 2012; Rizzolatti and Sinigaglia 2010), except that the network activated by action performance/observation is typically bilateral and symmetric, whereas the body part word network is strongly left-lateralized. In keeping with an embodied cognition perspective, we propose that this left hemisphere network of high-level association cortices supports the retrieval of body part concepts through partial activation of relevant multimodal experiential information, including the visual shapes, sizes, and movements of body parts; tactile, proprioceptive, and relative spatial location information supporting the body schema and body structural description; and motor programs associated with particular body parts and object-directed actions. Based solely on their spatial proximity to different sensory-motor systems, we hypothesize that the SMG component of the network (areas PF and PFop) plays a larger

role in retrieving somatosensory information relevant to the body schema and body structural description, whereas the pMTG node has a larger role in retrieving visual shape and visual motion information, and the precentral sulcus a larger role in retrieving information about motor programs associated with certain body parts.

Evidence from lesion studies provides some support for these distinctions. Vascular lesions involving the EBA (Moro et al. 2008) and repetitive transcranial magnetic stimulation (rTMS) applied to the EBA (Urgesi et al. 2004; Urgesi et al. 2007) are associated with selective impairments of body part visual form discrimination without impairment of body part action discrimination. In contrast, Moro et al. (2008) found evidence linking body part action discrimination to ventral premotor cortex and posterior IFG in a lesion-symptom mapping study of 28 patients with vascular lesions. Damage in this region produced larger deficits on a body part action discrimination task compared to a body part form discrimination task. Damage to anterior parietal somatosensory cortex has long been linked with impairments of touch and haptic perception, including impairments of touch localization (autopagnosia) believed to reflect damage to a “superficial body schema” (Critchley 1953; Head and Holmes 1911; Longo et al. 2010). Somatosensory inputs also provide dynamic proprioceptive information that allows representation of body part positions and postures via a “postural schema” that stores knowledge about the dimensions of body parts (Longo et al. 2010). Schwoebel & Coslett (2005) tested patients on their ability to perform imagined hand movements and to judge the laterality (left or right) of visually presented hand pictures, both of which the authors considered to depend on these dynamic somatosensory representations of the body.

Patients with deficits on these tasks tended to have lesions involving either the left or right supramarginal gyrus.

Notably, the SMG showed greater selective body part activation in the current study for body parts that are external or movable compared to those that are internal or not movable. This finding is consistent with the claim that the SMG supports a dynamic somatosensory representation of the body, in that neither internal nor spatially fixed body parts would be expected to have a dynamic somatosensory representation. That is, internal body parts cannot be touched and so have no superficial body schema, and immobile body parts do not change position and so have little proprioceptive or postural representation.

Most relevant for identifying the critical substrate for a *spatial* representation of the body are studies of individuals with autotopagnosia. Patients with this rare disorder are unable to locate named body parts either on their own or other bodies, but typically can name individual body parts and describe their function (Pick 1922), suggesting specific damage to a spatial structural description of the body (Buxbaum and Coslett 2001). Some patients with autotopagnosia have had relatively diffuse vascular or neurodegenerative lesions (Buxbaum and Coslett 2001; Pick 1922; Sirigu et al. 1991). Those with focal damage have tended to have left parietal lobe lesions (Denes et al. 2000; Guariglia et al. 2002; Ogden 1985; Schwoebel et al. 2001; Semenza 1988), although Schwoebel & Coslett (2005) observed a marginally higher frequency of damage in the temporal lobe than in the parietal lobe in patients who were impaired on body part localization tests.

Other evidence on parietal lobe involvement in body part spatial representation comes from two functional imaging studies. Le Clec'H et al. (2000) reported activation in the left intraparietal sulcus (IPS) and several left frontal regions during a spatial judgment task on body part names ("Is it above or below the shoulders?") compared to a size judgment task on numbers ("Is it less than 12?"). The lack of comparison with other body part tasks in this study, however, leaves open the question of whether these activations are specific to the spatial judgment task. Corradi-Dell'Acqua et al. (2008) compared tasks requiring either relative distance judgment or name retrieval with pairs of body part or building part images. A focus in the left posterior IPS showed stronger activation for the distance judgment task than the name retrieval task, but only for body parts, suggesting selective processing of body part location information in this area. We observed evidence for selective involvement of anterior portions of the IPS (HCP parcels AIP and IP2) in body part word processing in the univariate contrast, classifier, and intra-class correlation analyses (Figures 1 and 2). The IPS was also implicated in the semantic representation of body parts by a prior study that used body part images and whole-brain searchlight RSA with a distributionally derived semantic model (Bracci et al. 2015), although those authors did not compare activation in the IPS for body parts relative to other categories.

Other evidence suggests that the pMTG may also store more abstract semantic knowledge related to body parts. Schwoebel and Coslett (2005) found 16 patients (among a sample of 70) who had deficits on matching body parts by function and matching body parts to associated clothing and objects, both tasks designed to assess higher-level semantic knowledge about body parts. Of the 13 with available imaging data, 12 had lesions involving the temporal lobe. Three of these had deficits confined to these

semantic measures, without impairment on lower-level body schema (action imagery and hand laterality judgments) or body part localization tasks, and the lesions in these patients overlapped mainly in the left posterior temporal lobe. The pMTG (MNI: -57, -54, -10) was also implicated in the semantic representation of body parts in the study by Bracci et al. (2015).

Studies on regions involved in the semantic processing of verbs have also repeatedly implicated the pMTG and surrounding area. There is an inherent connection between body parts and verbs given that actions are performed using body parts. An ALE meta-analysis (Faroqi-Shah et al. 2018) of 23 studies contrasting verb vs. noun processing (33 contrasts, 190 activation foci) showed a reliable activation cluster for verbs greater than nouns in the left pMTG (MNI: -56, -46, 7) near our reported peak, as well as smaller clusters in left IFG (MNI: -42, 21, 5) and premotor eye field (MNI: -48, 6, 40) close to those observed in our univariate analysis. Thus, the network we associate with preferential processing of body part knowledge overlaps to a large extent with the network of regions previously implicated in verb > noun processing. Motion is also an intrinsic component of many verb concepts. A study on the representation of verbs found that of five verb classes containing a motion component, all had overlapping activation in the posterior lateral temporal cortex (Figure 3; Kemmerer et al. 2008). Given the inherent correlations between body part concepts, actions represented by verbs, and motion content of verbs, comparing the differential contributions of these knowledge types to activation in these regions will be challenging. It seems likely to us that this network represents all of these inter-related forms of knowledge, which collectively contribute to our understanding of body movements and body-object interactions.

The data in Goodglass et al. (1966) and a thorough review by Capitani et al. (2003) of reported category-selective deficits indicate typical sparing of body part knowledge relative to other categories of concrete object concepts in brain-damaged individuals. There are at least three possible reasons for this typical sparing. First, the evidence presented by Schwoebel and Coslett (2005) suggests that the pMTG and adjacent posterior ITG may be the most critical zone for supporting body part concepts. This zone lies at the posterior edge of the middle cerebral artery territory and is often spared in stroke. A second possibility is that body part concepts may have a more distributed (i.e., redundant) representation compared to other categories and thus be more resilient to focal damage. Finally, many body part concepts (e.g., hand, arm, leg, head, mouth, eye, nose) are acquired at an early age and are highly familiar. Body part knowledge might therefore appear to be more resilient to brain damage when items are not matched on these variables. These accounts are not mutually exclusive, and all may be true to some degree. Our results favor a key role for the preferential representation of body parts in the left pMTG, consistent with the lesion evidence implicating this region (Schwoebel and Coslett 2005).

We observed body part > other category effects in the left precentral sulcus in several model free analyses (univariate contrasts, classifier contrasts, neural RDM reliability contrasts) that does not appear to be well accounted for in the semantic models, as it did not show a significant category effect in RSA and only weakly in the encoding results. This could potentially be the result of a feature or set of features being represented in this region but not accounted for in the semantic models. Alternatively, the features represented in this region might be as relevant for distinguishing between

members of other categories as they are for distinguishing body parts, resulting in RSA values for one of the other categories that are as high as for body parts. For example, if this region mainly represents motor programs associated with actions, as we hypothesize, then it may be just as important for representing the similarity structure among tool concepts (which differ markedly in the actions associated with their use) as the similarity structure among body part concepts.

Defining category effects on brain activation patterns entails comparisons between categories, and the results naturally depend on the range and specific type of categories that are included for comparison. A limitation of the current study is that it involved just four categories, leaving open the possibility that some other category might exist that activates the identified areas as much as body part words. We believe this is unlikely given that two of the categories we included – tools and food/plants – are associated with object-directed actions and thus conceptually related to body parts, and the tool category, in particular, is known to preferentially activate closely adjacent or overlapping regions of the pMTG. In addition, many exemplars in the animal category have body part features (head, mouth, eyes, arm, leg, torso, etc.) not dissimilar from human body parts. Another general limitation of fMRI studies is the variety of analysis techniques available and the inherent degrees of freedom this introduces. We believe our study mitigates this by focusing on the convergent results across four very different analysis approaches. Given that the regions we identified showed stronger responses to body part words in comparison to other categories across all these analyses, and given the salience of body parts to many aspects of daily life, we believe the present evidence supports the claim for a brain network that is relatively specialized for representing body

part knowledge. Future research should examine comparisons between body part concepts and a wider variety of other categories.

## Conclusions

Our results show that lexical-semantic representations of body parts are distributed in a multi-lobar, left-lateralized network primarily involving posterior temporal, inferior parietal, and precentral cortex. Two commonly used models of semantic content accounted for within-category representational geometry and for univariate contrasts implicating the pMTG and SMG in body part representation. The results are broadly consistent with available neuropsychological evidence from individuals with focal brain damage and contribute to a developing mechanistic account of body part knowledge impairments.

## Chapter 5: Selected Experiential Feature Maps

### Introduction

An overarching goal of cognitive neuroscience is to understand the computations subsumed by different regions of the cortex. As elaborated on in the introduction to this dissertation, experiential models of semantics are founded on the idea that semantic representations are built up from the representations generated in the perceptual neural processors (e.g., MT processing visual motion, gustatory cortex processing taste).

Grounded models of cognition posit multimodal convergence zones as being central in the representation of semantic content. While previous chapters of this dissertation have shown that the CREA model can act as a common representational basis for concepts that span across categories, the hypothesis that perceptual neural processors overlap with and give rise to organizational constraints of the semantic system has not yet been examined in this dissertation. This chapter focuses on analyzing the regions of overlap and spatial distributions for select features from the CREA model for which there are strong prior expectations about localization.

The idea that the function of primary sensory cortices is given by their connection to the periphery, and that the function of ‘higher level’ cortex is given by its connection to primary cortex, is far from new in neuroscience<sup>13</sup>. It lies at the heart of grounded cognition models and, importantly, implies clear testable predictions. For example, to quote a seminal paper by Mesulam “Transmodal areas in the midtemporal cortex,

---

<sup>13</sup> For example, in 1882 Hughlings Jackson wrote: “The very highest of all nervous centres are but complex rearrangements of lower centres, and these of still lower centres unto the lowest, which last directly represent impression and movements” (An introduction to the life and work of john hughlings jackson: Introduction 2007).

Wernicke's area, the hippocampal-entorhinal complex and posterior parietal cortex provide critical gateways for transforming perception into recognition, *word-forms into meaning [...]”* (emphasis added; Mesulam 1998). One idea identified with grounded cognition models is that the more distance there is between sensory afferent and motor efferent neurons, the more flexibility there is to have a complex relation between sensory input and motor output. Extrapolation of this view leads to the prediction that regions of the brain higher up on the sensory-motor hierarchy should contain more complex, semantic representations. As this prediction relates to the experiential model examined in this dissertation, this would amount to the expectation that the peripheries of the semantic system should overlap with lower-level perceptual cortex, and that more central regions of the cortical hierarchy should represent a larger number of experiential features.

Critical for testing this hypothesis is an assessment of ‘distance’ from primary sensory-motor cortices. One network of regions that is commonly considered to meet the criteria of distant or relatively ‘disconnected’ from the primary sensory-motor cortices is the so-called ‘default mode network’ (DMN). While fMRI is often used to map regions of the brain that become more active during a particular task, one early observation on fMRI research was a consistent network of regions that appear to be *de-activated* by many tasks. These regions were sometimes called the ‘task negative network’ but they are now more commonly known as the DMN. This network of regions has come under intense study in part due to their robust identifiability in task free or ‘resting state’ fMRI data.

Several papers implicated the regions of the DMN at the top of the hierachal structure in the brain. Within the papers that support this view, one that provides a

particularly parsimonious account defines and visualizes what the authors term as the “principal gradient” of the cortex (Margulies et al. 2016). The principal gradient represents a decomposition of functional connectivity data that can be thought of as applying a non-linear version of principal component analysis to fMRI resting state connectivity data. The result of their analysis is a cortical gradient “that situates the default-mode network at the opposite end of a spectrum from primary sensory and motor regions.” (Margulies et al. 2016). In an analysis that analyzes this gradient with respect to distances along the surface, the authors write:

“These DMN regions exhibit the greatest geodesic distance along the cortical surface – and are precisely equidistant – from primary sensory/motor morphological landmarks. The principal gradient also provides an organizing spatial framework for multiple large-scale networks and characterizes a spectrum from unimodal to heteromodal activity in a functional metaanalysis”

This remarkable finding offers strong support for a view that “the contribution of the DMN to cognition relates to its physical and functional distance from the sensory and motor systems” (Smallwood et al. 2021).

While the functional role of DMN regions is not entirely clear, one early account posits that the regions are “engaged almost continuously in adaptive processes involving semantic knowledge retrieval, representation in awareness, and directed manipulation of represented knowledge for organization, problem-solving, and planning” (Binder et al. 1999). From this view, it follows that if these processes are “interrupted or inhibited by task performance”, that the DMN would exhibit task-induced deactivation (Binder et al. 1999). Binder and colleagues found support for this view when testing whether a

semantic task activates the DMN (Binder et al. 1999), and very strong support was also implicit in a seminal meta-analysis by Binder and colleagues (2009). Quoting their description of the observed general-semantic network, Binder and colleagues state “the semantic system identified here is also strikingly similar to the human brain ‘default network’” (Binder et al. 2009). Elaborating on a similar idea, Fernandino and Binder (2024) suggest that the DMN “implements an integrated simulation of phenomenological experience – that is, an embodied situation model – constructed from various modalities of experiential memory traces.” Based on this background, we predict that there will be a hierarchical convergence of experiential features, with a larger number of experiential features being represented closer to the peaks of the principal gradient.

### Generating Feature Maps

There are several approaches one can use to create feature maps to understand where the contents of a model are represented. Throughout this chapter, feature maps are defined as regions of cortex whose degree of activation is (positively) modulated by the experiential content associated with different words. An implicit assumption in this analysis is that reading words evokes relevant experiential content, but this hypothesis is already strongly supported by the previous chapters of this dissertation. As already used throughout the previous chapters of this dissertation, vertex-wise encoding models are a popular approach to examine where information is represented, that capitalize on the spatial resolution of the data and have direct interpretability. However, other approaches, such as searchlight RSA, also exist. While searchlight approaches introduce additional algorithm choices into the analysis (e.g., the searchlight size, distance metric used to generate RDMs, and correlation metric used to compare RDMs), several papers have

used this approach to examine where some semantic features are represented (Liuzzi et al. 2020; Peelen et al. 2014). We note that our results may differ with respect to those obtained using a combination of multiple regression and RSA (sometimes called feature reweighted RSA). We avoid an RSA based approach due to the aforementioned additional algorithm choices as well as interpretational difficulties such as obscuring the direction of univariate effects. For example, if a feature has a significant positive RSA correlation in a region, the underlying univariate activation could be *either* positively or negatively correlated with the relevant feature. To focus on interpretability and model first order effects, we chose to use vertex-wise encoding models to directly model fMRI time series using an estimation of the experiential content in each presented word.

Following the choice to use vertex-wise encoding models, another analysis decision is whether to examine features on their own, independent of other features in a model, or whether *marginal* contributions of each feature should be assessed. Given that our focus is on the *specific* neural substrate for each experiential feature, we focus on multivariate results so that each map is more specific to a given feature and less likely to be the result of spurious correlation. However, for the cases where no vertices were significant in a multivariate analysis, follow-up analysis was performed using a single feature as a predictor. While having multiple predictors that have some correlation with each other can increase the variance in parameter estimates, this problem was mitigated by using cross validated ridge regression.

## Feature Selection

One key component of this chapter is the *interpretation* of experiential feature maps. Homing in on the task of interpretation, one prerequisite for interpreting regions of

activation is knowledge about the relevant neural processer. The components of the CREA model differ with respect to how much prior literature there is regarding relevant functional localizers. For example, some components such as “Scene” have strong prior expectations about the involved regions, but some such as “Music” have less. As a result, this chapter assesses 14 experiential features for which there are strong priors, each of which is motivated below. On top of the theory-driven benefits of having a smaller feature set, an additional benefit of reducing the feature set is increased power to detect smaller effect sizes for the retained features.

Several previous studies have taken a similar approach to testing the embodied hypothesis that is taken in this chapter. Two studies closely related to the work done in this chapter are a study by Fernandino and colleagues who investigated 5 experiential features (Fernandino et al. 2016), and a separate study that examined 6 experiential features (Lin et al. 2024). Where relevant, their results are discussed in more detail below, but the general finding in each study was that many feature maps had an interpretable pattern in light of experimental data on the neural substrates for perceiving these features. The primary contribution of the work undertaken here is to simultaneously examine a larger and different feature set than has been used in prior analyses. Further, our analysis uses a large dataset with 81 participants, 522 unique stimuli, and except for three participants, 6 stimulus repetitions. Each of the selected CREA features is motivated below.

*Motion* - “*Showing a lot of visually observable movement*”: The study of motion processing in the cortex has a rich history, with a region of the brain known as MT or V5

thought to process motion (Grill-Spector and Malach 2004; Maunsell and Essen 1983)<sup>14</sup>.

Given the relatively well characterized and selective response of this region to visual motion, it has been the subject of several studies investigating grounded cognition. One study found:

“Retrieval of conceptual information from action pictures causes greater activation than from object pictures bilaterally in the human motion areas (MT/MST) and nearby temporal regions. By contrast, retrieval of conceptual information from action words caused greater activation in the left middle and superior temporal gyrus, anterior and dorsal to the MT/MST” (Kable et al. 2005).

