UC Berkeley UC Berkeley Electronic Theses and Dissertations

Title

Investigating Relationships Between Semantic Representations in the Human Brain

Permalink

https://escholarship.org/uc/item/3f74x5cc

Author

Popham, Sara Frances

Publication Date

2021

Peer reviewed|Thesis/dissertation

Investigating Relationships Between Semantic Representations in the Human Brain

By

Sara Frances Popham

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jack Gallant, Chair Professor Michael Silver Professor Frederic Theunissen Professor Nina Dronkers

Summer 2021

Investigating Relationships Between Semantic Representations in the Human Brain

Copyright 2021 by Sara Frances Popham

Abstract

Investigating Relationships Between Semantic Representations in the Human Brain

by

Sara Popham

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Jack Gallant, Chair

Humans are constantly taking in information from a multitude of dissimilar sources to integrate and update their existing mental representations, but the series of computations that underlie this process are still largely unknown. We do know that these complex processes recruit multiple brain networks that interface with each other, but the precise roles that these networks play can change depending on task demands, and the full extent of this is only partially understood. This research presented in this dissertation focuses on furthering this knowledge specifically in the domain of semantic representation in humans.

In this dissertation, I first summarize the existing literature on semantic representations in the human brain. In addition to this, I describe how careful computational modeling methods alongside naturalistic experimental conditions allow us to answer very precise questions about representations in the human brain. I then present two functional magnetic resonance imaging (fMRI) experiments that explore new avenues of research in semantic representation and further show the strength of these methods. The first of these experiments illustrates that the visual and linguistic semantic networks of the brain are more precisely aligned than had previously been hypothesized. This pattern would not have been discovered using most typical fMRI analysis methods. In the second experiment, I analyze how linguistic semantic representations update after learning and memorization. Specifically, after participants have spent many hours memorizing songs, we can see that their semantic representations of the content of the song lyrics can shift dramatically.

Together, these two experiments show the importance of studying semantic representations not only in a single modality or setting. Without looking at the relationships between different types of representations or how representations are updated with learning and experience, we will never have a full picture of how the brain is functioning in complex, naturalistic environments.

LIST OF TABLES ACKNOWLEDGEMENTS CHAPTER 1 1.1 INTRODUCTION CHAPTER 2 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC COMPREHENSION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5	LIST	OF FIGURES	ii
ACKNOWLEDGEMENTS CHAPTER 1 1.1 INTRODUCTION CHAPTER 2 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC COMPREHENSION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5	LIST OF TABLES		
ACKNOWLEDGEMENTS CHAPTER 1 1.1 INTRODUCTION CHAPTER 2 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5			
CHAPTER 1 1.1 INTRODUCTION CHAPTER 2 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 1 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 5.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5	ACK	NOWLEDGEMENTS	iv
CHAUTERT 1.1 INTRODUCTION CHAPTER 2 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5			
 1.1 INTRODUCTION CHAPTER 2 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODAL ITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 			1
CHAPTER 2 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 1 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 CHAPTER 5	1.1	INTRODUCTION	1
 2.1 ABSTRACT 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	CHA	PTER 2	3
 2.2 INTRODUCTION 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	2.1	ABSTRACT	3
 2.3 MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES 	2.2	INTRODUCTION	3
 2.4 AMODAL SEMANTIC REPRESENTATIONS 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES 	2.3	MODALITY-SPECIFIC SEMANTIC REPRESENTATIONS	4
 2.5 CONTROL PROCESSES FOR SEMANTIC COMPREHENSION 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES 	2.4	AMODAL SEMANTIC REPRESENTATIONS	5
 2.6 RECENT STUDIES OF SEMANTIC REPRESENTATION 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	2.5	CONTROL PROCESSES FOR SEMANTIC COMPREHENSION	6
 2.7 IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES 2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	2.6	RECENT STUDIES OF SEMANTIC REPRESENTATION	7
2.8 SUMMARY AND CONCLUSION CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5	2.7	IMPLICATIONS OF RECENT STUDIES FOR CURRENT THEORIES	12
CHAPTER 3 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5	2.8	SUMMARY AND CONCLUSION	14
 3.1 ABSTRACT 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	CHA	PTER 3	16
 3.2 INTRODUCTION 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	3.1	ABSTRACT	16
 3.3 MATERIALS AND METHODS 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	3.2	INTRODUCTION	16
 3.4 RESULTS 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	3.3	MATERIALS AND METHODS	19
 3.5 DISCUSSION 3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES 	3.4	RESULTS	25
3.6 SUPPLEMENTAL FIGURES AND TABLE CHAPTER 4 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5	3.5	DISCUSSION	37
CHAPTER 44.1ABSTRACT4.2INTRODUCTION4.3MATERIALS AND METHODS4.4RESULTS4.5DISCUSSION4.6SUPPLEMENTAL FIGURESCHAPTER 5	3.6	SUPPLEMENTAL FIGURES AND TABLE	39
 4.1 ABSTRACT 4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES 	CHA	PTER 4	47
4.2 INTRODUCTION 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5	4.1	ABSTRACT	47
 4.3 MATERIALS AND METHODS 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	4.2	INTRODUCTION	47
 4.4 RESULTS 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	4.3	MATERIALS AND METHODS	48
 4.5 DISCUSSION 4.6 SUPPLEMENTAL FIGURES CHAPTER 5 	4.4	RESULTS	55
4.6 SUPPLEMENTAL FIGURES CHAPTER 5	4.5	DISCUSSION	74
CHAPTER 5	4.6	SUPPLEMENTAL FIGURES	76
	CHA	PTER 5	84
5.1 CONCLUSIONS	51	CONCLUSIONS	84
5.2 FUTURE DIRECTIONS	5.2	FUTURE DIRECTIONS	84
			01
REFERENCES	REFI	ERENCES	85

Table of Contents

i

List of Figures

2.1	Voxelwise modeling procedure.	9
2.2 2.3	Semantic maps obtained from participants who listened to narrative stories. Relationship between visual and linguistic semantic representations along	10
	the boundary of visual cortex.	13
3.1	Voxels with correlated visual and linguistic semantic representations.	18
3.2	Visual and linguistic representations of semantic concepts known to be well- represented in visual cortex	27
33	Mathad for datacting catagory-specific modality shifts	27
3.5	Locations of category-specific modality shifts across cortex	32
3.5	Quantitative summary of semantic correspondence across the boundary.	34
3.6	Alignment of semantic selectivity along the boundary between vision and	01
2.0	language.	36
3. S1	Evaluation of the visual and linguistic semantic models.	39
3.S2	Visual and linguistic representations of place concepts.	40
3.83	Visual and linguistic representations of body part concepts.	41
3.S4	Visual and linguistic representations of face concepts.	42
3.S5	Analysis region around the boundary of the occipital lobe.	43
3. S6	Locations of category-specific modality shifts across cortex for alternate	
	parameter set 1.	44
3.87	Locations of category-specific modality shifts across cortex for alternate parameter set 2.	45
4.1	Learning task structure and voxelwise modeling procedure.	56
4.2	Differences in temporal response profiles of individual voxels across stages	60
	of the experiment.	
4.3	Shifting semantic representations due to learning.	62
4.4	Relationship between temporal response profiles, task demands, and semantic tuning stability.	65
4.5	Semantic tuning shifts are partially shared across participants and partially driven by individual differences.	68
4.6	Shared semantic tuning shifts align with the classical language network.	70
4.S1	Stimulus-driven brain activity for additional participants.	76
4.S2	Temporal response profiles across all stages of the experiment for	77
	additional participants.	
4.S3	Comparison of preparatory activity for additional participants.	78
4.S4	Shifting semantic representations due to learning for participant S1.	79
4. S5	Shifting semantic representations due to learning for participant S3.	80
4.S6	Semantic model weights and model performance for participant S4.	81
4. 87	Relationship between temporal response profiles, task demands, and semantic tuning stability for additional participants.	82
4.S8	Correlation values between semantic shift PCs and other semantic PCs.	83

List of Tables

3.S1	Qualitative observations of modality shifts around category-specific ROIs.	46
4.1	Stimuli used in all stages of the experiment.	50
4.2	Words with the strongest projections onto the first three principal	72
	components of the semantic shift space.	