Another study that scanned participants while they performed semantic judgments about sentences describing motion or static events found “the more posterior parts of LOTC, including motion perception cortex, respond differently to motion vs. static events.” (Zhang et al. 2022). An ALE metaanalysis by Kuhnke et al. (2023) examined motion-related conceptual processing ( $n=24$ )<sup>15</sup> and observed left hemisphere activation peaks in the IFG, pSTS, pMTG, and aSMG.

Of note, one previous study examined a visual motion feature similar to the current definition and used a wide-ranging set of 900 stimuli (Fernandino et al. 2016). The authors found that “the typical location of area MT+ [...] was not modulated by the visual motion ratings”. Instead, motion-related activation was observed in “a more dorsal

---

<sup>14</sup> As with most complex phenomena, there are likely many regions in the brain involved with processing motion. For example, Grill-Spector (2004) write: “Along the dorsal stream many regions are activated more strongly when subjects view moving versus stationary stimuli. Included are areas MT, MST, V3a, and even low-level areas such as V1 and V2.”

<sup>15</sup> This is the “n” where “n represents the number of experiments per modality (where one article could contain multiple experiments)” (Kuhnke et al. 2023)

region located in the ventrolateral portion of the left angular gyrus". Provided this background, our hypothesis is that the motion feature will likely be represented in posterior MTG, inferior parietal cortex, or in relatively anterior lateral occipitotemporal cortex (LOTC).

Color - "*Having a characteristic or defining color*": Color perception is generally thought to involve a specialized area of the visual hierarchy called V4<sup>16</sup> (Bartels and Zeki 2000; Lueck et al. 1989; McKeefry and Zeki 1997; Zeki et al. 1991)<sup>17</sup>. Similar to motion, color perception has received significant attention in the cognitive neuroscience literature, especially with regard to the grounded theories of cognition. This is in part due to it being considered a strongly 'unimodal' feature, in contrast to some other 'multi-modal' features (e.g., texture and shape, which are both somatosensory and visual). Previous studies have investigated regions involved in the representation of color knowledge using various tasks. Some tasks have included: assessing hue perception through a color sequencing task (Beauchamp et al. 1999)<sup>18</sup>, having participants name the color associated with different (achromatic) objects presented as line drawings (Chao and Martin 1999; Martin et al. 1995)<sup>19</sup> or as words (Goldberg et al. 2006b; Martin et al. 1995; Simmons et al. 2007), and chromatic judgment tasks using words (Hsu et al. 2012)<sup>20</sup>. The overall result from many of the studies in the vision literature has been either overlapping

---

<sup>16</sup> While area V4 is the area classically associated with color processing, it is not the only region as previous work "has established that the colour centre in the fusiform gyrus is one part of a cortical colour processing system that extends from the striate cortex through area V2 to V4, and beyond to the inferior temporal cortex" (Bartels and Zeki 2000). Some texts also highlight the role of smaller region anterior to area V4 that is termed V4α, and use V4 to refer to the posterior component of the complex (Bartels and Zeki 2000).

<sup>17</sup> Related to the previous discussion of color-agnosia, there was a report by Swiss ophthalmologist Louis Verrey (1888) of a "center for the chromatic sense" in the lingual and fusiform gyri (Zeki 1993)

<sup>18</sup> This is a modified version of the clinically used Farnsworth-Munsell 100-Hue test

<sup>19</sup> Sometimes called a 'property verification' task

<sup>20</sup> E.g., Which item is lighter? A lemon or a basketball

or adjacent activation between color perception and color ‘knowledge’ regions. An ALE metaanalysis on color conceptual knowledge (n=12) found left hemisphere activation peaks in the ventral fusiform gyrus (FG) and IPS (Kuhnke et al. 2023).

Color was also included as a feature in the analysis by Fernandino and colleagues (2016), who found that “the color attribute modulated activity in the left ventromedial occipital cortex at the junction of lingual and fusiform gyri” (Fernandino et al. 2016). Given the large variation in results, the network involved in color perception is difficult to predict, but it is expected to include ventral temporal or occipitotemporal regions.

Complexity - “*Visually complex*”: While visual complexity is a feature that is not typically localized, it was included due to the expectation that early visual regions might activate more to stimuli that are rated higher on degree of associated visual complexity.

Face - “*Having a human or human-like face*”: One region that has shown increased activation in response to human faces relative to other objects is known as the fusiform face area (FFA; Downing et al. 2001; Kanwisher et al. 1997). This region has been the topic of many studies in the visual sciences. Most studies of face processing have used picture stimuli, with fewer using lexical stimuli. Of those studies that have used names as stimuli, one reported activation in bilateral hippocampus and parahippocampal gyrus “during recognition of familiar names [...] compared to non-famous names” (Douville et al. 2005), and a different study found that “bilateral temporal poles and anterolateral temporal cortices, as well as the left temporoparietal junction” were more activated for familiar and famous names relative to unfamiliar names (Sugiura et al. 2006). However, no studies have specifically assessed Face knowledge using lexical stimuli. Our expectation is that Face should involve regions around the FFA and, due to

the implicit social content often involved, perhaps anterior superior temporal lobe (Olson et al. 2013; Zahn et al. 2007).

Sound - “*Having a characteristic or recognizable sound or sounds*”: Primary and secondary auditory cortex are extensively studied in human neuroscience. However, while visual, somatosensory, and motor cortices have seen extensive testing with respect to the grounded cognition hypothesis, sound features have seen less interrogation. Sound was selected as a feature because there is evidence that damage to auditory association cortex is associated with impairment in the processing of sound related concepts (Trumpp et al. 2013). A Sound feature was examined in the study by Fernandino and colleagues (2016). Those authors found “the sound attribute was associated with robust activation of the left ventrolateral prefrontal cortex”. An ALE meta-analysis of conceptual sound processing (n=16) found that “Sound-related conceptual processing consistently activated the bilateral pSTS (extending into pMTG in the left hemisphere), as well as left pSMG and dorsomedial prefrontal cortex (dmPFC)” (Kuhnke et al. 2023). Despite the variation in regions that have been implicated in processing semantic Sound features, we expect the Sound feature to implicate pMTG and pSTS regions, which are implicated in high-level auditory perceptual processing.

Taste - “*Having a characteristic or defining taste*”: The primary cortex involved in the perception of taste is thought to be the insula (Avery et al. 2019; Small 2010). Drawing on a study that used picture stimuli, the authors found that “compared to location pictures, food pictures activate the right insula/operculum and left orbitofrontal cortex” (Simmons et al. 2005). Related to taste, another study found that, compared to reading neutral language items, “odour-related terms elicited activation in the primary

olfactory cortex” (González et al. 2006). Another study that examined the grounding of “words whose meaning is primarily related to taste” found specific activation in the “anterior insula, frontal operculum, lateral orbitofrontal gyrus, and thalamus among others” (Barrós-Loscertales et al. 2012). An ALE metaanalysis (n=17) found that “Conceptual processing related to olfaction-gustation was associated with consistent activation in left orbitofrontal cortex (OFC)” (Kuhnke et al. 2023). This then suggests an important role for the insula and OFC regions in the semantic representation of foods. The Taste attribute is expected to involve activations in the insula region, and orbitofrontal cortex.

*Action:* Many studies have investigated the brain networks that underlie action observation and imitation, and one ALE meta-analysis of this literature found an extensive bilateral network involved in action observation (Caspers et al. 2010). The primary regions implicated in both observation and action were left pMTG, left IFG (BA 44 and BA 6), left PMC, left SMA, and left IPS. Primary motor and motor association cortices have received significant attention regarding their centrality to the grounded cognition hypothesis. Writing in 2015, David Kemmerer said that “one of the most controversial issues in the cognitive neuroscience literature on concepts is whether the motor features of verb meaning are represented in the precentral motor cortices” (Kemmerer 2015). While some studies have shown activation in motor cortices, with a particularly famous paper showing somatotopic representation of action words in motor and premotor cortex (Hauk et al. 2004), the evidence is mixed as involvement is often dependent on the task. The study by Fernandino and colleagues included a Manipulation attribute that is similar to the action attribute here. Those authors found “A single cluster

located at the junction of the left posterior MTG (pMTG) and anterior occipital cortex was modulated by the manipulation rating”. An ALE study of brain activation associated with action concepts (n=74) found that:

“Across neuroimaging studies, action-related conceptual processing consistently engaged the left inferior frontal gyrus (IFG) and premotor cortex (PMC), anterior supramarginal gyrus (aSMG) extending into primary somatosensory cortex (S1) and intraparietal sulcus (IPS), the lateral temporal-occipital junction (LTO) including parts of posterior middle and inferior temporal gyri (pMTG/ITG) and superior temporal sulcus (pSTS), as well as the bilateral (pre-)supplementary motor area (SMA)” (Kuhnke et al. 2023)

Given the extent of this network, it is thought that all of these regions should be involved in the semantic representations of relevant nouns.

The CREA model decomposes the motor action domain into several constituent features. Three of these features, representing different somatotopic effector locations, were selected for analysis. While somatotopy in the primary motor and sensory cortices dictate the expectations for where these features will localize in lower sensory-motor cortex, it is unknown whether higher level association cortices maintain somatotopy at a detectable resolution. Further, covariance among the motor features might mask regions that are less distinctly representing parts of the homunculus.

*Head* – “Associated with actions using the face, mouth, or tongue”

*Lower Limb* – “associated with actions using the leg or foot”

*Upper Limb* – “associated with actions using the arm, hand, or fingers”

Number - “Associated with a specific number or amount”: Closely related to quantity, numbers are used to quantify items in our environment. Some of the earliest evidence for the localization of deficits related to numbers is the description of Gerstmann syndrome, in which patients have acquired calculation deficits (acalculia) following damage “to the angular gyrus in its transition to the second occipital convolution” (Gerstmann 1940). In more recent times, many more neuroimaging studies have been performed, and one recurrent finding has been “the intraparietal sulcus as key node for the representation of the semantic aspect of numerical quantity” (Nieder and Dehaene 2009). As such, we expect the IPS to show involvement in the representation of this feature.

Scene - “Bringing to mind a particular setting or physical location”: The study of scenes has a rich history and is typically understood in contrast to object perception. That is, “whereas objects are spatially compact entities that one acts upon, scenes are spatially distributed entities that one acts within” (Epstein 2005). Scenes have been a popular topic of study and but in brief, one review states “neuroimaging studies have identified three cortical regions that respond selectively to scenes: parahippocampal place area, retrosplenial complex/medial place area, and occipital place area” (Epstein and Baker 2019). Consequently, we predict increased activation in these areas with increased “scene” association ratings.

Social - “An activity or event that involves an interaction between people”: It is sometimes said that humans are social creatures, and that operating in the world requires navigating social environments. It is generally thought that large parts of cortex are involved in social-semantic processing, which has been the topic of many studies. The

studies that have focused on the representation of lexicalized social concepts have consistently reported bilateral superior anterior temporal lobe involvement (Pobric et al. 2016; Zahn et al. 2007).

Similar to the ‘Social’ feature in CREA, one fMRI study analyzed activation differences between verbs based on their rated sociality<sup>21</sup>. The authors found a sociality effect in anterior superior temporal sulci, angular gyri, dorsomedial prefrontal cortex, and left PCC (Lin et al. 2018). A separate study that also examined the representation of a similarly defined social feature found “the socialness effect in the anterior lateral temporal lobe, temporoparietal junction, medial prefrontal cortex, posterior cingulate gyrus, and precuneus” (Lin et al. 2024).

Related to social processing, and considerably more studied, is the capacity to understand the mental states of other people, which is often termed ‘theory of mind’ (ToM; Gallagher and Frith 2003; Mar 2011). Quoting from Gallagher and Frith: “Three areas are consistently activated in association with theory of mind. These are the anterior paracingulate cortex, the superior temporal sulci and the temporal poles bilaterally.” Given this background, we expect an expansive network of regions including at a minimum the bilateral anterior temporal lobes, superior temporal sulci and angular gyri, and dorsomedial prefrontal cortex to be implicated in representing the Social feature.

*Valence* – “Someone or something that makes you feel happy or sad”: A prominent theory of emotional behavior divides emotion along two dimensions: valence (from positive to negative) and arousal (from low to high) (Russell 1980). While there is

---

<sup>21</sup> “Participants were asked to rank verbs on a 5-point scale according to how often an event that a verb refers to involves interaction between people.”

some evidence that these may be dissociable (Anders et al. 2004), many more studies have examined valence relative to arousal. In contrast to the other features considered so far, this feature was decomposed into two prompts, with the difference between them (i.e., Happy - Sad) used as the marker of valence. Studies that have examined the neural correlates of fear (which has negative valence) have had mixed results, with some evidence that “subjective fear [...] is encoded in distributed systems rather than isolated ‘fear centers’” (Zhou et al. 2021). Although fear conditioning is a separate topic, the association between the fear response and amygdala is perhaps one of the most commonly recognized structure-function associations in neuroscience. An ALE metaanalysis the processing of emotional concepts (n=27) reported that “processing emotional concepts was associated with consistent activity in the vmPFC, alongside the left amygdala and pSTS/TPJ” (Arioli et al. 2021). A separate ALE metaanalysis also examined the neural correlates of emotion-related conceptual processing (n=47) and found that it “consistently activated the bilateral amygdala and medial prefrontal cortex (mPFC), as well as the left temporo-parietal junction (TPJ) / angular gyrus (AG) and temporal pole (TP)” (Kuhnke et al. 2023).

A previous study used a similar emotion feature to the one used in this study (Lin et al. 2024). The primary difference was that valence was calculated as the absolute value of the Happy and Sad feature ratings. Those authors found that the results on the emotion dimension were “weak and inconsistent”, the general effects were found in “the medial prefrontal cortex and inferior parietal lobule” (Lin et al. 2024). We expect valence to be represented in limbic regions, possibly including medial prefrontal or orbitofrontal cortex, posterior cingulate gyrus, and periamygadalar regions of the ATL.

Shape – “Having a characteristic or defining visual shape or form”: This is a broadly defined attribute whose processing is thought to involve several large regions of cortex. Shape is important for the perception and identification of objects and large regions of the cortex have been implicated in object recognition, sometimes referred to as object selective regions (Grill-Spector and Malach 2004). One of these implicated regions is called the lateral occipital complex (LOC), and it has been shown to respond more to “images of objects, compared to a wide range of texture patterns” (Malach et al. 1995)<sup>22</sup>. Other places where the Shape attribute might be represented include regions of the ventral occipito-temporal (VOT) cortex that have shown selectivity to faces (Kanwisher et al. 1997), places (Epstein and Kanwisher 1998), and other categories of man-made objects (Haxby et al. 2001).

An analysis of a similar Shape attribute by Fernandino and colleagues (2016) found widespread activation in regions that included “the ventral occipitotemporal (VOT) cortex and the lateral occipital complex (LOC)” as well as “the superior occipital gyrus and adjacent posterior intraparietal sulcus”. One likely reason for this widespread activation is that Shape is a somewhat non-specific feature that is thought to correlate with other features. For example, quoting from Binder et al. (2016) “Shape information is generally the most important attribute of concrete object concepts and distinguishes concrete concepts from a range of abstract (e.g., affective, social, and cognitive) entities and event concepts.” While collinearity often be problematic in regression models (discussed below), the Shape feature was included as it can also help to “absorb” non-

---

<sup>22</sup> Emphasizing the Shape processing that occurs in this region, the authors write “A striking demonstration that activity in LO is uniquely correlated to object detectability was produced by the "Lincoln" illusion, in which blurring of objects digitized into large blocks paradoxically increases their recognizability. Such blurring led to significant enhancement of LO activation.”

specific variance. This can help to increase the specificity of the results for the remaining features.

## Materials and Methods:

*Experiential features:* All pairwise Pearson correlations were calculated using the 522 unique stimuli for the chosen set of 14 features. To better understand the degree of collinearity amongst all predictors, variance inflation factor (VIF) scores were calculated as well using the python statsmodels toolbox (Seabold and Perktold 2010). A VIF score for a given predictor (in our case, a feature) is a measure of how well a linear combination of the other predictors can estimate the target predictor. That is, one can imagine a leave-one-out analysis amongst the predictors of a model where a single left out predictor is estimated by all remaining predictors. A commonly used ‘rule-of-thumb’ is that VIF scores above 10 indicate an excessively high degree of collinearity amongst the predictors (O’Brien 2007). Having a high degree of collinearity results in more instability amongst the estimated beta values, which can also be interpreted as increased susceptibility to noise.

*Participants:* The participants are similar to those included in previous chapters. Motivated by results from chapter 2 of this dissertation, three additional participants were included who only completed two scanning sessions (i.e., 4 presentations of stimuli instead of 6). In total, 41 participants from study 1 and 42 participants from study 2 were included for analysis.

*Stimuli and task:* The stimuli and task were the same as used in previous chapters of this dissertation.

*FMRI processing:* Pre-processing of fMRI data was the same as previous chapters. However, post-processing of data differed. In contrast to previous chapters where fMRI processing used AFNI's 3dREMLFIT with a canonical HRF to model word activation in each vertex, this chapter instead directly estimates features maps on surface time-series data using cross-validated ridge regression. This was done by performing two sequential regressions. First, volumes that had an associated framewise displacement greater than .9 mm or that were not at steady-state, were censored and removed from analysis. Following censoring, the time-series data had nuisance variables controlled for by ordinary least squares regression. Nuisance variables were 6 motion regressors, their derivatives, CSF, and 4<sup>th</sup> order baseline cosine regressors. The residuals from the first regression were then z-scored and used as targets to generate feature maps.

Feature maps were generated by creating a finite impulse response (FIR) model for each of the 14 experiential features. The FIR model was made using the Nilearn python package and consisted of a set of 4 delayed Dirac functions with delays of 1, 2, 3, and 4 repetition times (TRs). Coefficients for each feature and delay were calculated using cross-validate ridge regression using the python Himalaya package (Tour et al. 2022). The ridge penalty was estimated using cross validation because each of the stimuli were presented multiples times in each study. For example, for most participants estimates were fit using 5 presentations of all stimuli, and the runs of the left-out stimulus presentation were used to estimate the quality of the fit. The ‘univariate’ feature maps were generated using the 4 lagged predictors for a single feature. For each feature, for each vertex, sensitivity to a feature was determined by averaging the 4 coefficients together. Significance at the group level was determined by first smoothing the individual

feature maps with a 4 mm FWHM kernel and then testing the averaged beta-coefficient against zero using a 1-sided t-test with significance set as FDR corrected  $p < .01$ . All visualized results were set to have a minimum cluster size of  $60 \text{ mm}^2$ .