Acknowledgements

Graduate school was an exhausting six years of my life, and I need to thank so many people for helping me survive it all. I am especially grateful to the weird cult-like family I have found in the Gallant Lab. Jack, thank you so much for believing in the crazy projects I proposed throughout the years, and encouraging me through the hard times. The nice things that you said about me will stay in my head forever.

I have loved spending time with every single member of this lab, from the in-depth discussions in lab meetings, to the relaxing times at the Russian River, to the countless lunches in downtown Berkeley. Alex Huth, thank you for guiding me through my first few years of graduate school, and basically acting as a second advisor to me. Cathy Chen, you are the best mentee that I ever could've asked for, and I hope that I was able to help you as much as you deserved. Christine Tseng, thank you for all of the wonderful conversations while running the scanner and showing me that badminton is definitely awesome. Anwar Nunez-Elizalde, I was constantly impressed by your deep knowledge of so many things, and I also want to thank you for just being such a great friend. Overall, I am so grateful that I got to be part of such a big lab and share ideas with all of these other awesome people: Mark Lescroart, Fatma Deniz, Lydia Majure, Leila Wehbe, Natalia Bilenko, James Gao, Michael Oliver, Storm Slivkoff, Robert Gibboni, Michael Eickenberg, Carson McNeil, Tianjiao Zhang, Matteo Visconti (dOC), Tom Dupre La Tour, Michele Winter, Lily Gong, Emily Meschke, and Alicia Zeng. And of course, a special thank you to the members of the lab who went on other adventures with me as members of the Voodoo Rangers.

I also need to thank my family back on the east coast for their support. I wish you had been able to visit more, but sometimes global pandemics get in the way. Dad, thank you for helping me navigate the weird intricacies of academia. Mom, you have to call me Dr. now too, no matter what you think of that! Erik, I'm excited to be able to celebrate this big milestone with you and I'm so happy that you'll be coming out to California more often now.

Thank you to my chosen family across the country, especially Kayla Kipps and Albany Carlson who have always reassured me that I could do this, even if they didn't understand everything that I was doing.

The entire Helen Wills Neuroscience community has been a wonderful support system throughout my time in Berkeley. I honestly don't know if I would've gotten a PhD if I had tried to do it somewhere else. Thank you to Michael Silver, Frederic Theunissen, Nina Dronkers, Bruno Olshausen, and Marty Banks for serving on my qualifying exam and thesis committees and giving amazing advice every time we met. And a huge thank you to Candace Groskreutz for making sure that the administrative nightmares in the university were always solved.

Thank you to the incredible people who I lived with at the Hot Pony Social Club over the years: Alex Naka, Charles Frye, Dan Mossing, Bob Cail, Storm Slivkoff, Sophie Obayashi, and Davis Goodnight. I haven't lived anywhere else that long in my entire adult life, and I will remember my time there with all of you fondly.

Thank you to the teams (sports and otherwise) I was able to be a part of during grad school: The Bunsen Burners, The Volley Llamas, and The Goldheart Companions. Being able to do those kinds of things outside of lab with other brilliant scientists was so important to my sanity.

The closeness of my cohort from the beginning of Neuro Boot Camp made me feel like I belonged somewhere for the first time in a long time. I'm certain that I've started friendships here that will last the rest of my life. Amanda Tose, thank you for never judging me for being weird around you. I'm so happy that we just kept getting closer over the years. Mathew Summers, thank you for introducing me to D&D and causing a terrible obsession. I hope that you stay in California after graduate school so that we can go to a hundred more concerts together. Adam Eichenbaum... I have way too many things to thank you for. Thank you for struggling through that MCB class with me our first semester, thank you for cooking me amazing mac and cheese on my worst days, thank you for encouraging me to always stand up for myself, and thank you for trusting me to help you too.

CHAPTER 1

1.1 Introduction

Humans are constantly bombarded by information from a variety of diverse sources, but somehow we can make sense of all of it. This information comes through all of our sensory modalities and through language, and all of these individual pieces are tied together to come up with coherent semantic percepts with structured meaning. Not only that, but we are able to update these concepts with information that we have gathered throughout our entire lives. The goal of this dissertation is to add just a small amount of knowledge to this huge subfield of cognitive neuroscience, specifically in terms of how multiple semantic systems interact with each other and the manner in which this information is updated with experience. For the purposes of this dissertation, I will define "semantic" information as all of the knowledge that we have about a given object, concept, or word.

In Chapter 2, I provide a detailed background on the state of research on semantic representation in the human brain. This covers a wide range of research modalities, starting with neuropsychological patient work (Desgranges et al., 2007; Diehl et al., 2004; Galton et al., 2001; Hodges et al., 1992; Mummery et al., 2000; Nestor, Fryer, & Hodges, 2006; Snowden, 2015; Snowden et al., 2018; Snowden, Goulding, & Neary, 1989; Rosen et al., 2002; Warrington, 1975), early positron emission tomography (PET; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996), and functional magnetic resonance imaging (fMRI) work (Mummery et al., 2000; Visser, Jefferies, & Lambon Ralph, 2010), and then building up to more recent work that makes use of voxelwise modeling (VM) and more naturalistic stimuli (Çukur, Nishimoto, Huth, & Gallant, 2013; Huth et al., 2012, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019; Naselaris et al., 2011). In this chapter, I pay special attention to these modeling methods and dive deep into the benefits that these modes of study bring to the field. This chapter also briefly introduces the study which is the focus of the following chapter, but does not delve into any methodological details yet.

In Chapter 3, I build off of the naturalistic research studies that were introduced in the previous chapter. Here, I show how data from two separate studies can be combined in a novel manner to look at fine-grain representational details along the boundary where the visual and linguistic semantic networks meet. One of those experiments used silent movies as stimuli and the other used narrative stories. By examining the patterns of semantic selectivity across these two maps, I was able to test the hypothesis that there is a systematic spatial correspondence between the semantic networks selective for visual and linguistic categories. This required that I develop a new method for detecting these shifts from visual to linguistic selectivity of the same semantic category across a broad span of the cortex. Through this method, I was able to find strong evidence that for each location along the anterior border of visual cortex that is selective for that same semantic category in language. This suggests that visual areas are passing information to the linguistic semantic system through semantically-selective channels aligned at the border of visual cortex. This architecture may even support the integration of visual perception and semantic memory.

Finally, in Chapter 4, I focus solely on lexical semantics and how cortical representations of these shift as a result of learning. Using similar techniques to the other naturalistic studies described in this document, I was able to analyze how brain responses to the same set of stimuli shift through time and experience. In this experiment, songs and their lyrics were used as stimuli in a three-stage fMRI study. In the first stage of the experiment, participants listened to a series of 22 songs. Then, over the course of a few months, participants memorized the lyrics to that entire set of 22 songs. In the second stage of the experiment, the participants were instructed to internally sing the lyrics to each song while listening to instrumental-only tracks. Finally, in the third stage of the experiment, participants listened to all 22 songs again. By analyzing the shifts in temporal response profiles across the three stages of the experiment, we saw that after learning, much more of the brain was engaged well before the stimulus was presented. However, it was also clear that the brain networks which exhibited this anticipatory response were distinct for the second and third stages of the experiment. This indicates that it is not only learning that influences the shifts in the brain's responses, but also task demands. Furthermore, I was able to infer how semantic representations shifted as a result of learning. This is because the only difference between stages one and three of this experiment was that the participants had memorized the stimuli. I find that some of these shifts were shared across all participants in the study, but there is also a great deal of individual variability. While further study is needed to better understand these individual differences, we find preliminary evidence that the differences may be related to pre-experiment familiarity with the stimuli.

I hope that the evidence provided from these two experiments illustrate the importance of studying the complex relationships between different types of semantic representations. Studying these representations in isolation will only allow us to understand a small fraction of how the brain is actually functioning in complex, naturalistic environments.