## Results:

|                   | Color | Complexity | Shape | Face  | Sound | Taste | Head  | Upper Limb | Lower Limb | Number | Scene | Social | Valence |
|-------------------|-------|------------|-------|-------|-------|-------|-------|------------|------------|--------|-------|--------|---------|
| <b>Motion</b>     | 0.03  | 0.70       | 0.27  | 0.19  | 0.51  | -0.32 | -0.47 | -0.06      | 0.48       | -0.13  | 0.42  | -0.05  | 0.01    |
| <b>Color</b>      |       | 0.22       | 0.73  | -0.04 | -0.21 | 0.62  | 0.05  | 0.12       | -0.08      | 0.03   | 0.14  | -0.59  | 0.19    |
| <b>Complexity</b> |       |            | 0.40  | 0.14  | 0.23  | -0.14 | -0.42 | -0.07      | 0.32       | 0.00   | 0.52  | -0.11  | 0.10    |
| <b>Shape</b>      |       |            |       | -0.06 | -0.13 | 0.25  | -0.24 | 0.32       | 0.12       | 0.13   | 0.29  | -0.69  | 0.17    |
| <b>Face</b>       |       |            |       |       | 0.14  | -0.16 | 0.12  | -0.17      | 0.09       | 0.05   | -0.05 | 0.22   | 0.02    |
| <b>Sound</b>      |       |            |       |       |       | -0.32 | -0.10 | -0.15      | 0.09       | -0.32  | 0.11  | 0.06   | -0.08   |
| <b>Taste</b>      |       |            |       |       |       |       | 0.47  | 0.13       | -0.24      | -0.07  | -0.05 | -0.28  | 0.26    |
| <b>Head</b>       |       |            |       |       |       |       |       | 0.01       | -0.33      | -0.17  | -0.23 | 0.23   | 0.10    |
| <b>Upper Limb</b> |       |            |       |       |       |       |       |            | 0.12       | 0.03   | 0.12  | -0.05  | 0.16    |
| <b>Lower Limb</b> |       |            |       |       |       |       |       |            |            | 0.05   | 0.22  | 0.09   | 0.08    |
| <b>Number</b>     |       |            |       |       |       |       |       |            |            |        | -0.02 | -0.15  | 0.04    |
| <b>Scene</b>      |       |            |       |       |       |       |       |            |            |        |       | 0.05   | 0.19    |
| <b>Social</b>     |       |            |       |       |       |       |       |            |            |        |       |        | 0.07    |

**Table 9: Correlation between feature ratings.** The table shows the Pearson correlation coefficients between each pair of the fourteen features included in the analysis. The color scale is centered on 0.

Shown in **Table 9**, while the average correlation between features was low, several of the features included in the analysis had high correlations with each other. Collinearity amongst the features is quantified in **Table 10**. The feature most colinear with the other features was Shape (5.48) and all remaining features had VIF scores below 5.

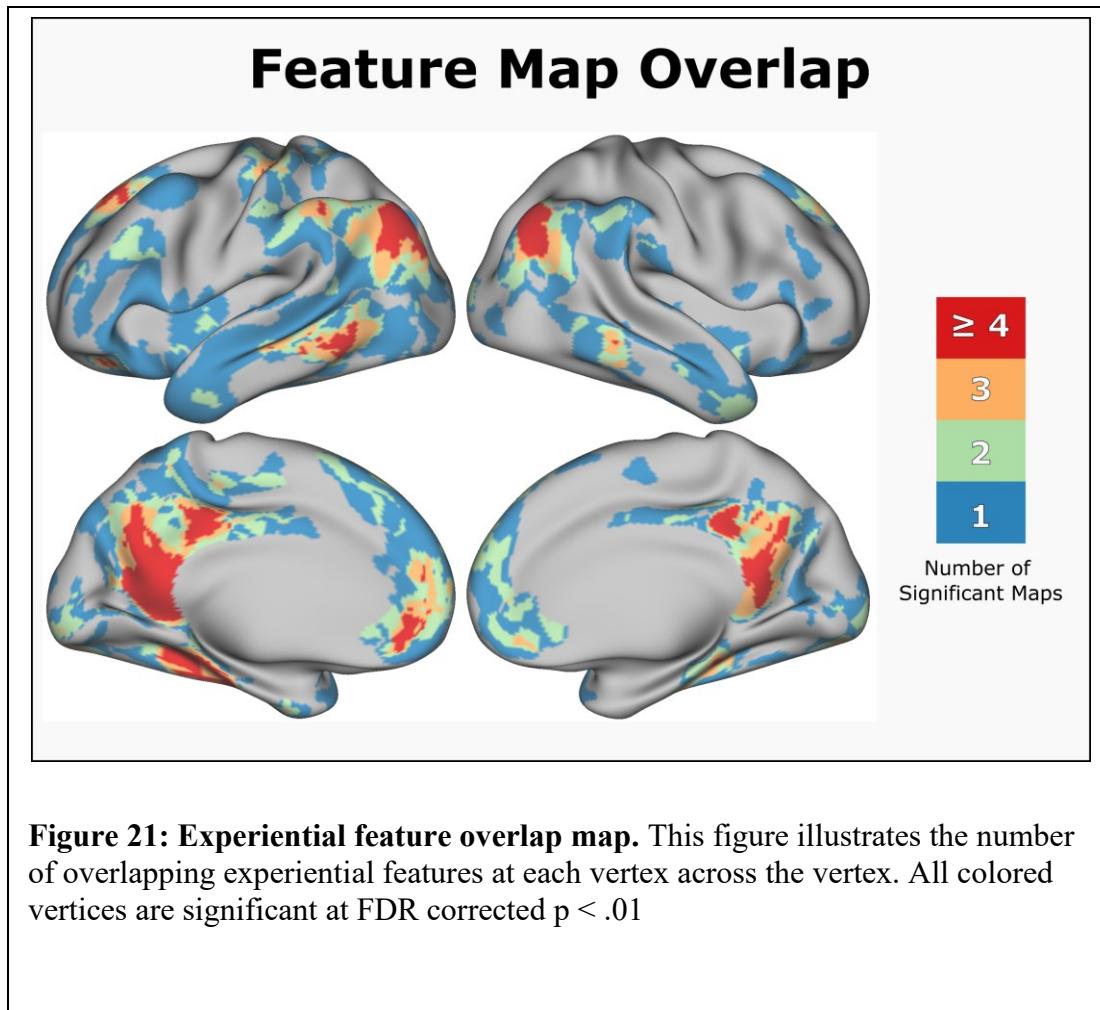
Shown in **Figure 21**, large regions of cortex had time-series that were significantly predicted by at least one experiential feature. The regions of cortex with the

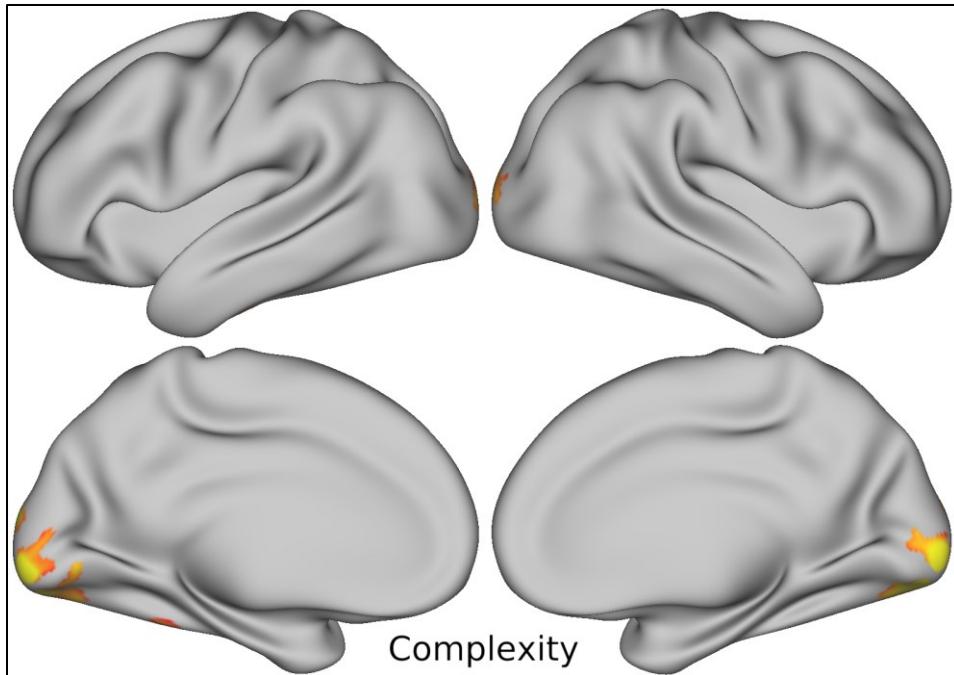
highest number of significant features were bilateral pMTG, AG, PCC, and PH, with larger regions in the left hemisphere compared to right in all cases. In addition, the left dorsomedial prefrontal cortex and dorsal cortex also had regions with 4 or more significant features.

As stated in the introduction to this chapter, due to the large number of intercorrelations between features, written description of the results focuses on the multivariate analysis. Eleven of the fourteen features had some significant vertices in the multivariate analysis. However, in follow-up univariate analysis the three initially non-significant features (Motion, Face, and Lower Limb) all had some significant vertices. All figures showing results from the multivariate analysis have vertices colored based on standardized beta-coefficients with the same color scale across all figures. Colored vertices are all significant at FDR corrected  $p < .01$ .

| Feature    | VIF Score |
|------------|-----------|
| Motion     | 3.69      |
| Color      | 4.33      |
| Complexity | 2.64      |
| Shape      | 5.48      |
| Face       | 1.34      |
| Sound      | 1.92      |
| Taste      | 3.13      |
| Head       | 2.23      |
| Upper Limb | 1.45      |
| Lower Limb | 1.48      |
| Number     | 1.26      |
| Scene      | 1.61      |
| Social     | 3.27      |
| Valence    | 1.24      |

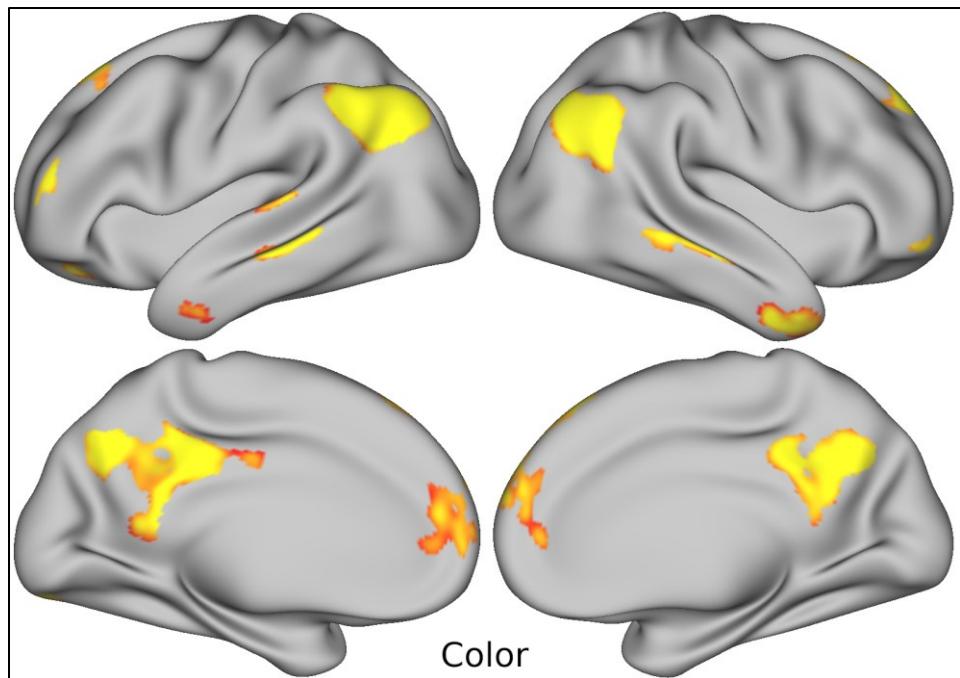
**Table 10: Variance inflation factors for selected experiential features.**





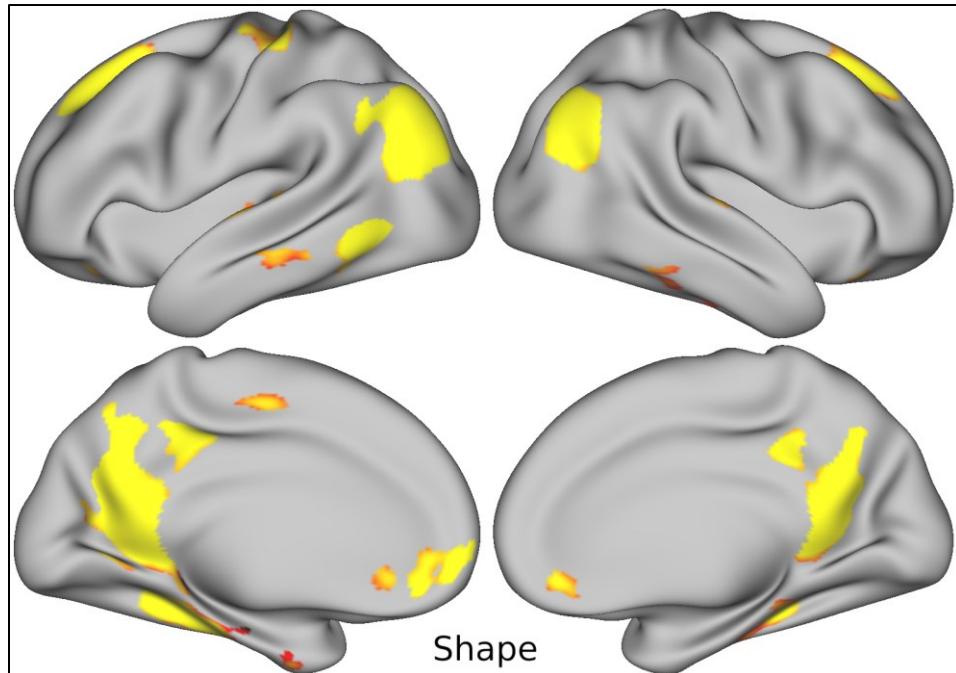
**Figure 22: Complexity feature map**

Visual Complexity was associated with increased activation in the bilateral occipital poles, and ventral occipital cortex. Complexity was also associated with smaller activation foci in the left middle FG and anterior FG.



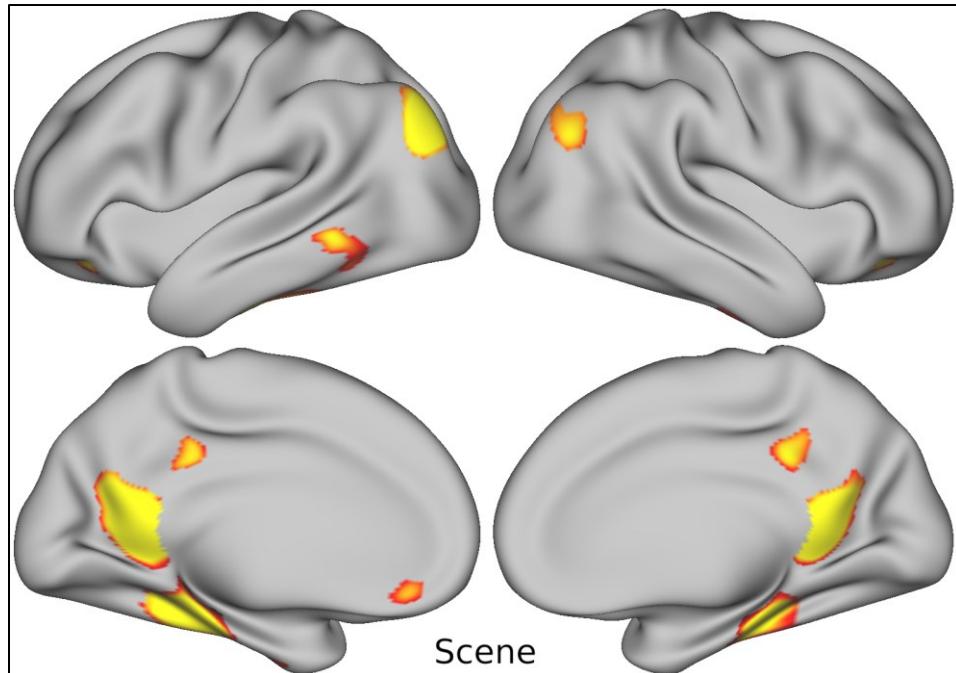
**Figure 23: Color feature map**

A network of regions was found to have significant sensitivity to the Color feature. Most of these are components of the canonical default mode network. The specific regions include the bilateral: lateral AT poles, PCC, dmPFC, AG, MTG, superior frontal gyrus, and middle frontal gyrus, as well as the left STG, left ventral occipital, and left OFC.



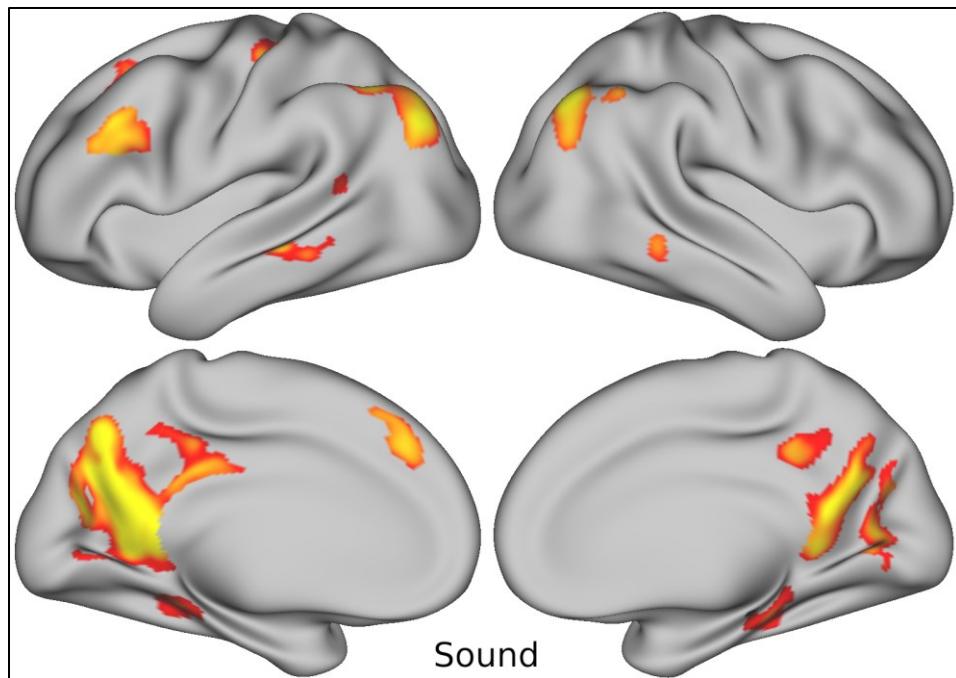
**Figure 24: Shape feature map**

The Shape feature was associated activation in the bilateral: AG, PCC, collateral sulcus, vmPFA, OFC, STG, and superior frontal sulcus, as well as activation the left pMTG, left sensory-motor hand area of the central sulcus, left middle temporal gyrus, right inferior temporal sulcus, left Heschle's gyrus, and left inferior paracentral lobule.



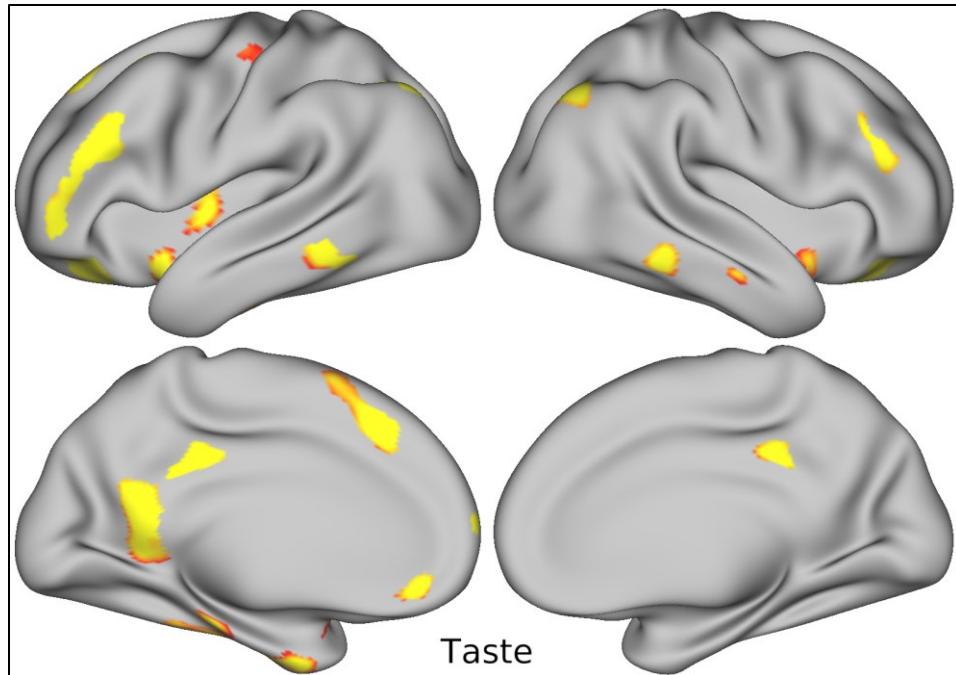
**Figure 25: Scene feature map**

The Scene feature was associated with activation in the bilateral: posterior parahippocampal gyrus, middle fusiform gyrus (FG), isthmus of the PCC, and OFC, as well as left > right AG, left pMTG, and left ventromedial prefrontal cortex.



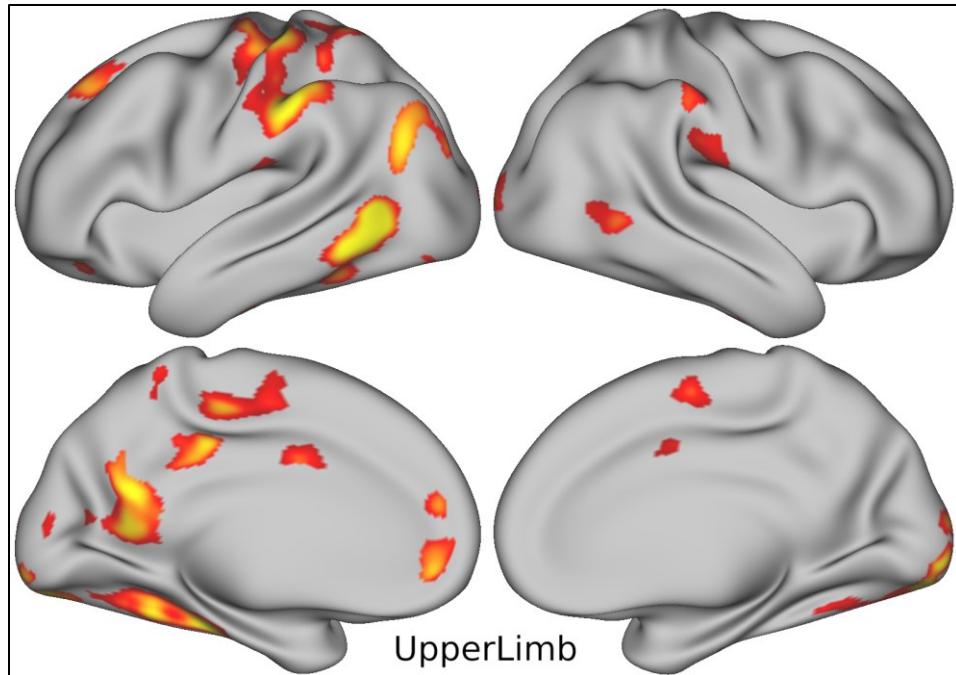
**Figure 26: Sound feature map**

The Sound feature was associated with activation in the bilateral: PCC, RSC, mid-MTG, parieto-occipital sulcus into the precuneus, posterior parahippocampal gyrus, and AG, as well as left superior and inferior frontal sulci, left sensory-motor hand area of the central sulcus, left STS, and left middle frontal gyri.



**Figure 27: Taste feature map**

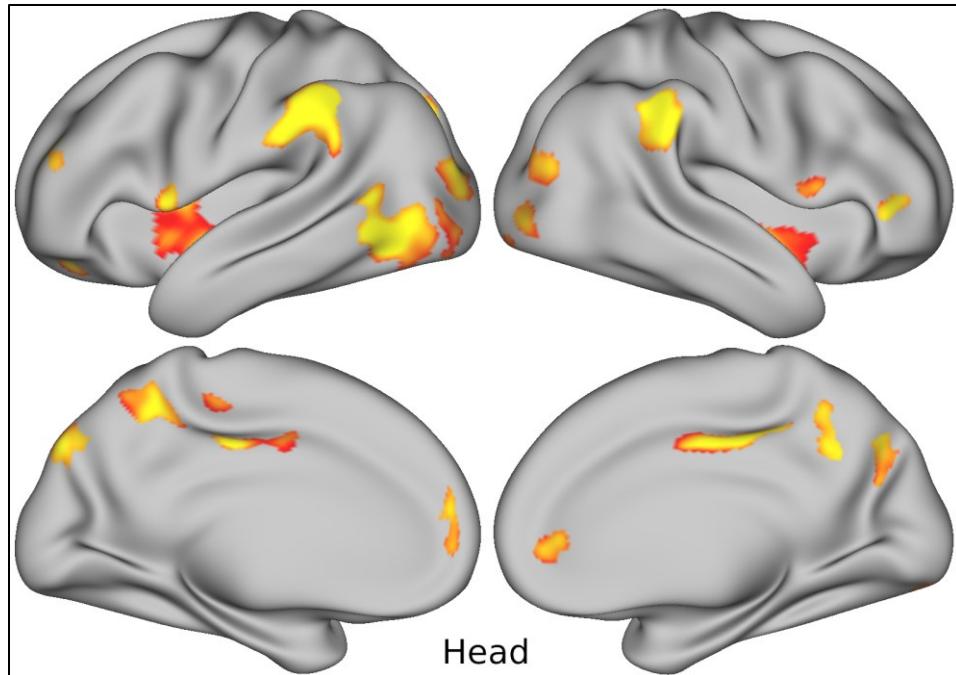
The Taste feature was associated with activation in the bilateral: OFC, pMTG, bilateral ventral insula, PCC, dorsal AG, and inferior frontal sulci, as well as left dorsal insula, left isthmus of the cingulate gyrus, left posterior middle frontal gyrus, and left superior frontal gyrus. Small foci were also observed in the left vmPFC, left frontal pole, and left collateral sulcus into the left uncus.



**Figure 28: Upper Limb feature map**

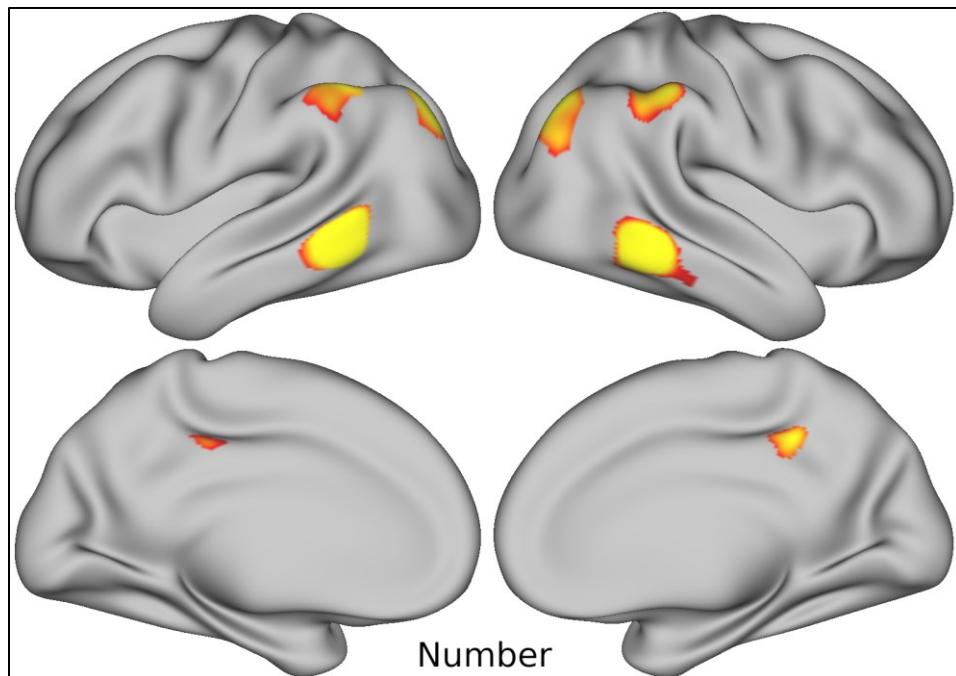
The Upper Limb action feature was associated with widespread areas of activation.

There was overall leftward lateralization in all bilateral regions. The areas of associated activation include the bilateral: pMTG, dorsal SMG, mid cingulate gyrus, and occipital pole, as well as left AG, left paracentral lobule, left collateral sulcus, left vmPFC, left PCC, left isthmus of the cingulate gyrus, left OFC, and left superior sensorimotor cortex (including hand representation area in the central sulcus).



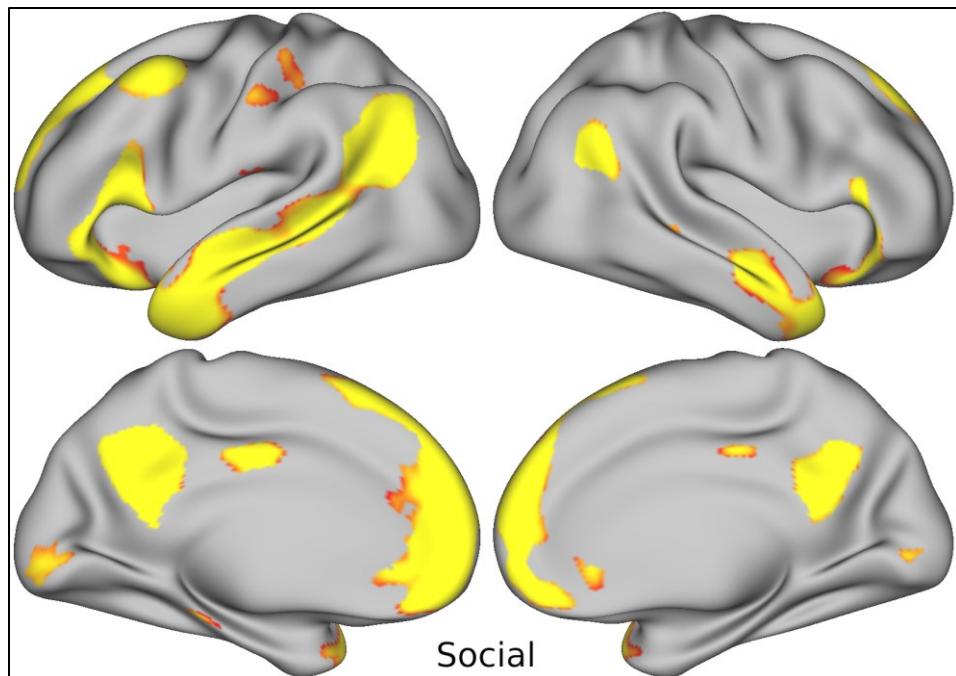
**Figure 29: Head feature map**

The Head action feature was associated with activation in the bilateral: middle and inferior occipital gyri, SMG, insula, ventral premotor cortex, posterior cingulate cortex, precuneus, and dorsal parietal occipital sulcus, as well as left LOTC, right rostral cingulate gyrus, left rostral medial superior frontal gyrus, left caudal IPS, and right anterior IFG. A small activation foci was also observed in the left middle frontal gyrus.



**Figure 30: Number feature map**

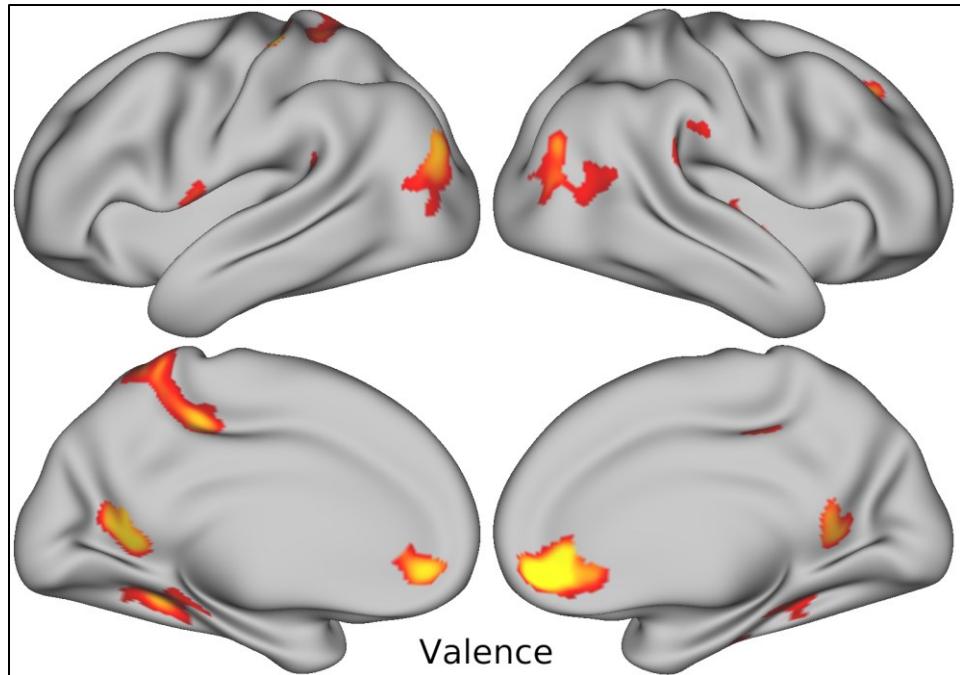
The Number feature was associated with activation in the bilateral: pMTG, dorsal AG into the IPS, dorsal SMG into the IPS and bilateral PCC.



**Figure 31: Social feature map**

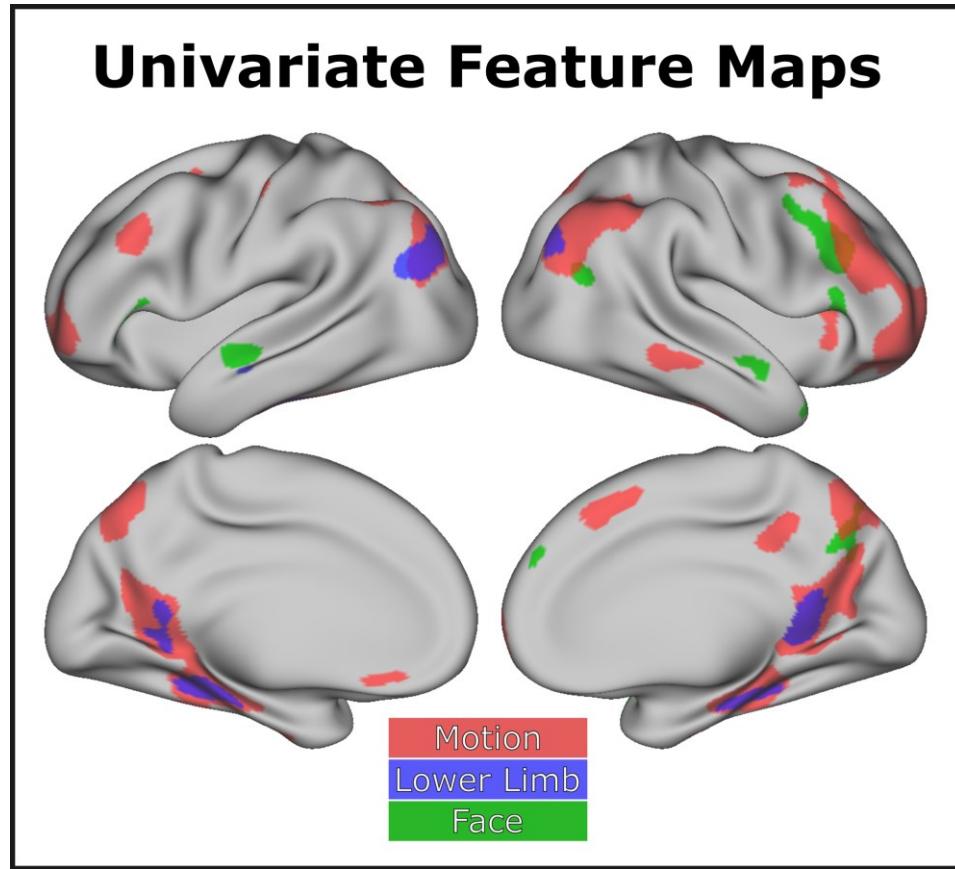
The Social feature was associated with activation in the bilateral: temporal poles, and anterior STS, as well as left posterior STS, left > right AG, left > right IFG, left > right dorsomedial and ventromedial prefrontal cortex, and left > right PCC and precuneus.