CHAPTER 2

Semantic representation in the human brain under rich, naturalistic conditions

2.1 Abstract

Conceptual understanding of the world is mediated by a broadly distributed network of brain areas that represent semantic information about our current experience and prior knowledge. Several decades of cognitive neuroscience research suggest that semantic processing in the natural world is supported by three distinct subsystems: modality-specific semantic representations are located in sensory and motor areas; amodal semantic representations are located in association areas; and the prefrontal cortex exercises the cognitive control required to understand rich semantic content in context. In this chapter we briefly review the large body of work on semantic representation. We then examine current views of semantic representation in light of a recent series of studies in which brain activity was recorded while individuals performed naturalistic tasks, such as listening to stories or watching movies. These studies revealed that semantic information is represented in an intricate mosaic of semantically selective regions that are mapped continuously across much of the human cerebral cortex and are highly consistent across individuals. These data have two profound implications for current views of semantic representation. First, they indicate that modal sensory information likely enters the amodal semantic system through multiple routes. Second, they suggest that current views that the prefrontal cortex does not directly represent semantic information need to be revised. These data suggest that the semantic system is a hybrid network in which connections between modal sensory areas and amodal semantic representations bind information about current experience, in parallel with a separate system for semantic memory access mediated by the anterior temporal lobes.

2.2 Introduction

Natural human behavior is based on a complex interaction between immediate sensory experience, stored knowledge about the natural world, and continuous evaluation of the world relative to our own plans and goals. Even seemingly simple tasks, such as watching a movie or listening to a story, likely involve a range of different perceptual and cognitive processes whose underlying circuitry is broadly distributed across the brain. When watching a movie, we integrate visual and auditory information into a perceptual whole; we recognize the objects and actions in the movie and the intentions of the actors; and we understand the narrative arc of the story as it develops over time. When reading a book, we can still comprehend the story and its narrative arc even though the perceptual information available to us is greatly reduced compared with a film of the same story. A large body of research indicates that these remarkable capacities are underpinned by a broadly distributed network of brain areas that represents and processes information relevant to different parts of these tasks (Binder, Desai, Graves, & Conant, 2009; Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Huth, Nishimoto, Vu, & Gallant, 2012). In this review we focus on one specific aspect of this system, the representation of conceptual information about the world: semantics (Binder et al., 2009; Martin & Chao, 2001; Patterson, Nestor, & Rogers, 2007; Ralph, Jefferies, Patterson, & Rogers, 2017).

The question of how the brain represents semantic information has been an intense topic of research in cognitive neuroscience for the past 40 years. Much of the early work on this topic involved neurological patients with temporal lobe degeneration, which causes a syndrome called semantic dementia (Hodges, Patterson, Oxbury, & Funnell, 1992; Snowden, 2015; Warrington, 1975; Wilkins & Moscovitch, 1978). About 25 years ago, researchers began to use neuroimaging to investigate this issue, first with positron emission tomography (PET; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Diehl et al., 2004) and later with functional magnetic resonance imaging (fMRI; Mummery et al., 2000; Visser, Jefferies, & Lambon Ralph, 2010). These studies, and the subsequent research reviewed below, support the idea that semantic processing in the natural world is supported by three distinct subsystems. First, modality-specific semantic representations are located in sensory and motor areas. Second, amodal semantic representations are located in areas, though the precise location and nature of these representations are more controversial. Third, prefrontal cortex appears to be involved in the cognitive control required to understand rich semantic content in context.

In this chapter we will first review the existing literature on each of these three aspects of semantic representation. Then we will summarize findings on semantic representation that have grown out of recent naturalistic experiments and evaluate how these data fit into existing theories.

2.3 Modality-specific semantic representations

Both lesion studies and neuroimaging experiments support the view that modalityspecific semantic representations are distributed in a network of distinct sensory and motor areas. Lesion studies have shown that individuals who have suffered stroke often exhibit modalityspecific comprehension deficits, such as pure word deafness (Auerbach, Allard, Naeser, Alexander, & Albert, 1982; Kussmaul, 1877) or visual agnosia (Farah, 2004; Riddoch & Humphreys, 1987). Neuroimaging studies using positron emission tomography (PET; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996) and functional magnetic resonance imaging (fMRI; Chao, Haxby, & Martin, 1999; Goldberg, Perfetti, & Schneider, 2006; Hauk, Johnsrude, & Pulvermüller, 2004) both indicate that modality-specific semantic information is represented in a network of brain areas broadly distributed across sensory and motor cortex. For example, watching a close-up of a Western gunfighter pulling his weapon out of its holster would produce activity in visual areas that represent body parts (Nishimoto et al., 2011) and in premotor areas that represent the hand (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004). Modality-specific representations have been identified in the visual and auditory systems, around the precentral and postcentral gyri, and across much of the ventral temporal cortex.

These data have been used to support the view that semantic information is represented in a distributed form in the network of sensory and motor areas that serve as the source and sink for all human interactions with the world (Barsalou, 1999; Martin, 2007; Pulvermüller, 2013). According to this view, semantic concepts arise from connections between these distributed modality-specific representations (Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012). This family of theories is usually called embodied or grounded cognition. While the theory of embodied cognition is broadly consistent with a large body of data, one area of contention

concerns how such a system can represent abstract semantic concepts that have no direct sensory or motor correlates, such as truth, justice, and love (Meteyard et al., 2012; Vigliocco, Meteyard, Andrews, & Kousta, 2009).

2.4 Amodal semantic representations

Other lesion and imaging data suggest that semantic information is also represented in an amodal form that is not closely tied to sensory or motor representations. Most importantly, some neurodegenerative diseases or brain lesions appear to affect semantic judgment regardless of modality. The most profound of these disorders is semantic dementia, which causes a progressive bilateral atrophy of the anterior temporal lobes (ATL; Desgranges et al., 2007; Diehl et al., 2004; Galton et al., 2001; Hodges et al., 1992; Mummery et al., 2000; Nestor, Fryer, & Hodges, 2006; Snowden, 2015; Snowden et al., 2018; Snowden, Goulding, & Neary, 1989; Rosen et al., 2002; Warrington, 1975). ATL degeneration results in deficits in the amodal conceptual representations of words, pictures, sounds, smells, and actions (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000; Bozeat, Ralph, Patterson, & Hodges, 2002; Garrard & Carroll, 2006; Jefferies, Patterson, Jones, & Lambon Ralph, 2009; Luzzi et al., 2007; Schwartz, Marin, & Saffran, 1979; Wilkins & Moscovitch, 1978). Individuals with ATL degeneration also suffer from anomia and cannot name concepts based on the sensory evidence provided. For example, a patient with anomia might identify a zebra as a horse and express confusion about the presence of stripes (Patterson, Nestor, & Rogers, 2007). However, other aspects of cognition (syntax, numerical abilities, executive function) appear to be relatively spared (Jefferies, Patterson, Jones, Bateman, & Lambon Ralph, 2004; Hodges et al., 1992, 1999; Kramer et al., 2003). These profound semantic deficits are not observed in other neurodegenerative diseases that affect the hippocampus, parahippocampal cortex, and limbic structures, areas more closely involved with autobiographical memory than with semantic memory (Chan et al., 2001). In sum, degeneration of the temporal lobe is a key cause of semantic dementia. However, several aspects of this disorder are still in dispute.

First, there is some controversy about the organization of semantic representations along the temporal lobe. Some studies argue that degeneration of the most anterior regions of the temporal lobe produce the most profound deficits of semantic comprehension and that the degeneration of more posterior regions does not affect semantic judgment (Nestor, Fryer, & Hodges, 2006). Others have argued that the degradation of posterior regions is involved in semantic dementia (Galton et al., 2001) or rather that connections between posterior and anterior temporal lobe regions are in fact more critical for semantic judgment than the anterior regions (Martin & Chao, 2001; Mummery et al., 1999).

Another point of contention in studies of semantic dementia concerns whether this disease affects semantic comprehension in general (Lambon Ralph, Graham, Patterson, & Hodges, 1999) or whether it is mainly a deficit of lexical semantics (Lauro-Grotto, Piccini, & Shallice, 1997). The answer to this question has profound implications for any theory of semantic representation. The first case would indicate that the ATL is a critical hub for semantic comprehension, while the second would imply that the ATL is a critical interface mediating between perceptual and language systems. However, the evidence bearing on this issue is still mixed. Some studies have argued that this disorder impairs representations of categories of

concrete objects but that verbs and abstract concepts are relatively spared (Breedin, Saffran, & Branch Coslett, 1994; Silveri, Brita, Liperoti, Piludu, & Colosimo, 2018). Others argue that representations of concrete categories, verbs, and abstract concepts are all degraded equally in semantic dementia if the base-rate frequencies for the exemplars used in testing are all equated (Bird, Lambon Ralph, Patterson, & Hodges, 2000; Ralph, Graham, Ellis, & Hodges, 1998). However, whether this impairment occurs at the level of concepts or the linguistic representations of those concepts is still unclear (Caramazza & Mahon, 2003; Kiefer & Pulvermüller, 2012).