Smaller foci were observed in the left post-central gyrus, bilateral V1, left parahippocampal gyrus, left parietal operculum, and left posterior middle frontal gyrus.



**Figure 32: Valence feature map**

The Valence feature was associated with activation in the bilateral: vmPFC, isthmus of the cingulate gyrus, left > right marginal sulcus, posterior AG, posterior collateral sulcus, and posterior planum temporale, as well as left sensory-motor hand area, right superior frontal sulcus, and left ventral premotor area.

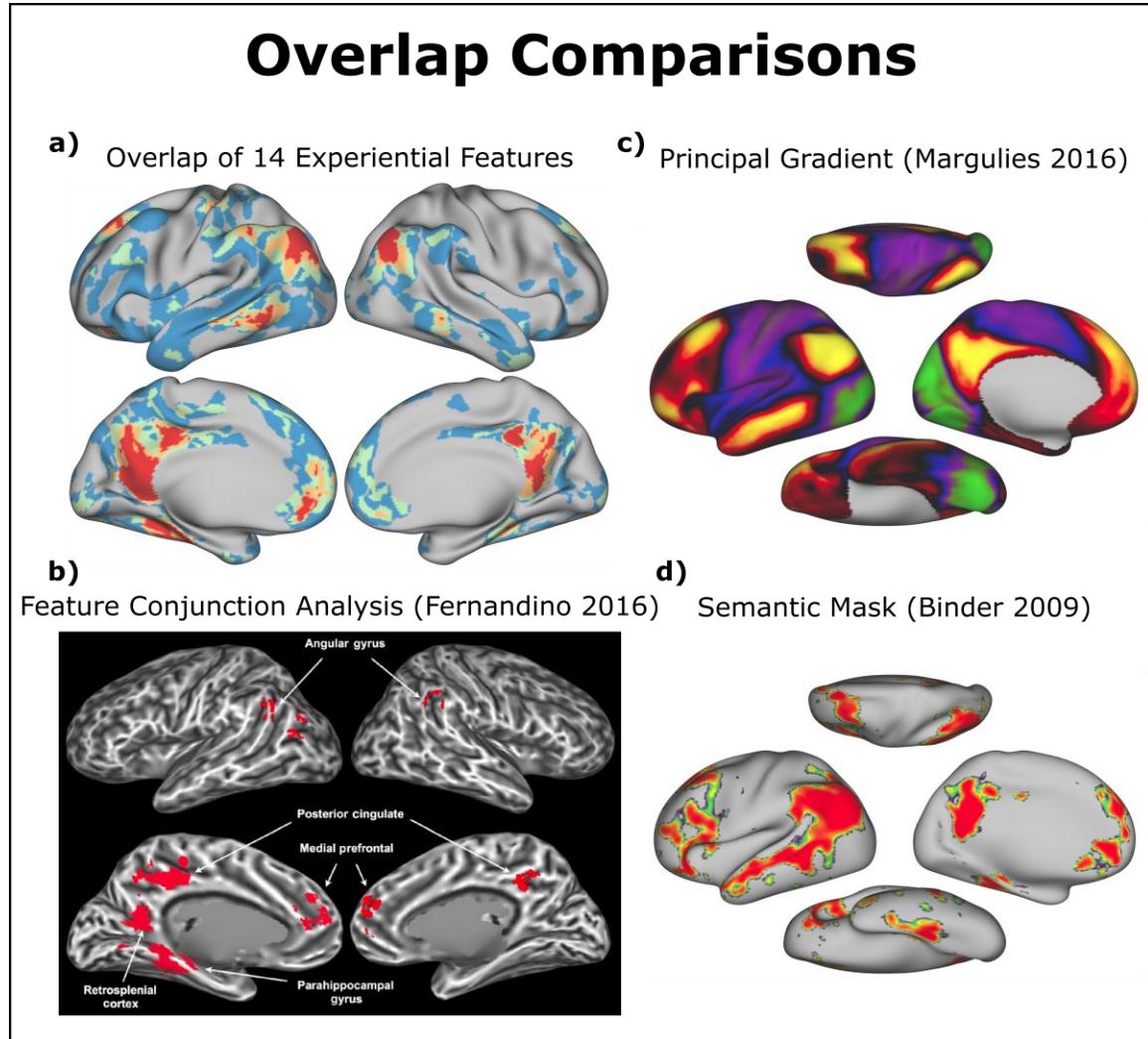


**Figure 33: Follow-up univariate feature maps**

The three features not significant in multivariate analysis all had areas of significant activation when they were the sole regressors in the regression model. The Motion feature was associated with significant activation in the bilateral: AG into the IPS, FG into parahippocampal gyri, preCun, RSC, middle frontal gyrus, superior frontal sulcus, as well as left vmPFC, right vmPFC, right PCC, right MTG, right anterior insula, and right anterior FG. A small focus was also observed in the left post central gyrus. The Lower Limb feature was associated with activation in the bilateral: left > right AG, RSC, and collateral sulcus into the parahippocampal gyri. Face was associated with activation in the bilateral: anterior STS as well as right middle frontal sulcus, and right preCun.

Smaller foci of activation were also observed in the right dmPFC, right AT pole, right AC and two small foci bilaterally in the IFG.

## Discussion



**Figure 34: Comparison of feature overlaps to prior work.** **a)** Shows the number of features that were significant at each vertex (reproduction of Figure 21). **b)** A reproduction of Figure 5 from Fernandino et. al. (2016). The colored areas correspond to “areas associated with all 5 semantic attributes in the conjunction analysis” (Fernandino 2016). The figure is licensed under Oxford University Press and permission to reproduce the figure here was provided by Copyright Clearance Center. **c)** A visualization of the first principal gradient from Margulies et. al. (2016). **d)** A visualization of the general semantic mask reported in Binder et. al. (2009). All colored vertices are significant at  $p < .01$ .

A set of fourteen experiential features were simultaneously fit to surface time series data and analyzed at the group level. A hierarchical convergence of the features that closely matched the principal gradient was observed (**Figure 34**). The areas of highest overlap occurred at the pMTG, AG, and PCC, all bilaterally, and left > right FG. For the feature maps that had significant vertices, the results were mixed with respect to prior expectations. Some individual feature maps, such as Complexity, Scene, and Social closely matched expectations. Some other feature maps, such as Head, Number, Shape, Sound, Taste, Upper Limb, and Valence had overlap with the expected areas, but also included some other unanticipated areas of activation. Lastly, the Color feature map included broad areas of activation that were not anticipated.

The prediction of hierarchical feature overlap as distance from primary sensory-motor cortices increased was well supported. The regions of cortex with a high degree of feature overlap are also strikingly similar to those observed in a similar prior study by Fernandino et al. (2016). Given the large amount of literature on the DMN and convergent evidence for areas of feature overlap, we find compelling support for a view of semantic representation that includes a hierarchy of modal processing.

We note, however, that while there was close similarity between the principal gradient and the observed regions of experiential feature, there were some differences. The MTG appears as a large high-level area in the principal gradient, yet the feature overlap map shows only the pMTG as having a high degree of feature overlap. One likely reason for this difference is that the lateral temporal lobes are known to have low signal-to-noise ratio as a result of ‘signal dropout’. That is, due to bone/air/tissue interfaces in this area, inhomogeneities in the magnetic field lead to decreased MR signal. We think

that it is likely that the MTG would show highly semantic and multimodal representations if there was improved signal quality in this region.

Another area of difference between the principal gradient and semantic feature overlap map was in the fusiform gyrus and parahippocampal gyrus. Although the FG does show increased high-level characteristics in the principal gradient, the feature overlap map emphasizes the hubness of this region more. One likely reason for this difference is related to our choice of experiential features. Given the strong representation of concrete features in our analysis (e.g., Shape, Color, Upper Limb, Taste, Sound), it is likely that there is a ‘visual bias’<sup>23</sup> that emphasizes the ‘hubness’ of visual areas.

A difference in magnitude was also observed in the left IFG region. One likely reason for this is the converse of the effect in the FG/parahippocampal gyrus, and the same reason the Social feature was remarkably widespread. The lack of ‘abstract’ experiential features in our analysis de-emphasized the hubness of the IFG. Evidence in support of this comes from the RSA performed by Tong et al. (2022) in a subset of the current participants. In that study, the authors performed searchlight RSA with the full 65 dimensional CREA model. The resulting map shows large RSA correlations in the IFG and a lesser degree of correlation in the parahippocampal gyrus, closer to the pattern observed in the principal gradient.

Transitioning to the interpretation of the individual feature maps, eleven of the fourteen analyzed features had significant vertices. Each of the features are discussed in turn, with non-significant features discussed first.

---

<sup>23</sup> Although Upper Limb, Taste, and Sound might not appear to be ‘Visual’ features, they likely are highly correlated with concreteness along with the visual feature.

### Non-significant Feature Maps:

*Motion*: Although previous studies have found activation associated with Motion features, our analysis did not. However, there were large regions associated with Motion in the follow-up univariate analysis. Although the ratings for the Motion feature are not particularly skewed to the low end, the lack of significant Motion vertices in the multivariate analysis could be because of limitations in the underlying stimulus set. For example, the Motion feature could be difficult to detect because no verbs are included in the study. As such, it could be that within the current stimulus set Motion is ‘masked’ by features such as Upper Limb, and a broader stimulus set could detect some unique variance associated with Motion. Another consideration is that it could be that short phrases evoke more activation relevant to this feature (e.g., “a boy hit a ball”). Support for this hypothesis comes from analyzing the studies included in the Kuhnke et al metaanalysis (2023). Of the studies included in their analysis of conceptual processing related to motion, only 5 of the 24 studies had written nouns as stimuli, with most using either verbs (11 of 24), written sentences, or pictures.

*Face*: No region was found to be significantly modulated by the Face attribute. Notably, no activation was observed in the fusiform face area. This result is similar to a prior study that used an fMRI adaptation paradigm “to explore the integration of information about familiar faces and names” (Harris et al. 2015). Their results suggested that, within the fusiform gyrus, “information about familiar faces and names is not integrated at this stage of processing” (Harris et al. 2015). One possible reason no activation appeared modulated by the Face attribute is the same as Motion. Similar to how the feature map for motion was likely limited due to choices made in the creation of

the stimulus set, the lack of proper nouns likely limited the amount of activation that the ‘Face’ feature would be able to isolate. Indeed, one study that used famous names found “increased MR signal activity was observed on a bilateral basis for both the hippocampus and parahippocampal gyrus (PHG) during recognition of familiar names [...] compared to non-famous names” (Douville et al. 2005). Further, pointing to a sensitivity problem, there was significant univariate activation for the face feature in the middle temporal lobe, but this disappeared when controlling for all other features.

*Lower Limb*: No region was significant for the Lower Limb feature in the full model. One likely reason for this is that few words in the study were rated highly on the lower limb feature. While this could be the result of variance being masked by other features, the univariate result also did not show much activation correlated with lower limb ratings. Given the non-significant results for both Motion and Lower Limb, one important consideration for future studies could be the inclusion of words with higher ratings on the Lower Limb feature (e.g., verbs along with nouns).

## **Significant Feature Maps**

*Complexity*: While this feature is not typically investigated, the activation in the primary visual cortex matches the expectation. This provides strong support for the power of this analysis approach and sensitivity to modality specific features.

*Color*: There was a large portion of the cortex that was associated with color. This finding was somewhat unexpected. Bilateral AG and medial PCC were previously observed to encode color information in Fernandino et al. (2016). The left IPS was also found in a meta-analysis (Kuhnke et al. 2023). Interestingly, an RSA study with both

blind and sighted individuals, when controlling for several visual features, found the left STG to have a unique object color effect slightly anterior to our observed STG peak (Figure 4C; Wang et al. 2020). One potential cause for the large and distributed network is that not many comparable visual or somato-motor features were included such as “Bright”, “Dark”, “Pattern”, “Manipulation”, “Touch”, or “Weight”.

Interpreting the results of a regression analysis with co-linear features is difficult, but some degree of collinearity is likely to exist across any interpretable semantic feature set. For example, a major division in the stimuli of Study 1 is between event nouns and object nouns. Event nouns are rated as more abstract than concrete nouns (a la Brysbaert et al. 2014) and many features in the experiential model capture this distinction. All ‘concrete’ features (Shape, Color, ...) and ‘abstract’ features (Social, Valence) are correlated with each other. This problem is similar to the problem of cross-cut categories discussed in chapter 3. Given the many possible features likely to co-occur with the Color feature (especially those not already included in the model), we think that the large activation for color likely represents non-specific variance. This reasoning is supported by the observation that Color has the second highest VIF score (behind Shape) in our analysis.

Shape: The shape feature was significant across a large region of cortex, this was similar to the finding in Fernandino et al. (2016) and several of the involved regions were expected. As described in the introduction, the large area of involvement is likely reflecting a large amount of non-specific variance that the feature accounts for, as indicated by Shape having the highest VIF score. It is likely that the specific contributions of Shape would be reduced if other features were included in the model.

**Sound:** A notable observation was the finding of bilateral MTG activation associated with the sound feature. This observation fits well with prior literature (e.g., Lewis et al. 2004) and appears similar to the activation reported in Kuhnke et al. (2023). The finding of strong RSC activation was unexpected. Activation in this area for the Sound attribute has been previously reported as Fernandino and colleagues (2016) write: “. The analysis in which sound was the only predictor in the model revealed additional activations in the retrosplenial cortex, parahippocampal gyrus, and postcentral gyrus”. Also, a post-hoc search of the literature reveals a study that used fear-conditioned rats and found that the RSC “is required for the retrieval of remote memory for auditory cues”<sup>24</sup>(Todd et al. 2016). These authors performed a retrograde labeling study to determine projections to the RSC and found “consistent labeling in the primary (Au1) and secondary (AuD and AuV) auditory cortices of all animals, indicating direct projections from these cortices to the RSC”. This connection is interesting, as Fernandino et al. (2016) also find medial activation for the Sound attribute in the PCC.

**Taste:** The regions associated with the Taste feature include areas in the primary and secondary gustatory cortices. This result is impressive as it persists even with features such as color or upper limb being controlled for, which are likely to correlate with whether objects have a “characteristic or defining taste”. The other peaks on the medial surface, while not directly expected, could in the ventral case, reflect the concreteness of the food items, and in the dorsal case, reflect emotional attachment to some items with taste.

---

<sup>24</sup> Remote condition is 28 days following initial fear conditioning task.

Head: Head demonstrated widespread activation in an expected, but interpretable pattern. Bilateral LOTC activation was close to where the extrastriate body area typically localizes, which is a known processor for body parts. Middle insula activation was surprising, but likely could result from shared variance in items from the “Food” category having both taste and typical head movements (e.g., chewing) associated with it. The bilateral SMA is typically associated with the action perception network, and this could be the effect that is noticed. The medial peaks are of unknown importance, but there is a small activation in the ventral portion of the precentral sulcus.

Upper Limb: An extensive network was observed for the Upper Limb action feature that overlapped with many of the anticipated areas. The network of activated regions contains the large pMTG and IPS, known to be involved in action perception. Further, the large activation across the left central sulcus near the hand area bodes well with the possibility of semantic involvement of this primary sensory-motor cortex. While the large medial regions of activation were not entirely expected, they are partly reconciled in considering concrete object preferring regions over abstract regions. Ventrolateral and posterior cingulate cortex have demonstrated this preference. One possible mechanistic understanding is that the RSC is thought to play a role in spatial memory.

Number: The finding of bilateral IPS activation associated with the Number feature matches the expectation from prior literature. This is a remarkable finding, and also appears to be partially observable in the PrAGMATIC atlas generated by Huth et al. (2016). The bilateral pMTG activation was not anticipated, but also appears to be

noticeable in the PrAGMATIC<sup>25</sup> atlas (especially in the right hemisphere). This activation is interesting as it appears well circumscribed and warrants further investigation. The Number feature was also associated with bilateral activation in the PCC, and this focus is of unknown significance.

*Scene*: The Scene feature showed remarkable similarity to the regions implicated in the perceptual processing of scenes, especially with regard to the activations on the medial surface and posterior AG. The activations in the left pMTG were not anticipated and are of unknown importance. As with all features, it is possible that the Scene feature ‘cuts across’ categories and the lateral activations are the result of non-specific Scene activation that would not remain if other features were included in the analysis.