Additionally, individuals with semantic dementia appear to lose finer categorical distinctions first and then coarser categorical distinctions at later stages of the disease (Ralph, Sage, Jones, & Mayberry, 2010; Lambon Ralph & Patterson, 2008). For example, someone with mild semantic dementia might be able to identify a picture of a robin as a bird but could be confused when presented with an ostrich (see Patterson, Nestor, & Rogers, 2007). Then, with further progression of the disease, the person would become unable to identify any bird. This pattern of deficits has been used to support the idea that semantic dementia impairs access to information about the hierarchical categorical structure of the world (Garrard, Ralph, Hodges, & Patterson, 2001; Laisney et al., 2011).

Much of the recent work on semantic dementia has proposed that the modality-specific semantic representations in sensory and motor areas serve as spokes that feed into a single semantic hub located in the ATL (Ralph et al., 2017). However, an older, alternative view suggests that multiple semantic convergence zones outside of the ATL serve as interfaces between different areas of unimodal semantic representations (A. R. Damasio, 1989; Damasio et al., 1996; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Devereux, Clarke, Marouchos, & Tyler, 2013; Fairhall & Caramazza, 2013). This earlier idea proposes that different convergence zones mediate the interaction of different kinds of information, based on anatomical constraints and individual life experiences. A meta-analysis of over 120 studies of semantic representation in the brain identified a set of putative high-level convergence zones, including the angular gyrus; middle temporal gyrus; precuneus, fusiform, and parahippocampal gyri; and some portions of frontal cortex (Binder et al., 2009). When tested directly, the posterior middle temporal gyrus, angular gyrus, and precuneus were found to be responsive to both visual and linguistic stimuli of the same categories, lending support to the argument that they may function as high-level convergence zones (Fairhall & Caramazza, 2013). At this time it remains unclear whether these convergence zones support sensory integration or memory access and precisely how their functional properties differ from the ATL.

2.5 Control processes for semantic comprehension

Substantial evidence suggests that regions of prefrontal cortex, particularly the inferior frontal gyrus (IFG), play a role in controlling the processes that mediate semantic judgments. Early PET and fMRI studies of semantic processing suggested that some prefrontal cortex areas are specifically involved in semantic retrieval, rather than serving as general-purpose cognitive-control regions (Demb et al., 1995; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995). This theory was further supported by reports that neurodegenerative diseases and lesions that affect

the prefrontal cortex but leave the temporal cortex intact sometimes cause semantic deficits (Jefferies & Lambon Ralph, 2006). A more recent study argued that the IFG mediates decisionmaking only in semantic contexts but is not involved in other difficult decision-making processes (Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011). Finally, it has been argued that the prefrontal cortex contains specific regions that mediate semantic judgments but remain completely separate from the regions involved in cognitive control (Fedorenko, Behr, & Kanwisher, 2011).

In contrast, other studies of patients with lesions to the prefrontal cortex have reported that semantic deficits tend to be expressed only in tasks with relatively greater executive demands, such as comprehension of a complex narrative (Jefferies & Lambon Ralph, 2006). This suggests that prefrontal lesions do not affect semantic representations directly. Instead, they affect control processes that govern how semantic information is accessed, sequenced, and integrated (Jefferies & Lambon Ralph, 2006; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). Consistent with this, in cognitively normal subjects the IFG is engaged during the comprehension of sentences that are semantically ambiguous (Bedny & Thompson-Schill, 2006; Rodd, Davis, & Johnsrude, 2005), and its activity is modulated by the difficulty of a semantic decision-making task (Roskies, Fiez, Balota, Raichle, & Petersen, 2001). A meta-analysis also revealed that the IFG is recruited in language tasks that require nonsemantic judgments (Bookheimer, 2002). Finally, there is evidence that the left inferior prefrontal cortex (LIPC) is important for the retrieval of task-relevant information, regardless of whether the task requires semantic information (Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). This is supported by the finding that the LIPC is more engaged when participants are presented with semantic violations and violations of factual knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004).

In sum, a wide variety of lesion and neuroimaging studies suggest that prefrontal cortex is involved in cognitive-control and selection processes rather than semantic representation per se (Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005; Gold et al., 2006). However, this interpretation has not received unanimous support (Nozari & Thompson-Schill, 2016).

2.6 Recent Studies of Semantic Representation

Until recently, much of the debate regarding semantic representation has focused on where semantic information is represented (Humphries, Binder, Medler, & Liebenthal, 2007; Patterson, Nestor, & Rogers, 2007; Visser, Jefferies, & Lambon Ralph, 2010), rather than precisely how semantic information is mapped across the cerebral cortex. Furthermore, the studies that have attempted to understand where some specific type of semantic information is represented have used classical experimental paradigms that manipulate a few semantic parameters under highly controlled and simplistic conditions (Binder, Westbury, McKiernan, Possing, & Medler, 2005; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). While simple controlled studies have ample statistical power to identify specific semantic representations, they lack the power to support broad mapping of the semantic space. Our lab has taken a different approach to understanding semantic representations by using brain activity evoked by complex, naturalistic stimuli to create quantitative, high-dimensional models of semantic selectivity (Naselaris, Kay, Nishimoto, & Gallant, 2011; Wu, David, & Gallant, 2006).

This approach allows us to create rich, high-dimensional maps of semantic selectivity across the entire cerebral cortex (Çukur, Nishimoto, Huth, & Gallant, 2013; Huth et al., 2012, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019; Naselaris et al., 2011; Popham et al. 2021).

Our experiments are based on a naturalistic, data-driven approach designed to reveal how semantic information is represented in individuals watching movies or listening to stories. Thus, our experiments are quite different from those usually used to study semantic representation, which often involve very reduced tasks such as naming pictures or defining words (Patterson, Nestor, & Rogers, 2007). We analyze these rich data by means of a powerful statistical approach called voxelwise modeling (Naselaris et al., 2011). The procedure proceeds in several steps (see Figure 2.1). First, semantic features-objects and actions in movies and stories-are extracted from the stimuli and encoded in an appropriate semantic feature space. Each of the semantic features is used as a regressor in a regularized (ridge) regression procedure run separately for each of the approximately 50,000-100,000 voxels in each individual's brain. Our methods allow us to model thousands of semantic features simultaneously, providing a means to answer many questions about semantic representations in parallel. Second, the output of this procedure produces a separate weight vector for every voxel that describes how each semantic feature contributes to measured brain activity within that voxel. Features present in a movie or story that tend to elicit activity from a voxel will be given positive weights; features whose presence or absence has no effect on a voxel's response will be given zero weights; and features that tend to suppress a voxel's response when present will be given negative weights. Third, the semantic model of each voxel is tested using a separate data set reserved for this purpose. The model predicts how the voxel will respond to the new stimulus, and this prediction is compared to the voxel's actual response to the stimulus as measured by fMRI. Prediction accuracy is quantified by the correlation between the prediction and the observed response, and statistical significance is assessed by permutation testing. The end result is a list of semantic features that significantly modulate activity in each cortical voxel, ordered by the influence of each feature on voxel responses. This entire procedure is performed separately for each voxel in each participant. Finally, the fit voxelwise models are examined to understand how semantic features are represented across the cerebral cortex. The simplest method for this is to use principal component analysis to find a low-dimensional semantic space that best accounts for the data. An inspection of these principal components reveals the relative importance of each semantic feature within the semantic space. The principal components can also be visualized on the cortical surface to reveal how the dimensions of the semantic space are mapped across the surface of the cerebral cortex. Comparing these maps across participants shows which aspects of semantic representation are common at the group level and which reflect individual differences.



Figure 2.1. Voxelwise modeling procedure.

Functional MRI data are recorded while participants listen to natural stories or watch natural movies. These data are separated into two sets: a training set used to fit voxelwise models and a separate test set used to validate the fit models. Semantic features are extracted from the stimuli in each data set. Left, For each separate voxel, ridge regression is used to find a model that explains recorded brain activity as a weighted sum of the semantic features in the stories. Right, Prediction accuracy of the fit voxelwise models is assessed by using the model weights obtained in the previous step to predict voxel responses to the testing data and then comparing the predictions of the fit models to the obtained brain activity. Statistical significance of predictions and of specific model coefficients is assessed through permutation testing.

We have used voxelwise modeling to recover semantic representations from brain activity recorded during several different naturalistic paradigms: while participants were presented with a series of natural photographs (Naselaris et al., 2011); while they watched a series of very short (~20 seconds each) natural movie clips (Huth et al., 2012); while they listened to natural narrative short stories (Huth et al., 2016); while they read a text version of these same narrative stories (Deniz, Nunez-Elizalde, Huth, & Gallant, 2019) while they watched natural short films with sound (Nunez-Elizalde, Deniz, Gao, & Gallant, 2018); and while they watched short films while attending to the presence of vehicles or humans (Çukur et al., 2013). All these studies show that semantic information is represented in an intricate mosaic of semantically selective regions that are mapped continuously across much of the human cerebral cortex and which are highly consistent across individuals. (For the purposes of this chapter, a semantic region is a patch of cortex with fairly uniform semantic tuning, whether unimodal or

amodal.) For example, numbers appear to be represented in a collection of semantic regions distributed broadly across the cerebral cortex (dark green patches, Figure 2.2). Social concepts appear to be represented in a different collection of semantic regions distributed broadly across the cerebral cortex (bright red patches, Figure 2.2). However, there is no obvious systematic relationship between the distribution of the semantic regions pertaining to one domain versus another.

Furthermore, the semantic maps produced in these studies appear to be largely consistent regardless of whether they were acquired during listening to stories or during reading (Deniz, Nunez-Elizalde, Huth, & Gallant, 2019) This consistency is found across a broadly distributed set of regions, including posterior cingulate cortex, parahippocampal cortex, the temporal lobes, posterior parietal cortex, the temporo-parietal junction, dorsolateral prefrontal cortex, ventromedial prefrontal cortex, and orbitofrontal cortex. The only regions that produce inconsistent maps across reading and listening are primary sensory and motor regions, an unsurprising result.



Figure 2.2. Semantic maps obtained from participants who listened to narrative stories.

Principal components analysis of voxelwise model weights reveals four important semantic dimensions in the brain.

(A) An RGB color map was used to color both words and voxels based on the first three dimensions of the semantic space. Words that best matched the four semantic dimensions were found and then collapsed into 12 categories using k-means clustering. Each category was manually assigned a label. The 12 category labels (large words) and a selection of the 458 best words (small words) are plotted here along four pairs of semantic dimensions. The largest axis of variation lies roughly along the first dimension and separates perceptual and physical categories (tactile, locational) from human-related categories (social, emotional, violent).

(B) Voxelwise model weights were projected onto the semantic dimensions and then colored using the same RGB color map. Projections for one participant (S2) are shown on that participant's cortical surface. Semantic information seems to be represented in intricate patterns across much of the semantic system. White lines show conventional anatomical and/or functional ROIs. Labeled ROIs in prefrontal cortex reflect the typical anatomical parcellation into seven broad regions: dorsolateral prefrontal cortex (dlPFC), ventrolateral prefrontal cortex (vlPFC), dorsomedial prefrontal cortex (dmPFC), ventromedial prefrontal cortex (vmPFC), orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), and the frontal pole (FP). Each of these conventional prefrontal cortex in semantic comprehension is more complicated than the current cognitive-control view would suggest. Reproduced and modified from Huth et al. (2016).

Finally, we find evidence for both modal and amodal semantic regions (Huth et al., 2012, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). Modal regions appear to be located in higher-order sensory areas in the occipital and temporal lobes and in motor areas between the motor strip and prefrontal cortex (see Figure 2.2). Amodal regions are located predominantly in the posterior parietal cortex, temporo-parietal junction, dorsolateral prefrontal cortex, ventromedial prefrontal cortex, and orbitofrontal cortex.

As discussed earlier, many previous studies have identified semantically selective regions of interest (ROIs) in many different locations across the cerebral cortex, such as the fusiform face area (FFA; Kanwisher, McDermott, & Chun, 1997), the parahippocampal place area (PPA; Epstein & Kanwisher, 1998), and so on. These regions identified previously also appear in our functional maps. However, our studies also reveal a rich, continuous pattern of semantically selective regions that have not been identified previously. Furthermore, we find that many of the classical functional ROIs located within visual cortex are actually composed of several subdivisions. For example, the FFA contains three spatially segregated functional subregions that differ primarily in their responses for non-face categories, such as animals, vehicles, and communication verbs (Çukur et al., 2013). Three place-selective ROIs—the PPA, the retrosplenial cortex (RSC), and the occipital place area (OPA, also called the transverse occipital sulcus)—each contain two functional subregions, one selectively biased toward static stimuli and one biased toward dynamic stimuli (Çukur, Huth, Nishimoto, & Gallant, 2016). The temporoparietal junction (TPJ) is a broad region usually thought to represent information related to theory of mind and social meaning (Saxe & Kanwisher, 2003), but our data suggest that the TPJ encompasses many separate semantic regions that represent different aspects of social information (Huth et al., 2016). Cognitive-control regions within the prefrontal cortex, such as the dorsolateral prefrontal cortex (DLPFC), are quite large, but our data show that each of these ROIs may contain several distinct semantic regions (see Figure 2.2B).

2.7 Implications of Recent Studies for Current Theories

Taken together, the results from our studies have important implications for two key aspects of current theories regarding semantic representation: the role of the ATL as a semantic hub and the role of prefrontal areas in semantic processing.

The anterior temporal lobe as a semantic hub

As explained earlier, the current hub-and-spoke theory of semantic representation holds that the ATL serves as a hub that integrates distributed semantic representations. This view proposes that all information flowing between the unimodal and amodal semantic systems passes through the ATL. The studies from our laboratory do not offer much new information about semantic representation in the ATL itself. The ATL is difficult to image using fMRI (Binder et al., 2011; Visser, Jefferies, & Lambon Ralph, 2010), and correlations between ATL lesions and semantic deficits seen with PET are not readily apparent with fMRI (Devlin et al., 2000). Functional imaging of the ATL requires specialized protocols that can reveal ATL function but substantially lower image quality in the rest of the brain. Our laboratory chooses imaging protocols designed to optimize image quality across the entire cortex and thus the image quality in the ATL in our previous studies has been poor. For this reason, our data are agnostic about semantic representation within the ATL.

However, our data suggest that the ATL may not be the sole route for information flow through the semantic system and between modal and amodal representations. Instead, we suspect that there are multiple routes for modal semantic information to enter the amodal semantic system. In recent work we compared maps obtained when individual participants watched brief movie clips versus when they listened to stories (Huth et al., 2012, 2016; Popham et al., 2021). We found that the representations of semantic information received through the visual modality and information received through the linguistic modality abut one another just anterior to occipital cortex (see Figure 2.3). Furthermore, the arrangement of semantically selective regions along this border corresponds between vision and language. That is, for each patch of semantically selective visual cortex lying posterior to this border, there is another patch of semantically selective cortex immediately anterior to the border that responds to the same semantic content when it occurs in stories. It seems unlikely that this very specific arrangement would arise by chance; it seems more likely that some relationship exists between semantically selective regions on each side of this border. A well-known principle of cortical anatomy holds that nearby structures are relatively more likely to be anatomically connected than more distant structures. Therefore, we suspect that this arrangement is evidence of a direct parallel pathway that connects visual to lexical representations in the same semantic regions. This conflicts with a

basic assumption of the hub-and-spoke model of the ATL, which holds that all modal semantic information must pass through the ATL in order to enter the amodal system (Ralph et al., 2017). Our result is more in line with the theory of multiple high-level convergence zones (Damasio & Damasio, 1994; Devereux et al., 2013; Fairhall & Caramazza, 2013).



Figure 2.3. Relationship between visual and linguistic semantic representations along the boundary of visual cortex.

The black boundary indicates the border between cortical regions activated by brief movie clips versus stories. Voxels posterior to the boundary (i.e., nearer the center of the figure) are activated by movie clips but not stories. Voxels anterior to the border are activated by stories but not movie clips. Each of the voxels activated by only one modality is colored based on fit model weights that indicate the semantic category for which it is selective (legend at right; data from Huth et al. [2012] and Huth et al. [2016]). For almost all semantic concepts, the semantic selectivity of voxels posterior to the boundary is similar to the semantic selectivity of voxels anterior to the boundary. The only exception seems to be "mental" concepts (purple voxels located in the dorsal region of the boundary in the right hemisphere), which appear to be represented only in the stories. However, these concepts were not labeled explicitly in the movies and therefore cannot be found in the visual semantic map.