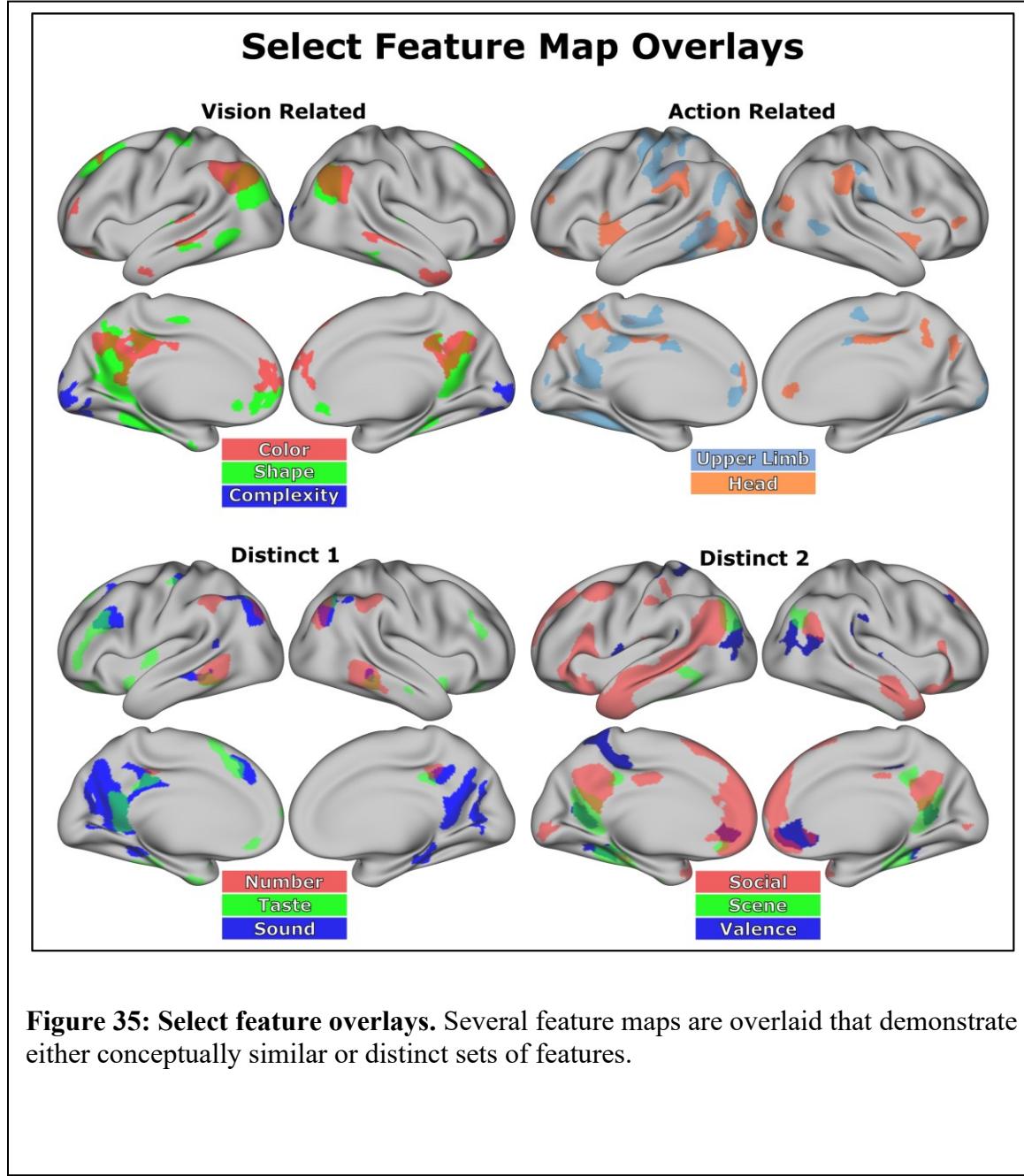
*Social*: A large network of regions was associated with the Social attribute including bilateral AT pole and left STS. This large network of regions aligns well with previous results of Lin et al. (2024) as well as the typical regions associated with the processing of Social concepts. The observed network of regions is extensive and the lateral regions are similar to the ‘language network’ as defined by some other researchers (e.g., Fedorenko et al. 2024). As expected, the Social attribute also closely matches the areas implicated in ToM literature. Given that the VIF score for the social feature is not much higher than for comparable features, Social is likely absorbing non-specific variance for many other abstract features.

*Valence*: Several regions were significant for valence. Activations in the vmPFC and PCC match expectations. Activation in the bilateral AG and left preCun are more

---

<sup>25</sup> Link to interactive viewer: <https://gallantlab.org/viewer-huth-2016/>

difficult to interpret. The involvement of more subcortical regions, such as the amygdala



**Figure 35: Select feature overlays.** Several feature maps are overlaid that demonstrate either conceptually similar or distinct sets of features.

can be difficult to assess in surface-based analyses so volumetric based analysis is required to better understand potential involvement of any subcortical structures.

Shown in **Figure 35**, an interesting observation of the individual feature maps is that some feature maps from the same domain (e.g., Vision related or Action related)

have broad areas of non-overlapping activation. Conversely, some features that might appear conceptually distinct, such as Number, Taste and Sound, have several large areas of overlap. Lastly, some conceptually distinct features showed remarkable non-overlap in some hub areas. For example, although Social, Scene, and Valence, all have associated activations in the bilateral AG, there is minimal overlap between these feature maps in this region. Although the number of possible feature overlay maps is large, one general finding appears to be that there are complex intercorrelations between features that lead to highly distributed representations that span the cortex.

### ***Methodological Considerations***

While several reasons were provided for why some feature maps might not be significant or match prior expectations, one reason has yet to be discussed: the specification of the feature queries. Collecting and curating tens of thousands of human ratings is a major endeavor. The collection of any large set of normative data opens up an almost endless array of data analysis strategies. For example, analysis could be stratified based on inter-rater consistency as poorly specified feature queries would be expected to have low inter-rater consistency. Although inter-rater agreement was high across the 535 words initially collected with the publication of the CREA model<sup>26</sup>, subsets of the data are likely to have varying degrees of inter-rater agreement. For example, it is likely that not only will individual features meaningfully differ with respect to how consistent participants rate words on the feature, but for a given feature, different categories of words are likely to have differing degrees of inter-subject agreement. For example, such an effect is seen in

---

<sup>26</sup> Quoting: “the final data set consisted of 15,268 rating vectors with a mean intra-word individual-to-group correlation of .785 (median .803) [...]”

the normative ratings collected in a study by Grand et al. (2022). In that study, while Size is generally a meaningful semantic feature, human ratings of clothing items by size were not included in the analysis because of low inter-rater reliability.

The reasons some features on some word-sets might have low inter-rater reliability are varied. A likely candidate that was already mentioned is poor specification of the feature. That is, participants simply did not understand what exactly it was they were being asked to rate. However, it is important to note that it is also possible that a feature is difficult to measure (i.e., has low inter-rater reliability), but is neurobiologically relevant with meaningful neural correlates. Another reason for low inter-rater reliability could be a feature not being applicable to a set of words. Given that the CREA features have the general prompt of “to what degree ...”, feature inapplicability is likely to manifest as low scores. As a result, features with low average values or low variance in ratings are also less likely to have detectable neural correlates.

Another reason feature maps could not show any significant vertices at the group level is variability across participants in the locations where a feature is represented. For example, it could be that some features have a meaningful representation at the neural level, but if the activation is circumscribed and the precise anatomical location varies among individuals, no group level effect may be observed.

### ***Future Directions***

As mentioned in the discussion of the Motion feature, adding more diverse stimulus set could help to unveil the involvement of more cortical regions in semantic representation. For example, as already mentioned, the addition of verbs into the stimulus

set could reveal an even more distributed set of regions, especially for the non-significant features of Motion and Lower Limb. Adding additional stimuli can help to reveal whether, for example, whether the Motion feature as currently is not a detectable experiential feature in the current paradigm, or just a limitation of the stimulus set.

Another future direction is the use of different feature mapping strategies. Although a portion of the introduction to this chapter was spent motivating the use of vertex-wise encoding models, it remains to be tested whether more complex non-linear encoding techniques offer significant improvements in explainability. For example, suppose the brain happens to use population coding to represent a feature. In that case, the activation of multiple voxels might need to be simultaneously considered to decode the presence or absence of an experiential feature.

One limitation of more complex methods is that there are many of them, and it can sometimes be difficult to determine the trade-off between more complex models and meaningful improvements in model fit. In this case, it is important to emphasize that there are multiple senses of the word ‘explainability’: there is both the formal sense with typical ‘percent variance explained’ metrics and their related metrics such as Akaike information criterion (AIC) or Bayesian information criterion (BIC), as well as the intuitive sense where the resulting maps can be understood in light of previous literature. This latter sense is necessarily difficult to formulate but is meant more as a heuristic as the existence of “low-level” sensory motor cortices is not strongly contested<sup>27</sup>. However,

---

<sup>27</sup> The exact nature and function of “low-level” cortex is of course not entirely known, and there are likely “top-down” influences muddying these distinctions. However, those objections notwithstanding, it is generally agreed that primary sensory cortex is a reasonable construct.

further investigation is needed to determine whether more complex models in fact do add explainability over and above vertex-wise encoding models.

This chapter of the dissertation focused on analyzing the CREA features that, perhaps subjectively, have the strongest prior spatial expectations. Given that all 65 experiential features in the CREA model have neurobiological motivation, next steps are to analyze more features of the model. However, this task is not without methodological challenges. There is, perhaps, *the* limiting problem with the inclusion of additional features: collinearity. Discussion of collinearity can and does fill textbooks, and a myriad of methods have been put forward to mitigate the problem of having similar, but distinct predictors. Discussion of these methods, however, is outside the scope of this chapter. It remains critical future work to determine what methods can faithfully relate a large number of experiential features to their neural correlates.

### Conclusions:

This chapter provides resounding support for the hypothesis that semantic representation involves a large region of cortex and that several semantic hub regions exist at the convergence of lower-level representations. As a result, a parsimonious explanation of several lines of work come together in the experiential account of semantic representation. Regarding the interpretability of the generated experiential feature maps, results were mixed with how well the maps matched prior expectations. Several of the maps, such as Scene, Complexity, and Taste showed remarkable overlap with known perceptual processors. However, some of the maps were either more expansive than expected (e.g., Color and Sound), or did not show any activation (Motion, Lower Limb,

Face) motivating further investigation. As a result, the findings of this chapter warrant further investigation into the neural correlates of experiential information.

## General Discussion

The chapters of this dissertation have outlined an experiential account of concept representation with a specific focus on the emergence of category selective regions of the cortex.

One point that has not been made explicit so far is that although the phenomena of CRSDs implies spatial segregation in the semantic representation of items from different categories, it does not imply that the pattern of spatial segregation is similar across individuals. As such, while the presence of significant category contrasts at the group level supports spatial separation in the representation of relevant conceptual knowledge, the absence of a group effect would not have implied overlapping category representations. Given that the semantic system is likely to have individual variation, it means that the group level univariate effects are likely to *underestimate* the degree of a category effect. Although this dissertation has some analyses that depend on consistent spatial organization across individuals (e.g., pairwise univariate category contrasts and generation of category maps), other analyses in this dissertation are less dependent on a shared organizational layout (e.g., multivariate clustering within a large mask, vertex-wise encoding for individual words in individual stimuli, and searchlight RSA encoding). Given the consistent results from all analyses throughout this dissertation, there is support not only for the claim that the organization of the semantic system has category preferring regions, but also, that the general organizing principles are consistent across individuals.

Another aspect of this dissertation that has not been discussed is the choice of the experiential model. Although CREA arguably represents the most comprehensive

experiential model, it is likely that some features could be better specified, some important features are likely absent, and some features might sufficiently overlap as to have indistinguishable substrate at the resolution of fMRI. Elaborating on features that might be missing from the CREA model of Binder and colleagues (2016), one class of words not mentioned outside of the discussion of the discussion of chapter 5 is verbs. It is likely that some neurobiologically relevant features will apply to verbs (e.g., being unergative) but not have relevance for nouns<sup>28</sup>.

Reviewing the previous chapters, below are some chapter specific future directions followed by general discussion of broader future aims.

### Chapter Specific Future Directions

Chapter 2 examined the reliability of neural RDMs as a function of a number of participants and stimulus repetitions. While the qualitative pattern of results is important and the primary emphasis of the chapter, there are two future directions that were not mentioned that could be valuable to future researchers. The first is that no explicit statistical framework was discussed to relate ICCs to statistical power. The inclusion of such a formula could be helpful to future researchers collecting pilot data. The second derives from a consideration that is not typically given much attention. Due to the entries of neural RDMs not being independent (that is, there is an implicit structure in the off-diagonal entries of RDMs), standard statistical formulas are not necessarily valid. It is for this reason that most researchers perform significance testing using permutation tests. In practice, this is often not much of an issue, as it is empirically observed that when

---

<sup>28</sup> Kim, S. et al. Neural and Experiential Semantic Correlates of the Unergative-Unaccusative Verb Distinction. in Society for the Neurobiology of Language (2023).

Pearson correlation distances are used to generate neural RDMs, standard statistical formulas are approximately true (e.g., the null distribution for the relevant sum of squares ratio to calculate the ICC follows the F-distribution with the expected degrees of freedom in the numerator and denominator). However, when Euclidean distances are used, the loss of degrees of freedom is readily observed in the null distribution. While this might appear to be an intractable problem, the loss of the degrees of freedom in pairwise distance matrices can be analytically accounted for as is done in multivariate distance matrix regression (MDMR; Anderson 2001; Zapala and Schork 2012). Researchers might benefit from the derivation of a formula for calculating intraclass correlation coefficients of RDMs that is analogous to the classical framework but accounts for the implicit structure in (Euclidean derived) RDMs.

Chapter 3 examined the shared substrate of different categories of concepts. As stated throughout the introduction to that chapter, the specific hypothesis that was tested related to whether experiential feature maps *linearly* combined to predict regions of category selectivity. As implied in the introduction to that chapter, it could be the case that a non-linear combination of features better predicts selectivity across categories. The possibility of such was not pursued, in part because if it were to be the case that feature maps combined non-linearly, it would not directly bear on the central topic of this dissertation (i.e., that category selectivity can be accounted for by the spatial layout of experiential features). The primary benefit of performing a more complex analysis is that it might provide insight into the computational role of different regions of the cortex.

Chapter 4 examined the representation of body parts and showed that the CREA model can account for much of the observed specificity. One future analysis is an

assessment of *what experiential* information is important for the representation of this category. This precise question was initially investigated in the initial versions of chapter 3. However, untangling the contribution of different experiential features led to an array of difficult methodological considerations. For example, it was observed that the key (experiential) features that seemed to drive the response varied based on whether univariate analysis was used, ridge-regression, random-forest feature selection, etc. Also, independent of the neural data, it was also observed that reliable features that discriminated amongst body parts (e.g., upper-limb association, lower-limb association) did not have significant beta coefficients in feature re-weighted RSA, bringing up concerns of collinearity among the predictors. For these reasons, a comprehensive examination of what drives the response was not undertaken here.

Although chapter 4 focused on the body part category, subsequent analyses for each the 3 other categories of concrete nouns is also warranted.

## General Future Directions

Ultimately, while fMRI studies can provide a large and rich amount of data, fMRI represents only one avenue of data generation for model testing. Model verification will likely require convergent results from different measurement modalities. One modality that can also help to elucidate the functional role of regions of the cortex is magnetoencephalography (MEG). While fMRI ostensibly provides good spatial resolution, it is limited in its temporal resolution. Determining the order in which information is represented in different brain regions while comprehending a sentence or phrase has the potential to provide insight into the mechanisms of concept representation (e.g., Vignali et al. 2023).

Another modality of information that has high spatial and temporal resolution, but low coverage, is the use of electrocorticography (ECoG). Similar to MEG data, ECoG data has very high temporal resolution, as well as the ability to measure single neuron responses (Alahi et al. 2021). The use of ECoG has the potential to provide a mechanistic understanding of conceptual access and deepen our understanding of the neural correlates of the CREA model<sup>29</sup>.

Lastly, another critical avenue of support comes from the relation between lesions in the brain and changes in different behavioral measures. Since the advent of high-resolution brain imaging techniques, this is largely accomplished using a method called voxel-based lesion-symptom mapping (VLSM; Bates et al. 2003). The ability to predict a patient's deficits based on a structural MRI from when they enter a clinic is considered by some, to use a *façon de parler*, where the rubber hits the road. That is, some might say an ultimate end that theory building should serve is the generation of clinically actionable information that stands to benefit patients. To this end, it remains to be tested whether the CREA model can improve prognostication or treatment for patients. While VLSM studies have been informative for corroborating and better understanding the functional roles played by different regions of the cortex e.g., (e.g., Binder et al. 2020; Pillay et al. 2017; Pillay et al. 2014; Schwartz et al. 2011; Schwartz et al. 2009), a critical next step to support the CREA model is to relate the predictions implied by the fMRI analysis to VLSM findings.

---

<sup>29</sup> For an example of such work, see the poster by Gross WL, Fernandino L, Conant LL, Krucoff MO, Mueller WM, Raghavan M, Binder JR. Electrocortical activity during semantic retrieval is modulated by feature modality. (<https://wiredbrains.org/publications/>)

## General Conclusion

The analysis presented throughout this dissertation supports several claims related to experiential models of cognition. Namely:

1. There is general category selectivity throughout the cortex,
2. Experiential models can partially account for this category selectivity by weighted combinations of experiential feature maps,
3. Body parts represent a semantically distinct category of concepts with robust representation in the pMTG, SMG, and preCS,
4. Experiential feature maps hierarchically converge in a pattern that closely matches the default mode network and
5. Experiential feature maps are widely distributed across the cortex, and some experiential features have high overlap with known perceptual processors.

It is hoped that the efforts taken throughout the generation of this dissertation both advance the discussion on how sensory inputs become imbued with semantic meaning and generate insights that lead to improved care for those in need.

## References

- Alahi MEE, Liu Y, Xu Z, Wang H, Wu T, Mukhopadhyay SC. 2021. Recent advancement of electrocorticography (ecog) electrodes for chronic neural recording/stimulation. *Mater Today Commun.* 29:102853.
- Allefeld C, Haynes J-D. 2014. Searchlight-based multi-voxel pattern analysis of fmri by cross-validated manova. *NeuroImage.* 89:345-357.
- Allport DA. 1985. Distributed memory, modular subsystems and dysphasia. *Current perspectives in dysphasia.*
- Anders S, Lotze M, Erb M, Grodd W, Birbaumer N. 2004. Brain activity underlying emotional valence and arousal: A response-related fmri study. *Hum Brain Mapp.* 23(4):200-209.
- Anderson A, Binder JR, Fernandino L, Humphries CJ, Conant LL, Aguilar M, Wang X, Doko D, Raizada RDS. 2016a. Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cereb Cortex.*
- Anderson AJ, Binder JR, Fernandino L, Humphries CJ, Conant LL, Raizada RDS, Lin F, Lalor EC. 2019. An integrated neural decoder of linguistic and experiential meaning. *J Neurosci.* 39(45):8969-8987.
- Anderson AJ, Zinszer BD, Raizada RDS. 2016b. Representational similarity encoding for fmri: Pattern-based synthesis to predict brain activity using stimulus-model-similarities. *NeuroImage.* 128:44-53.
- Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26(1):32-46.