Cognitive control of semantic access and use in prefrontal cortex

As summarized earlier, it is well known that regions of prefrontal cortex become activated under conditions requiring the integration or use of complex semantic information but that prefrontal activation is much reduced under conditions requiring only simple semantic judgments. In contrast, lesions or degeneration of the ATL interferes with all semantic judgments, regardless of task complexity (Hodges et al., 1999). For these reasons, prefrontal cortex is not usually thought to be a primary site of semantic representation. Instead, it is thought to control the sequencing, ordering, access, and use of semantic information (Jefferies & Lambon Ralph, 2006). This idea is consistent with the common view of prefrontal cortex as a major site of cognitive control (Badre et al., 2005; Gold et al., 2006).

Several different lines of evidence from our studies suggest that this conventional view of the role of prefrontal cortex in semantic tasks may be oversimplified. The current view holds that the regions of prefrontal cortex responsible for cognitive control do not represent specific semantic information. However, our data show that prefrontal cortex is highly semantically selective during naturalistic semantic tasks (Huth et al., 2012, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). The intricate pattern of semantic selectivity found in prefrontal cortex varies on a scale much finer than would be predicted based on the conventional parcellations of prefrontal cortex (see Figure 2.2B). The current view predicts that activity in prefrontal cortex should depend only on the task requirements and not semantic content. In contrast, the semantic maps that we have obtained during reading and listening appear to be very similar (Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). Furthermore, unpublished preliminary data from our lab suggest that answering questions about specific semantic categories produces patterns of prefrontal activity that can be predicted by semantic selectivity during narrative comprehension. Finally, attention alters semantic selectivity in prefrontal cortex even under constant task conditions (Cukur et al., 2013). If prefrontal areas were involved in cognitive control exclusive of semantic content, then these results should not occur.

Taken together, our data suggest three different possibilities regarding the nature of semantic selectivity in prefrontal cortex. First, cognitive-control areas might be organized at a scale finer than currently believed so that each semantically selective region in prefrontal cortex has its own associated cognitive-control network. Second, cognitive-control areas might be interdigitated with semantically selective regions. Third, cognitive-control areas might be functionally distinct from, but overlap, semantically selective regions. Further studies will be required to determine which of these hypotheses is correct. One way to address this issue would be to obtain semantic maps simultaneously with cognitive-control localizers within the same set of participants.

2.8 Summary and conclusion

Data from naturalistic fMRI experiments in which participants watch movies or listen to stories largely support a distributed view of semantic knowledge. Semantic comprehension appears to involve a large network of surprisingly specific semantic regions that are distributed broadly across most of the cerebral cortex (Huth et al., 2012, 2016). Areas located nearer primary sensory areas appear to represent semantic information within a specific sensory modality, while those located farther from primary sensory areas and in prefrontal cortex appear to represent amodal semantic information. However, our experiments reveal that the structure of these semantic maps is far richer and more detailed than previously suspected. This detail is most prominent in areas that represent amodal semantic information outside the ATL, such as the

temporo-parietal junction, parietal cortex, and prefrontal cortex.

Parietal areas are thought to be a key part of the network for directed attention (Farah, Wong, Monheit, & Morrow, 1989; Lynch, Mountcastle, Talbot, & Yin, 1977; Posner, Walker, Friedrich, & Rafal, 1987), and we speculate that perhaps semantic selectivity in parietal regions reflects semantically selective attentional demands of perception under natural conditions (Çukur et al., 2013). Semantic selectivity in prefrontal cortex is thought to reflect the operation of cognitive-control processes required for sequencing and organizing semantic information under natural conditions (Badre et al., 2005; Gold et al., 2006; Jefferies & Lambon Ralph, 2006). However, this explanation cannot account for the rich organization of semantic domains within prefrontal cortex.

Our data also show a close correspondence between semantic maps along the anterior border of the visual system and along the posterior border of the semantic system that is activated during naturalistic comprehension (Popham et al., 2021). This correspondence suggests that these areas may communicate directly along pathways that are independent of the ATL. Given the strong evidence that the ATL serves as a semantic hub, it seems unlikely that these direct connections are sufficient to provide semantic assignment to sensory experience. We propose that these connections provide the pathways necessary to bind information from different sensory modalities to each other, in parallel to the memory access processes mediated by the ATL. This explanation would reconcile the results found in support of both the ATL as a semantic hub and the existence of multiple high-level convergence zones. In other words, the hub-and-spoke model and the convergence zone model of semantic representation may merely describe different phases of semantic comprehension.

CHAPTER 3

Visual and linguistic semantic representations are aligned at the border of human visual cortex

3.1 Abstract

Semantic information in the human brain is organized into multiple networks, but the fine-grain relationships between them are poorly understood. We compared semantic maps obtained from two fMRI experiments in the same participants: one that used silent movies as stimuli and another that used narrative stories. Movies evoked activity from a network of modality-specific, semantically-selective areas in visual cortex. Stories evoked activity from another network of semantically-selective areas immediately anterior to visual cortex. Remarkably, the pattern of semantic selectivity in these two distinct networks corresponded along the boundary of visual cortex: for visual categories represented posterior of the boundary, the same categories are represented linguistically on the anterior side. These results suggest that these two networks are smoothly joined to form one contiguous map.

3.2 Introduction

Humans can visually recognize thousands of objects and actions in the natural world, and they can communicate and reason about these semantic categories through language. This flexible language capacity suggests that there may be a rich connection between the functional networks that represent semantic information acquired directly through the senses, and semantic information conveyed in spoken language (Barsalou, 1999; Damasio, 1989; Ralph, Jefferies, Patterson, & Rogers 2017). There are currently two prevailing theories of how semantic information from vision, language, and other modalities are combined. The hub-and-spoke view (Ralph, Jefferies, Patterson, & Rogers 2017) holds that unimodal processing units in modalityspecific cortex are independent spokes that converge at the amodal hub in the anterior temporal lobe (ATL). This model is consistent with evidence that ATL degeneration results in semantic dementia (Snowden, Goulding, & Neary, 1989; Warrington, 1975; Wilkins & Moscovitch, 1978), while other aspects of cognition (syntax, numerical abilities, executive function) appear to be relatively spared (Jefferies, Patterson, Jones, Bateman, & Ralph, 2004; Kramer et al., 2003; Hodges, Patterson, Oxbury, & Funnell, 1992; Hodges et al., 1999). In contrast, the convergence zone view (Damasio, 1989; Damasio, Grabowski, Tranel, Hichwa, & Damasio; 1995; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004) holds that modality-specific and amodal semantic representations are combined at multiple points across the cortex, outside of the ATL. This view is supported by evidence that other regions such as the angular gyrus, precuneus, and middle temporal gyrus respond to the same semantic category whether presented either visually or through language (Devereux, Clarke, Marouchos, & Tyler, 2013; Fairhall & Caramazza, 2013).

Functional magnetic resonance imaging (fMRI) studies have provided substantial evidence that visual semantic information is represented as a mosaic of modality- and category-

specific functional areas that are distributed across anterior portions of occipital cortex and posterior temporal and parietal cortex (Kanwisher, McDermott, & Chun, 1997; Epstein & Kanwisher, 1998; Downing, Jiang, Shuman, & Kanwisher, 2001; Huth, Nishimoto, Vu, & Gallant, 2012). Furthermore, we have recently shown that semantic information during narrative language comprehension is represented as a mosaic of category-specific functional areas located anterior to visual cortex (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016), and further work from our lab has shown that the semantic selectivity for most of this mosaic is the same for both listening and reading (Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). Finally, past studies have reported that some regions along the border between these two networks appear to represent the same semantic category when it is presented either visually or through language (Devereux, Clarke, Marouchos, & Tyler, 2013; Fairhall & Caramazza, 2013), a finding we have replicated independently (Figure 3.1; (Huth, Nishimoto, Vu, & Gallant, 2012; Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016)). Therefore, one interesting possibility is that information from the modal visual semantic system enters the amodal semantic system along a set of parallel semantically-selective pathways that are arranged along the border between these networks. Evidence supporting this hypothesis would lend more support to the convergence zone view of semantic cognition, as this border would effectively form a network of convergence zones.