- Arbuckle SA, Yokoi A, Pruszynski JA, Diedrichsen J. 2019. Stability of representational geometry across a wide range of fmri activity levels. *NeuroImage*. 186:155-163.
- Arioli M, Gianelli C, Canessa N. 2021. Neural representation of social concepts: A coordinate-based meta-analysis of fmri studies. *Brain Imaging Behav*. 15(4):1912-1921.
- Avery JA, Liu AG, Ingeholm JE, Riddell CD, Gotts SJ, Martin A. 2019. Taste quality representation in the human brain. *The Journal of Neuroscience*. 40(5):1042-1052.
- Bandettini PA. 2012. Sewer pipe, wire, epoxy, and finger tapping: The start of fmri at the medical college of wisconsin. *NeuroImage*. 62(2):620-631.
- Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C et al. 2013. Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*. 80:169-189.
- Barrós-Loscertales A, González J, Pulvermüller F, Ventura-Campos N, Bustamante JC, Costumero V, Parcet MA, Ávila C. 2012. Reading salt activates gustatory brain regions: Fmri evidence for semantic grounding in a novel sensory modality. *Cereb Cortex*. 22(11):2554-2563.
- Barsalou LW. 1999. Perceptual symbol systems. *Behav Brain Sci*. 22(4):577-660.
- Bartels A, Zeki S. 2000. The architecture of the colour centre in the human visual brain: New results and a review \*. *Eur J Neurosci*. 12(1):172-193.
- Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. 2003. Voxel-based lesion-symptom mapping. *Nat Neurosci*. 6(5):448-450.

- Beauchamp MS, Haxby JV, Jennings JE, DeYoe EA. 1999. An fmri version of the farnsworth-munsell 100-hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cereb Cortex*. 9(3):257-263.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Royal Statistical Soc Ser B Methodol*. 57(1):289-300.
- Berker EA, Berker AH, Smith A. 1986. Translation of broca's 1865 report: Localization of speech in the third left frontal convolution. *Arch Neurol-chicago*. 43(10):1065-1072.
- Berthier ML. 2005. Poststroke aphasia. *Drug Aging*. 22(2):163-182.
- Binder JR, Conant LL, Humphries CJ, Fernandino L, Simons SB, Aguilar M, Desai RH. 2016. Toward a brain-based componential semantic representation. *Cogn Neuropsychol*. 33(3-4):1-45.
- Binder JR, Desai RH. 2011. The neurobiology of semantic memory. *Trends Cogn Sci*. 15(11):527-536.
- Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex*. 19(12):2767-2796.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW. 1999. Conceptual processing during the conscious resting state: A functional mri study. *J Cognitive Neurosci*. 11(1):80-93.

- Binder JR, Tong JQ, Pillay SB, Conant LL, Humphries CJ, Raghavan M, Mueller WM, Busch RM, Allen L, Gross WL et al. 2020. Temporal lobe regions essential for preserved picture naming after left temporal epilepsy surgery. *Epilepsia*.
- Biran I, Coslett HB. 2003. Visual agnosia. *Curr Neurol Neurosci Rep.* 3(6):508-512.
- Bobadilla-Suarez S, Ahlheim C, Mehrotra A, Panos A, Love BC. 2020. Measures of neural similarity. *Comput Brain Behav.* 3(4):369-383.
- Boccia M, Raimo S, Vita AD, Battisti A, Matano A, Guariglia C, Grossi D, Palermo L. 2020. Topological and hodological aspects of body representation in right brain damaged patients. *Neuropsychologia.* 148:107637.
- Borghesani V, Pedregosa F, Buiatti M, Amadon A, Eger E, Piazza M. 2016. Word meaning in the ventral visual path: A perceptual to conceptual gradient of semantic coding. *NeuroImage.* 143:128-140.
- Bracci S, Caramazza A, Peelen MV. 2015. Representational similarity of body parts in human occipitotemporal cortex. *The Journal of Neuroscience.* 35(38):12977-12985.
- Brambati SM, Myers D, Wilson A, Rankin KP, Allison SC, Rosen HJ, Miller BL, Gorno-Tempini ML. 2006. The anatomy of category-specific object naming in neurodegenerative diseases. *J Cognitive Neurosci.* 18(10):1644-1653.
- Bravo G, Potvin L. 1991. Estimating the reliability of continuous measures with cronbach's alpha or the intraclass correlation coefficient: Toward the integration of two traditions. *J Clin Epidemiol.* 44(4-5):381-390.

- Bruffaerts R, Dupont P, Peeters R, Deyne SD, Storms G, Vandenberghe R. 2013. Similarity of fmri activity patterns in left perirhinal cortex reflects semantic similarity between words. *The Journal of Neuroscience*. 33(47):18597-18607.
- Brysbaert M, Warriner AB, Kuperman V. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behav Res Methods*. 46(3):904-911.
- Buxbaum LJ, Coslett HB. 2001. Specialised structural descriptions for human body parts: Evidence from autotopagnosia. *Cogn Neuropsychol*. 18(4):289-306.
- Cai MB, Schuck NW, Pillow JW, Niv Y. 2019. Representational structure or task structure? Bias in neural representational similarity analysis and a bayesian method for reducing bias. *Plos Comput Biol*. 15(5):e1006299.
- Capitani E, Laiacoma M, Mahon B, Caramazza A. 2003. What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn Neuropsychol*. 20(3-6):213-261.
- Caramazza A, Hillis AE, Rapp BC, Romani C. 1990. The multiple semantics hypothesis: Multiple confusions? *Cogn Neuropsychol*. 7(3):161-189.
- Caramazza A, Mahon BZ. 2006. The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cogn Neuropsychol*. 23(1):13-38.
- Carota F, Nili H, Pulvermüller F, Kriegeskorte N. 2020. Distinct fronto-temporal substrates of distributional and taxonomic similarity among words: Evidence from rsfMRI of bold signals. *NeuroImage*. 117408.
- Caspers S, Eickhoff SB, Geyer S, Schepers F, Mohlberg H, Zilles K, Amunts K. 2008. The human inferior parietal lobule in stereotaxic space. *Brain Struct Funct*. 212(6):481-495.

- Caspers S, Geyer S, Schleicher A, Mohlberg H, Amunts K, Zilles K. 2006. The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. *NeuroImage*. 33(2):430-448.
- Caspers S, Zilles K, Laird AR, Eickhoff SB. 2010. A meta-analysis of action observation and imitation in the human brain. *NeuroImage*. 50(3):1148-1167.
- Chao LL, Martin A. 1999. Cortical regions associated with perceiving, naming, and knowing about colors. *J Cognitive Neurosci*. 11(1):25-35.
- Chen L, Ralph MA, Rogers TT. 2017. A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*. 1(3):0039.
- Chouinard PA, Goodale MA. 2010. Category-specific neural processing for naming pictures of animals and naming pictures of tools: An ale meta-analysis. *Neuropsychologia*. 48(2):409-418.
- Cicchetti DV. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assessment*. 6(4):284-290.
- Coltheart M, Inglls L, Cupples L, Michie P, Bates A, Budd B. 1998. A semantic subsystem of visual attributes. *Neurocase*. 4(4-5):353-370.
- Corradi-Dell'Acqua C, Hesse MD, Rumiati RI, Fink GR. 2008. Where is a nose with respect to a foot? The left posterior parietal cortex processes spatial relationships among body parts. *Cereb Cortex*. 18(12):2879-2890.
- Coslett HB. 2018. Apraxia, neglect, and agnosia. *Contin: Lifelong Learn Neurol*. 24(3, Behavioral Neurology and Psychiatry):768-782.

- Cox RW. 1996. Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res.* 29(3):162-173.
- Critchley M. 1953. The parietal lobes. Oxford, England: Williams and Wilkins.
- Damasio AR. 1989. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition.* 33(1-2):25-62.
- de Vet HCW, Mokkink LB, Mosmuller DG, Terwee CB. 2017. Spearman–brown prophecy formula and cronbach's alpha: Different faces of reliability and opportunities for new applications. *J Clin Epidemiol.* 85:45-49.
- Denes G, Cappelletti JY, Zilli T, Porta FD, Gallana A. 2000. A category-specific deficit of spatial representation: The case of autotopagnosia. *Neuropsychologia.* 38(4):345-350.
- Derderian KD, Zhou X, Chen L. 2021. Category-specific activations depend on imaging mode, task demand, and stimuli modality: An ale meta-analysis. *Neuropsychologia.* 161:108002.
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT et al. 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage.* 31(3):968-980.
- Dobbins IG, Schnyer DM, Verfaellie M, Schacter DL. 2004. Cortical activity reductions during repetition priming can result from rapid response learning. *Nature.* 428(6980):316-319.

- Dong J, Li A, Chen C, Qu J, Jiang N, Sun Y, Hu L, Mei L. 2021. Language distance in orthographic transparency affects cross-language pattern similarity between native and non-native languages. *Hum Brain Mapp.* 42(4):893-907.
- Douville K, Woodard JL, Seidenberg M, Miller SK, Leveroni CL, Nielson KA, Franczak M, Antuono P, Rao SM. 2005. Medial temporal lobe activity for recognition of recent and remote famous names: An event-related fmri study. *Neuropsychologia.* 43(5):693-703.
- Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A cortical area selective for visual processing of the human body. *Science.* 293(5539):2470-2473.
- Downing PE, Peelen MV, Wiggett AJ, Tew BD. 2006. The role of the extrastriate body area in action perception. *Soc Neurosci.* 1(1):52-62.
- Eggert GH. 2019. Early sources in aphasia and related disorders. De Gruyter.
- Epstein R. 2005. The cortical basis of visual scene processing. *Vis Cogn.* 12(6):954-978.
- Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature.* 392(6676):598-601.
- Epstein RA, Baker CI. 2019. Scene perception in the human brain. *Annu Rev Vis Sc.* 5(1):1-25.
- Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik IA, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M et al. 2019. Fmriprep: A robust preprocessing pipeline for functional mri. *Nat Methods.* 16(1):111-116.
- Farah MJ. 1996. The living/nonliving dissociation is not an artifact: Giving an a priori implausible hypothesis a strong test. *Cogn Neuropsychol.* 13(1):137-154.

- Farah MJ, McClelland JL. 1991. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *J Exp Psychology Gen.* 120(4):339-357.
- Faroqi-Shah Y, Sebastian R, Woude AV. 2018. Neural representation of word categories is distinct in the temporal lobe: An activation likelihood analysis. *Hum Brain Mapp.* 39(12):4925-4938.
- Fedorenko E, Ivanova AA, Regev TI. 2024. The language network as a natural kind within the broader landscape of the human brain. *Nat Rev Neurosci.* 1-24.
- Fernandino L, Binder JR. 2024. How does the “default mode” network contribute to semantic cognition? *Brain Lang.* 252:105405.
- Fernandino L, Binder JR, Desai RH, Pendl SL, Humphries CJ, Gross WL, Conant LL, Seidenberg MS. 2016. Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cereb Cortex.* 26(5):2018-2034.
- Fernandino L, Tong J-Q, Conant LL, Humphries CJ, Binder JR. 2022. Decoding the information structure underlying the neural representation of concepts. *Proc National Acad Sci.* 119(6):e2108091119.
- Finger S. 2001. *Origins of neuroscience: A history of explorations into brain function.* Oxford University Press.
- Firth J. 1957. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis.* 10-32.
- Fischer-Baum S, Bruggemann D, Gallego IF, Li DSP, Tamez ER. 2017. Decoding levels of representation in reading: A representational similarity approach. *Cortex.* 90:88-102.

- Fodor JA. 1975. *The language of thought*. Harvard University Press.
- Freud S. 1891. On aphasia. Stengel E, translator. New York: International Universities Press, Inc.
- Frisby SL, Halai AD, Cox CR, Ralph MAL, Rogers TT. 2023. Decoding semantic representations in mind and brain. *Trends Cogn Sci*.
- Funnel E, Sheridan J. 1992. Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cogn Neuropsychol*. 9(2):135-153.
- Gainotti G. 1996. Cognitive and anatomical locus of lesion in a patient with a category-specific semantic impairment for living beings. *Cogn Neuropsychol*. 13(3):357-390.
- Gainotti G. 2004. A metanalysis of impaired and spared naming for different categories of knowledge in patients with a visuo-verbal disconnection. *Neuropsychologia*. 42(3):299-319.
- Gallagher HL, Frith CD. 2003. Functional imaging of ‘theory of mind’. *Trends Cogn Sci*. 7(2):77-83.
- Gao Z, Zheng L, Gouws A, Krieger-Redwood K, Wang X, Varga D, Smallwood J, Jefferies E. 2022. Context free and context-dependent conceptual representation in the brain. *Cereb Cortex*.
- Gerstmann J. 1940. Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: Local diagnostic value. *Arch Neurol Psychiatry*. 44(2):398-408.
- Gessell B, Geib B, Brigand FD. 2021. Multivariate pattern analysis and the search for neural representations. *Synthese*. 1-21.

- Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M et al. 2016. A multi-modal parcellation of human cerebral cortex. *Nature*. 536(7615):171-178.
- Glenberg AM. 1997. What memory is for. *Behav Brain Sci*. 20(1):1-19.
- Goldberg RF, Perfetti CA, Schneider W. 2006a. Distinct and common cortical activations for multimodal semantic categories. *Cognitive Affect Behav Neurosci*. 6(3):214-222.
- Goldberg RF, Perfetti CA, Schneider W. 2006b. Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*. 26(18):4917-4921.
- González J, Barros-Loscertales A, Pulvermüller F, Meseguer V, Sanjuán A, Belloch V, Ávila C. 2006. Reading cinnamon activates olfactory brain regions. *NeuroImage*. 32(2):906-912.
- Goodglass H, Klein B, Carey P, Jones K. 1966. Specific semantic word categories in aphasia. *Cortex*. 2(1):74-89.
- Grafton ST, Fadiga L, Arbib MA, Rizzolatti G. 1997. Premotor cortex activation during observation and naming of familiar tools. *NeuroImage*. 6(4):231-236.
- Grand G, Blank IA, Pereira F, Fedorenko E. 2022. Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*. 6(7):975-987.
- Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn Sci*. 10(1):14-23.
- Grill-Spector K, Malach R. 2004. The human visual cortex. *Annu Rev Neurosci*. 27(1):649-677.

- Grosbras MH, Beaton S, Eickhoff SB. 2012. Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Hum Brain Mapp.* 33(2):431-454.
- Gross WL, Fernandino L, Conant LL, Krucoff MO, Mueller WM, Raghavan M, Binder JR. Electrocortical activity during semantic retrieval is modulated by feature modality.
- Guariglia C, Piccardi L, Allegra MCP, Traballesi M. 2002. Is autotopoagnosia real? Ec says yes. A case study. *Neuropsychologia.* 40(10):1744-1749.
- Guo W, Geng S, Cao M, Feng J. 2022. Functional gradient of the fusiform cortex for chinese character recognition. *Eneuro.* 9(3):ENEURO.0495-0421.2022.
- Harnad S. 1990. The symbol grounding problem. *Phys D Nonlinear Phenom.* 42(1-3):335-346.
- Harris RJ, Rice GE, Young AW, Andrews TJ. 2015. Distinct but overlapping patterns of response to words and faces in the fusiform gyrus. *Cereb Cortex.* 26(7):3161-3168.
- Harris ZS. 1954. Distributional structure. *WORD.* 10(2-3):146-162.
- Hart J, Gordon B. 1992. Neural subsystems for object knowledge. *Nature.* 359(6390):60-64.
- Hauk O, Johnsrude I, Pulvermüller F. 2004. Somatotopic representation of action words in human motor and premotor cortex. *Neuron.* 41(2):301-307.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science.* 293(5539):2425-2430.

- Head H, Holmes G. 1911. Sensory disturbances from cerebral lesions. *Brain*. 34(2-3):102-254.
- Heilman KM. 2006. Aphasia and the diagram makers revisited: An update of information processing models. *J Clin Neurol*. 2(3):149-162.
- Henderson SK, Dev SI, Ezzo R, Quimby M, Wong B, Brickhouse M, Hochberg D, Touroutoglou A, Dickerson BC, Cordella C et al. 2021. A category-selective semantic memory deficit for animate objects in semantic variant primary progressive aphasia. *Brain Commun*.
- Hillis AE, Caramazza A. 1991. Category-specific naming and comprehension impairment: A double dissociation. *Brain*. 114(5):2081-2094.
- Hodgson VJ, Ralph MAL, Jackson RL. 2022. The cross-domain functional organization of posterior lateral temporal cortex: Insights from ale meta-analyses of 7 cognitive domains spanning 12,000 participants. *Cereb Cortex*.
- Hsu NS, Frankland SM, Thompson-Schill SL. 2012. Chromaticity of color perception and object color knowledge. *Neuropsychologia*. 50(2):327-333.
- Huth AG, Heer WAd, Griffiths TL, Theunissen FE, Gallant JL. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*. 532(7600):453.
- An introduction to the life and work of john hughlings jackson: Introduction. 2007. *Med Hist Suppl*. (26):3-34.
- Iwamura Y. 1998. Hierarchical somatosensory processing. *Curr Opin Neurobiol*. 8(4):522-528.

- Jastorff J, Begliomini C, Fabbri-Destro M, Rizzolatti G, Orban GA. 2010. Coding observed motor acts: Different organizational principles in the parietal and premotor cortex of humans. *J Neurophysiol.* 104(1):128-140.
- Johnston M, Leslie S-J. 2019. Metaphysics and cognitive science. 183-205.
- Kable JW, Kan IP, Wilson A, Thompson-Schill SL, Chatterjee A. 2005. Conceptual representations of action in the lateral temporal cortex. *J Cognitive Neurosci.* 17(12):1855-1870.
- Kanwisher N. 2010. Functional specificity in the human brain: A window into the functional architecture of the mind. *Proc National Acad Sci.* 107(25):11163-11170.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci.* 17(11):4302-4311.
- Kemmerer D. 2015. Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychon B Rev.* 22(4):1068-1075.
- Kemmerer D, Castillo JG, Talavage T, Patterson S, Wiley C. 2008. Neuroanatomical distribution of five semantic components of verbs: Evidence from fmri. *Brain Lang.* 107(1):16-43.
- Kemmerer D, Tranel D. 2008. Searching for the elusive neural substrates of body part terms: A neuropsychological study. *Cogn Neuropsychol.* 25(4):601-629.
- Kraut R. 2022. Plato. In: Edward NZ, editor. *The Stanford Encyclopedia of Philosophy.* Spring 2022 ed.: Metaphysics Research Lab, Stanford University.