This possibility suggests a strong and novel prediction about the relationship between functional semantic maps anterior and posterior to the border: for each location along the anterior border of visual cortex that is selective for a particular visual category, there should be an area immediately anterior to it that is selective for that same semantic category in language. Hereafter, we will refer to this as the "semantic alignment hypothesis."



Figure 3.1. Voxels with correlated visual and linguistic semantic representations.

(A) Shown here is the flattened cortex around the occipital pole for one typical participant, along with inflated hemispheres. This map shows the correlation of the visual and linguistic model weights for each voxel. Only voxels with significant weight correlations (r[weights] > 0.0625, p < 0.05, two-sided, uncorrected) and high prediction performance of the semantic models (at least one with r[performance] > 0.1) are shown. High correlation values indicate candidate multi-modal regions of the brain. There are clusters of voxels with high correlation values in the precuneus (PrCu) and angular gyrus (AG), which replicate previous findings that these are high-level semantic convergence zones

(B) Flattened occipital lobes for the 10 other participants. All participants show high correlation values in PrCu and AG, further supporting the possibility that these are high-level semantic convergence zones.

3.3 Materials and Methods

MRI Data Collection

MRI data were collected on a 3T Siemens TIM Trio scanner at the UC Berkeley Brain Imaging Center using a 32-channel Siemens volume coil. Functional scans were collected using gradient echo EPI with repetition time (TR) = 2.0045 s, echo time (TE) = 31ms, flip angle = 70° , voxel size = $2.24 \times 2.24 \times 4.1$ mm (slice thickness = 3.5mm with 18% slice gap), matrix size = 100×100 , and field of view = 224×224 mm. Thirty axial slices were prescribed to cover the entire cortex and were scanned in interleaved order. A custom-modified bipolar water excitation radiofrequency (RF) pulse was used to avoid signal from fat. Anatomical data were collected using a T1-weighted multi-echo MP-RAGE sequence on the same 3T scanner.

Participants

Functional data were collected on eleven participants (three female, eight male), between the ages of 23-32. All participants were healthy and had normal or corrected-to-normal vision.

Separation of Exploratory and Confirmatory Analyses

The data analysis procedures were performed completely independently for every one of the 11 participants in the experiment. The only components that were shared across the participants were the stimuli they were presented and the features extracted from those stimuli. Finding these modality shifts along the visual cortex boundary in one individual therefore has no bearing on whether this effect will be seen in another individual.

Many of the pilot analyses were heavily exploratory before the final analysis pipeline was determined. To prevent over-fitting, exploratory analyses were run only on participants 1-5. No analyses were done on participants 6-11 until the workflow was set. Results shown here are the only iteration of analyses run on those 6 participants. All changes to the analysis pipeline that

occurred during the review process also followed this guideline. Alterations to the analysis were run only on participants 1-5 while editing the manuscript, and results for participants 6-11 were viewed only while preparing the revised figures.

Natural Movie Stimuli

Model estimation data were collected in 12 separate 10 minute scans. Movie stimuli consisted of color natural movies drawn from the Apple QuickTime HD gallery (http://trailers.apple.com/) and YouTube (http://www. youtube.com/). Movies were then clipped to 10–20 seconds in length, and the stimulus sequence was created by randomly drawing movies from the entire set. Validation data were collected in nine separate 10 minute scans, each consisting of ten 1 minute validation blocks, taken from the same stimulus set. Each 1 minute validation block was presented ten times within the 90 minutes of validation data. The movies were shown on a projection screen at 24 x 24 degrees of visual angle.

Natural Story Stimuli

The model estimation data set consisted of ten 10- to 15-min stories taken from *The Moth Radio Hour*. In each story, a single speaker tells an autobiographical story in front of a live audience. The ten selected stories cover a wide range of topics and are highly engaging. Each story was played during a separate fMRI scan. The length of each scan was tailored to the story, and included 10s of silence both before and after the story. The model validation data set consisted of one 10-min story, also taken from The Moth Radio Hour. Stories were played over Sensimetrics S14 in-ear piezoelectric headphones.

Fitting Encoding Models

Semantic features used to fit the encoding models were derived from a 985-dimensional word co-occurrence space. To create this, we first constructed a 10,470-word lexicon from the union of the set of all words appearing in the stories and the 10,000 most common words in the large text corpus. We then selected 985 basis words from Wikipedia's List of 1000 Basic Words (contrary to the title, this list contained only 985 unique words at the time it was accessed). This basis set was selected because it consists of common words that span a very broad range of topics. The text corpus used to construct this feature space includes the transcripts of 13 Moth stories (including the 10 used as stimuli in this experiment), 604 popular books, 2,405,569 Wikipedia pages, and 36,333,459 user comments scraped from reddit.com. In total, the 10,470 words in our lexicon appeared 1,548,774,960 times in this corpus.

Next, we constructed a word co-occurrence matrix, M, with 985 rows and 10,470 columns. Iterating through the text corpus, we added 1 to Mi,j each time word j appeared within 15 words of basis word i. A window size of 15 was selected to be large enough to suppress syntactic effects (e.g. word order) but no larger. Once the word co-occurrence matrix was complete, we log-transformed the counts, replacing Mi,j with log(1+ Mi,j). Next, each row of M was z-scored to correct for differences in basis word frequency, and then each column of M was z-scored to correct for word frequency. Each column of M is now a 985-dimensional semantic vector representing one word in the lexicon.

The matrix used for voxel-wise model estimation was then constructed from the stories: for each word-time pair (w,t) in each story we selected the corresponding column of M, creating a new list of semantic vector-time pairs, (Mw,t). These vectors were then resampled at times corresponding to the fMRI acquisitions using a 3-lobe Lanczos filter with the cut-off frequency set to the Nyquist frequency of the fMRI acquisition (0.249Hz).

For the movie stimuli, an observer manually labeled all objects and actions in each 1second clip of the movie using WordNet synsets (Miller, 1995). For each synset (e.g. *bank.n.02*) we then extracted the corresponding set of lemma names (e.g. "depository_financial_institution", "bank", "banking_concern", & "banking_company"), and for multi-word lemma names (e.g. "banking_company") split them on underscores. The resulting list of tokens for each synset were then concatenated with tokens from all other synsets appearing in the same 1-second movie clip. In addition to these annotations, we included textual descriptions of each 1-second scene provided by users on Amazon Mechanical Turk. Finally, to form a 985-dimensional semantic vector for each 1-second clip, we fetched the vector for each word in the full annotation list (including token lists from all the WordNet synsets and the Mechanical Turk annotations) that was in the set of 10,470 word lexicon, and then averaged all of these vectors together. The vectors for the two 1-second clips comprising each 2-second fMRI acquisition were then averaged together.

In addition, we extracted a set of low-level features from each set of stimuli to control for that type of information in each model. For the movies, we extracted a set of 2,139 motion energy features (Nishimoto et al., 2011), in which each filter consisted of a quadrature pair of space-time Gabor filters. For the stories, the 41 low level features were word rate (1 feature), phoneme rate (1 feature), and phonemes (39 features).

Prior to regression, each stimulus feature within each story or movie run was z-scored through time. This was done to match the features to the fMRI responses, which were also z-scored through time for each functional run.

To model our data, we used a modified version of ridge regression called banded ridge (Nunez-Elizalde, Huth, & Gallant; 2019). In this framework, each voxel in the brain is assigned two different ridge parameters: one for the semantic features and one for the low-level features. These pairs of ridge parameters can vary across each voxel, and this allows the model to effectively weight the two sets of features independently across the brain.

A separate linear temporal filter with four delays (1, 2, 3, and 4 time points) was fit for each feature. This was accomplished by concatenating feature vectors that had been delayed by 1, 2, 3, and 4 time points (2, 4, 6, and 8s). Thus, in the concatenated feature space one channel represents the word rate 2s earlier, another 4s earlier, and so on. Taking the dot product of this concatenated feature space with a set of linear weights is functionally equivalent to convolving the original stimulus vectors with linear temporal kernels that have non-zero entries for 1-, 2-, 3-, and 4-time-point delays.

As in previous publications (Huth, Nishimoto, Vu, & Gallant, 2012; Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019), data

were split into a training and test set, and ridge parameters were selected through crossvalidation within the training set. Model performance for both the semantic and low-level submodels was evaluated by multiplying the fit model weights by the features for the test story/movie, then taking the correlation coefficient of that predicted time course per voxel with the actual brain data.