- Kriegeskorte N, Mur M, Bandettini PA. 2008. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers Syst Neurosci.* 2:4.
- Kuhnke P, Beaupain MC, Arola J, Kiefer M, Hartwigsen G. 2023. Meta-analytic evidence for a novel hierarchical model of conceptual processing. *Neurosci Biobehav Rev.* 144:104994.
- Laiacona M, Allamano N, Lorenzi L, Capitani E. 2006. A case of impaired naming and knowledge of body parts. Are limbs a separate sub-category? *Neurocase.* 12(5):307-316.
- Le Clec'H G, Dehaene S, Cohen L, Mehler J, Dupoux E, Poline JB, Lehéricy S, Moortele PFvd, Bihan DL. 2000. Distinct cortical areas for names of numbers and body parts independent of language and input modality. *NeuroImage.* 12(4):381-391.
- LeBel A, Jain S, Huth AG. 2021. Voxelwise encoding models show that cerebellar language representations are highly conceptual. *The Journal of Neuroscience.* 41(50):10341-10355.
- Lee S-M, Henson RN, Lin C-Y. 2020. Neural correlates of repetition priming: A coordinate-based meta-analysis of fmri studies. *Front Hum Neurosci.* 14:565114.
- Lescroart MD, Stansbury DE, Gallant JL. 2015. Fourier power, subjective distance, and object categories all provide plausible models of bold responses in scene-selective visual areas. *Front Comput Neurosc.* 9:135.
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA. 2004. Human brain regions involved in recognizing environmental sounds. *Cereb Cortex.* 14(9):1008-1021.

- Li H, Cao Y, Chen C, Liu X, Zhang S, Mei L. 2022. The depth of semantic processing modulates cross-language pattern similarity in chinese–english bilinguals. *Hum Brain Mapp.*
- Li H, Qu J, Chen C, Chen Y, Xue G, Zhang L, Lu C, Mei L. 2019. Lexical learning in a new language leads to neural pattern similarity with word reading in native language. *Hum Brain Mapp.* 40(1):98-109.
- Lin N, Wang X, Xu Y, Wang X, Hua H, Zhao Y, Li X. 2018. Fine subdivisions of the semantic network supporting social and sensory–motor semantic processing. *Cereb Cortex.* 28(8):2699-2710.
- Lin N, Zhang X, Wang X, Wang S. 2024. The organization of the semantic network as reflected by the neural correlates of six semantic dimensions. *Brain Lang.* 250:105388.
- Lingnau A, Downing PE. 2015. The lateral occipitotemporal cortex in action. *Trends Cogn Sci.* 19(5):268-277.
- Lissauer H, Jackson M. 1988. A case of visual agnosia with a contribution to theory. *Cogn Neuropsychol.* 5(2):157-192.
- Liu X, Hu L, Qu J, Zhang S, Su X, Li A, Mei L. 2023. Neural similarities and differences between native and second languages in the bilateral fusiform cortex in chinese–english bilinguals. *Neuropsychologia.* 179:108464.
- Liu Z, Lin Y, Sun M. 2020. Representation learning for natural language processing. 1-11.
- Liuzzi A, Aglinskas A, Fairhall S. 2020. General and feature-based semantic representations in the semantic network. *Sci Rep-uk.* 10(1):8931.

- Longo MR, Azañón E, Haggard P. 2010. More than skin deep: Body representation beyond primary somatosensory cortex. *Neuropsychologia*. 48(3):655-668.
- Lueck CJ, Zeki S, Friston KJ, Deiber MP, Cope P, Cunningham VJ, Lammertsma AA, Kennard C, Frackowiak RSJ. 1989. The colour centre in the cerebral cortex of man. *Nature*. 340(6232):386-389.
- Machery E. 2016. The amodal brain and the offloading hypothesis. *Psychon B Rev*. 23(4):1090-1095.
- Mahon BZ, Caramazza A. 2009. Concepts and categories: A cognitive neuropsychological perspective. *Annu Rev Psychol*. 60(1):27-51.
- Mahon BZ, Caramazza A. 2011. What drives the organization of object knowledge in the brain? *Trends Cogn Sci*. 15(3):97-103.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc National Acad Sci*. 92(18):8135-8139.
- Mar RA. 2011. The neural bases of social cognition and story comprehension. *Annu Rev Psychol*. 62(1):103-134.
- Marcus DS, Harwell J, Olsen T, Hodge M, Glasser MF, Prior F, Jenkinson M, Laumann T, Curtiss SW, Essen DCV. 2011. Informatics and data mining tools and strategies for the human connectome project. *Front Neuroinform*. 5:4.
- Margolis E, Laurence S. 2023. Concepts. In: Edward N. Zalta UN, editor. *The Stanford Encyclopedia of Philosophy*. Fall 2023 ed.: Metaphysics Research Lab, Stanford University.

- Margulies DS, Ghosh SS, Goulas A, Falkiewicz M, Huntenburg JM, Langs G, Bezgin G, Eickhoff SB, Castellanos FX, Petrides M et al. 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc National Acad Sci.* 113(44):12574-12579.
- Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG. 1995. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science.* 270(5233):102-105.
- Maunsell JH, Essen DCV. 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol.* 49(5):1127-1147.
- McCarthy RA, Warrington EK. 2015. Past, present, and prospects: Reflections 40 years on from the selective impairment of semantic memory (warrington, 1975). *Q J Exp Psychology.* 69(10):1-28.
- McGraw KO, Wong SP. 1996. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1(1):30-46.
- McInnes L, Healy J, Saul N, Großberger L. 2018. Umap: Uniform manifold approximation and projection. *J Open Source Softw.* 3(29):861.
- McKeefry DJ, Zeki S. 1997. The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain.* 120(12):2229-2242.
- Meersmans K, Storms G, Deyne SD, Bruffaerts R, Dupont P, Vandenberghe R. 2021. Orienting to different dimensions of word meaning alters the representation of word meaning in early processing regions. *Cereb Cortex.* bhab416-.
- Mesulam MM. 1998. From sensation to cognition. *Brain.* 121(6):1013-1052.

- Miceli G, Fouch E, Capasso R, Shelton JR, Tomaiuolo F, Caramazza A. 2001. The dissociation of color from form and function knowledge. *Nat Neurosci.* 4(6):662-667.
- Mikolov T, Chen K, Corrado G, Dean J. 2013. Efficient estimation of word representations in vector space.
- Moro V, Urgesi C, Pernigo S, Lanteri P, Pazzaglia M, Aglioti SM. 2008. The neural basis of body form and body action agnosia. *Neuron.* 60(2):235-246.
- Moseley R, Kiefer M, Pulvermüller F. 2016. Grounding and embodiment of concepts and meaning: A neurobiological perspective. *Foundations of embodied cognition: Perceptual and emotional embodiment.* New York, NY, US: Routledge/Taylor & Francis Group. p. 93-113.
- Nee DE. 2019. Fmri replicability depends upon sufficient individual-level data. *Commun Biology.* 2(1):130.
- Nieder A, Dehaene S. 2009. Representation of number in the brain. *Annu Rev Neurosci.* 32(1):185-208.
- Nielsen JM. 1962. Agnosia, apraxia, aphasia: Their value in cerebral localization. New York: Hafner Publishing Company.
- Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. 2014. A toolbox for representational similarity analysis. *Plos Comput Biol.* 10(4):e1003553.
- Noirhomme Q, Lesenfants D, Gomez F, Soddu A, Schrouff J, Garraux G, Luxen A, Phillips C, Laureys S. 2014. Biased binomial assessment of cross-validated

- estimation of classification accuracies illustrated in diagnosis predictions.  
NeuroImage Clin. 4:687-694.
- O'brien RM. 2007. A caution regarding rules of thumb for variance inflation factors.  
Qual Quant. 41(5):673-690.
- Ogden JA. 1985. Autotopagnosia: Occurrence in a patient without nominal aphasia and with an intact ability to point to parts of animals and objects. Brain. 108(4):1009-1022.
- Oldfield RC. 1971. The assessment and analysis of handedness: The edinburgh inventory.  
Neuropsychologia. 9(1):97-113.
- Olson IR, McCoy D, Klobusicky E, Ross LA. 2013. Social cognition and the anterior temporal lobes: A review and theoretical framework. Soc Cogn Affect Neur. 8(2):123-133.
- Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE. 2010. Surface-based information mapping reveals crossmodal vision–action representations in human parietal and occipitotemporal cortex. J Neurophysiol. 104(2):1077-1089.
- Orlov T, Makin TR, Zohary E. 2010. Topographic representation of the human body in the occipitotemporal cortex. Neuron. 68(3):586-600.
- Paillard J. 1999. Body schema and body image-a double dissociation. Motor control, today and tomorrow. 197-214.
- Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci. 8(12):976-987.

- Peelen MV, He C, Han Z, Caramazza A, Bi Y. 2014. Nonvisual and visual object shape representations in occipitotemporal cortex: Evidence from congenitally blind and sighted adults. *The Journal of Neuroscience*. 34(1):163-170.
- Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK. 2019. Psychopy2: Experiments in behavior made easy. *Behav Res Methods*. 51(1):195-203.
- Pick A. 1922. Störung der Orientierung am eigenen Körper. *Psychol Forsch*. 1(1):303-318.
- Pillay SB, Binder JR, Humphries C, Gross WL, Book DS. 2017. Lesion localization of speech comprehension deficits in chronic aphasia. *Neurology*. 88(10):970-975.
- Pillay SB, Stengel BC, Humphries C, Book DS, Binder JR. 2014. Cerebral localization of impaired phonological retrieval during rhyme judgment. *Ann Neurol*. 76(5):738-746.
- Pobric G, Ralph MAL, Zahn R. 2016. Hemispheric specialization within the superior anterior temporal cortex for social and nonsocial concepts. *J Cognitive Neurosci*. 28(3):351-360.
- Popham SF, Huth AG, Bilenko NY, Deniz F, Gao JS, Nunez-Elizalde AO, Gallant JL. 2021. Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat Neurosci*. 24(11):1628-1636.
- Reilly J, Shain C, Borghesani V, Kuhnke P, Vigliocco G, Peelle JE, Mahon B, Buxbaum L, Majid A, Brysbaert M et al. 2023. What we mean when we say semantic: A consensus statement on the nomenclature of semantic memory.
- Renzi ED, Lucchelli F. 1994. Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex*. 30(1):3-25.

- Rescorla M. 2023. The language of thought hypothesis. In: Edward N. Zalta UN, editor. Winter 2023 ed.: Metaphysics Research Lab, Stanford University.
- Ritchie JB, Masson HL, Bracci S, Beeck HPOd. 2021. The unreliable influence of multivariate noise normalization on the reliability of neural dissimilarity. *NeuroImage*. 245:118686.
- Rizzolatti G, Sinigaglia C. 2010. The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nat Rev Neurosci*. 11(4):264-274.
- Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P. 1976. Basic objects in natural categories. *Cognitive Psychol*. 8(3):382-439.
- Rugg MD, Thompson-Schill SL. 2013. Moving forward with fmri data. *Perspect Psychol Sci*. 8(1):84-87.
- Russell JA. 1980. A circumplex model of affect. *J Pers Soc Psychol*. 39(6):1161-1178.
- Sacchett C, Humphreys GW. 1992. Calling a squirrel a squirrel but a canoe a wigwam: A category-specific deficit for artefactual objects and body parts. *Cogn Neuropsychol*. 9(1):73-86.
- Samson D, Pillon A. 2010. A case of impaired knowledge for fruit and vegetables. *Cogn Neuropsychol*. 20(3-6):373-400.
- Schwartz MF, Kimberg DY, Walker GM, Brecher A, Faseyitan OK, Dell GS, Mirman D, Coslett BH. 2011. Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proc National Acad Sci*. 108(20):8520-8524.
- Schwartz MF, Kimberg DY, Walker GM, Faseyitan O, Brecher A, Dell GS, Coslett HB. 2009. Anterior temporal involvement in semantic word retrieval: Voxel-based lesion-symptom mapping evidence from aphasia. *Brain*. 132(12):3411-3427.

- Schwoebel J, Coslett HB. 2005. Evidence for multiple, distinct representations of the human body. *J Cognitive Neurosci.* 17(4):543-553.
- Schwoebel J, Coslett HB, Buxbaum LJ. 2001. Compensatory coding of body part location in autotopagnosia: Evidence for extrinsic egocentric coding. *Cogn Neuropsychol.* 18(4):363-381.
- Seabold S, Perktold J. 2010. Statsmodels: Econometric and statistical modeling with python. *Proc 9th Python Sci Conf.* 92-96.
- Semenza C. 1988. Impairment in localization of body parts following brain damage. *Cortex.* 24(3):443-449.
- Shelton JR, Fouch E, Caramazza A. 1998. The selective sparing of body part knowledge: A case study. *Neurocase.* 4(4-5):339-351.
- Shields C. 2023. Aristotle. In: Edward N. Zalta UN, editor. *The Stanford Encyclopedia of Philosophy.* Winter 2023 ed.: Metaphysics Research Lab, Stanford University.
- Simmons WK, Barsalou LW. 2003. The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cogn Neuropsychol.* 20(3-6):451-486.
- Simmons WK, Martin A, Barsalou LW. 2005. Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cereb Cortex.* 15(10):1602-1608.
- Simmons WK, Ramjee V, Beauchamp MS, McRae K, Martin A, Barsalou LW. 2007. A common neural substrate for perceiving and knowing about color. *Neuropsychologia.* 45(12):2802-2810.
- Sirigu A, Grafman J, Bressler K, Sunderland T. 1991. Multiple representations contribute to body knowledge: Evidence from a case of autotagnosia. *Brain.* 114(1):629-642.

- Small DM. 2010. Taste representation in the human insula. *Brain Struct Funct.* 214(5-6):551-561.
- Smallwood J, Bernhardt BC, Leech R, Bzdok D, Jefferies E, Margulies DS. 2021. The default mode network in cognition: A topographical perspective. *Nat Rev Neurosci.* 22(8):503-513.
- Spiridon M, Fischl B, Kanwisher N. 2006. Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum Brain Mapp.* 27(1):77-89.
- Staples R, Graves WW. 2020. Neural components of reading revealed by distributed and symbolic computational models. *Neurobiology Lang.* 1(4):381-401.
- Stewart F, Parkin AJ, Hunkin NM. 1992. Naming impairments following recovery from herpes simplex encephalitis: Category-specific? *Q J Exp Psychology Sect.* 44(2):261-284.
- Sugiura M, Sassa Y, Watanabe J, Akitsuki Y, Maeda Y, Matsue Y, Fukuda H, Kawashima R. 2006. Cortical mechanisms of person representation: Recognition of famous and personally familiar names. *NeuroImage.* 31(2):853-860.
- Suzuki K, Yamadori A, Fuji T. 1997. Category-specific comprehension deficit restricted to body parts. *Neurocase.* 3(3):193-200.
- Tanaka K. 1996. Inferotemporal cortex and object vision. *Annu Rev Neurosci.* 19(1):109-139.
- Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline J-B. 2007. Analysis of a large fmri cohort: Statistical and methodological issues for group analyses. *NeuroImage.* 35(1):105-120.

- Tizard B. 1959. Theories of brain localization from flourens to lashley. *Méd Hist.* 3(2):132-145.
- Todd TP, Mehlman ML, Keene CS, DeAngeli NE, Bucci DJ. 2016. Retrosplenial cortex is required for the retrieval of remote memory for auditory cues. *Learn Mem.* 23(6):278-288.
- Tong J, Binder JR, Humphries C, Mazurchuk S, Conant LL, Fernandino L. 2022. A distributed network for multimodal experiential representation of concepts. *The Journal of Neuroscience.* 42(37):7121-7130.
- Tour TDI, Eickenberg M, Nunez-Elizalde AO, Gallant JL. 2022. Feature-space selection with banded ridge regression. *NeuroImage.* 264:119728.
- Trumpp NM, Kliese D, Hoenig K, Haarmeier T, Kiefer M. 2013. Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex.* 49(2):474-486.
- Turner BO, Paul EJ, Miller MB, Barbey AK. 2018. Small sample sizes reduce the replicability of task-based fmri studies. *Commun Biology.* 1(1):62.
- Urgesi C, Berlucchi G, Aglioti SM. 2004. Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Curr Biol.* 14(23):2130-2134.
- Urgesi C, Candidi M, Ionta S, Aglioti SM. 2007. Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nat Neurosci.* 10(1):30-31.
- Vallat R. 2018. Pingouin: Statistics in python. *J Open Source Softw.* 3(31):1026.

- Vignali L, Xu Y, Turini J, Collignon O, Crepaldi D, Bottini R. 2023. Spatiotemporal dynamics of abstract and concrete semantic representations. *Brain Lang.* 243:105298.
- Vignemont Fd. 2007. How many representations of the body? *Behav Brain Sci.* 30(2):204-205.
- Vignemont Fd. 2010. Body schema and body image—pros and cons. *Neuropsychologia.* 48(3):669-680.
- Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J. 2016. Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage.* 137:188-200.
- Wang X, Men W, Gao J, Caramazza A, Bi Y. 2020. Two forms of knowledge representations in the human brain. *Neuron.* 107(2):383-393.e385.
- Wang X, Xu Y, Wang Y, Zeng Y, Zhang J, Ling Z, Bi Y. 2017. Representational similarity analysis reveals task-dependent semantic influence of the visual word form area. *Sci Rep-uk.* 8(1):3047.
- Warrens MJ. 2015. Quantitative psychology research, the 78th annual meeting of the psychometric society. Springer P Math Stat.293-300.
- Warrington EK. 1974. The selective impairment of semantic memory. *Q J Exp Psychol.* 27(4):635-657.
- Warrington EK, McCarthy RA. 1987. Categories of knowledge: Further fractionations and an attempted integration. *Brain.* 110(5):1273-1296.
- Warrington EK, Shallice T. 1984. Category specific semantic impairments. *Brain.* 107(3):829-853.

- Yildirim FB, Sarikcioglu L. 2007. Marie jean pierre flourens (1794–1867): An extraordinary scientist of his time. *J Neurology Neurosurg Psychiatry.* 78(8):852.
- Zahn R, Moll J, Krueger F, Huey ED, Garrido G, Grafman J. 2007. Social concepts are represented in the superior anterior temporal cortex. *Proc National Acad Sci.* 104(15):6430-6435.
- Zapala MA, Schork NJ. 2012. Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Frontiers Genetics.* 3:190.
- Zeki S. 1993. The mystery of louis verrey (1854-1916). *Gesnerus.* 50 ( Pt 1-2):96-112.
- Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS. 1991. A direct demonstration of functional specialization in human visual cortex. *The Journal of Neuroscience.* 11(3):641-649.
- Zhang Y, Lemarchand R, Asyraff A, Hoffman P. 2022. Representation of motion concepts in occipitotemporal cortex: Fmri activation, decoding and connectivity analyses. *NeuroImage.* 259:119450.
- Zhou F, Zhao W, Qi Z, Geng Y, Yao S, Kendrick KM, Wager TD, Becker B. 2021. A distributed fmri-based signature for the subjective experience of fear. *Nat Commun.* 12(1):6643.