All model fitting was performed using custom software, available as a Python package called *tikreg* (https://github.com/gallantlab/tikreg) (Nunez-Elizalde, Huth, & Gallant; 2019).

Multi-modal Voxels

After separate encoding models were fit for the visual and language experiments in the semantic feature space, semantic tuning across those models was directly correlated within each participant. We first found the average tuning to each feature across delays as a 985-dimensional weight vector per voxel and model. Then, the voxel weights across the two models were directly correlated. The correlation values, which indicate an overlap in semantic tuning across the two modalities, are shown in Figure 3.1.

Calculation of MRI Signal Dropout

MRI signal dropout tends to occur in areas of the brain that are near tissues that are magnetically inhomogeneous, such as air sinuses. This phenomenon is generally quantified by the signal-to-fluctuation-noise ratio (SFNR; (Friedman et al., 2006)). This was calculated for each voxel in each functional run individually by dividing the mean of the signal by its standard deviation. That value was averaged across all runs, and then thresholded at a value of 15. Voxels below this value are tagged as dropout regions and are indicated with hash marks in Figures 3.4 and 3.5.

Modality Shifts of Chosen Categories

Our initial analysis of modality shifts examined only a few semantic categories that are known to be represented in anterior portions of the visual system: places, body parts, and faces. To look at the brain representations of each of these categories across the cortex, we created a generic representation of the categories within the 985-dimensional semantic feature space. First, we constructed a list of words related to each category:

Places: house, building, hotel, office, parking, lot, park, street, road, sidewalk, highway, path, field, mountain

Faces: *face, eyes, nose, mouth, hair, cheek, cheeks, smile, frown, teeth* Body parts: *body, arm, arms, leg, legs, hand, hands, foot, feet, torso, head, back, thigh*

Each word in each list was projected into the 985-dimensional word-embedding feature space. This location is based on each word's co-occurrence with each of the 985 basis words. Then, within each category, all vectors were averaged to get a general 985-dimensional category vector. Finally, the vision and language model weights for each voxel were projected onto each category vector: high visual projections onto the vector were voxels that represented that

category visually, voxels with high linguistic projections were voxels that represented that category linguistically, and voxels where both projections were high represented that category in both modalities. These projection values are shown in the 2-dimensional colormaps shown in Figure 3.2, 3.S2, 3.S3, and 3.S4 for each semantic category separately (i.e. places, faces, and body parts).

Surface-Based Analyses Surrounding Visual Cortex

In order to find modality shifts, we first selected the appropriate regions of the cortical surface using a software package developed in our lab called *pycortex* (Gao, Huth, Lescroart, & Gallant, 2015). In *pycortex*, each participant's cortical surface is formed from a triangular mesh that is made up of approximately 150,000 vertices per hemisphere. To save computation time, vertices were sub-selected from each surface such that any location on the surface was no more than 2.5mm away from a chosen vertex. (A pilot analysis run on one participant indicated that this sub-selection did not impact results; data not shown here.) Since we were only interested in the pattern that exists around the border of visual cortex, we further limited the analysis to a window within 50mm of the defined border of occipital cortex, as shown in Figure 3.S5. This process resulted in about 15,000-20,000 cortical locations per participant.

Next, we searched for modality shifts at all possible angles through each vertex. First, a circular patch on the cortical surface within 22.5mm of the starting vertex was selected. All vertices along the outer edge of the patch were then selected as endpoints of lines. Then, geodesic lines were drawn starting from each endpoint. Each line passed through the center vertex and continued in the same direction until leaving the patch. This was done using the *geodesic_distance* and *geodesic_path* functions in *pycortex* (Gao, Huth, Lescroart, & Gallant, 2015). For each vertex, this process resulted in about 200-300 lines passing through the center at all possible angles (see Figure 3.3A).

Finally, an window around each line was formed and we looked for modality shifts within each window. All vertices within 10mm of the line were selected. We formed a coordinate system within the window based on the vertices along the geodesic line. Each vertex in the window was given a coordinate along the geodesic line (see Figure 3.3B), which was a weighted sum of the line vertex coordinates and the distances to each line vertex:

$$W = \frac{e^{\left(\frac{-D_p}{s}\right)}}{\sum_i e^{\left(\frac{-D_{p,i}}{s}\right)}}$$

$$D_p = \text{pairwise distances between window vertices and line vertices}$$

$$D_p = \text{pairwise distance between line vertices}$$

$$S = \text{smoothing factor for exponential}$$

$$W = \text{weighting value for each vertex}$$

$$C = D_y W$$

$$C = \text{coordinate of each vertex along window}$$

Only the vertices with coordinates within 12.5mm of the original center vertex were retained. This resulted in a total window size of 20mm across by 25mm long, centered on the initial vertex.

Additional analyses were run where the size of the window was 10x25mm and

10x10mm. The results of these analyses are shown in Figures 3.S6 and 3.S7. The results with 10x25mm windows show no discernible difference to those shown in the main test with a 20x25mm window. The results using 10x10mm windows resulted in noisier maps, suggesting that a long axis across the boundary between networks is necessary to reliably detect the shifts. Nonetheless, a similar pattern of significant shifts are still clearly visible.

Definition of Modality Shift Summary Statistic

To quantify how representations of a semantic category changed within an window, we first calculated the average semantic tuning within each region by calculating the mean vision model weights and language model weights for all vertices, within each modality separately. These average tuning vectors were normalized such that their L2-norms were equal to one. Next, for each vertex, the vision and language model weights were projected onto that average visual model weight vector. (This entire process will also be repeated for the average linguistic model weigh vector.) Then, to measure representational shifts along the region, a linear regression model was fit for the weight projections as a function of coordinate along the line (see Figure 3.3C). This was done separately for the movie and story weight projections:

	C=coordinate of each vertex along window
$X - \begin{bmatrix} 1^T & C^T \end{bmatrix}$	w_A = average tuning vector within window
$X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ $V = W^T w$	W_V = visual tuning vector for each vertex within window
$I_V - W_V W_A$	W_L = linguistic tuning vector for each vertex within window
$Y_{L} = W_{L} W_{A}$	V_i = intercept term for visual fit line
$[V_i, V_m] = (X^T X)^{-1} X^T Y_V$	V_m = slope term for visual fit line
$[L_i, L_m] = (X^T X)^{-1} X^T Y_L$	L_i = intercept term for linguistic fit line
	L_m =slope term for linguistic fit line

Finally, we created a single metric that could describe locations where the visual representations were getting weaker and the linguistic representations were getting stronger along the length of the window. Thus, all windows were oriented such that first half of each region was visually responsive and the second half was linguistically responsive. If the results are consistent with the hypothesis, then the ratio of the slopes found above should be around -1. In order to find locations with large overall changes in selectivity, we opted to scale this ratio by the average magnitude of the slopes. Since we only wanted to identify locations at which there was a shift from visual to linguistic representation, we also included an indicator variable which was 1 when the fit lines cross within the analysis window, and was 0 otherwise. Lastly, because we only wanted to identify locations where the semantic tuning across modalities was similar and not diverging from each other, we included an indicator variable which was 1 when the average projection value across all vertices was a positive number, and was 0 otherwise. Thus, those are the components included in our summary statistic for modality shift magnitude (as well as a negation so that the strongest shifts are positive values):

$$R = sign(V_m L_m) min\left(\left|\frac{V_m}{L_m}\right|, \left|\frac{L_m}{V_m}\right|\right)$$

$$M = \frac{(|V_m| + |L_m|)}{2}$$

$$P = \frac{L_i - V_i}{V_m - L_m}$$

$$C = \begin{pmatrix} 1 & min(X) < P < max(X) \\ 0 & otherwise \end{pmatrix}$$

$$Y = \begin{pmatrix} 1 & \left(\left(\sum Y_v\right)_n + \left(\sum Y_i\right)_n\right) > 0 \\ 0 & otherwise \end{pmatrix}$$

$$S = -R * M * C * Y$$

$$V_m = slope term for visual fit line L_m = slope ter$$

This entire process was then repeated for the average linguistic model weight vector, and the stronger of the two metrics was retained.

Significance Testing

For each vertex on the cortical surface, the shift magnitude was evaluated. For each vertex, the strongest metric is plotted on the cortical flatmaps on Figure 3.4. Statistical significance of the putative shifts was evaluated through a permutation test. This test illustrated that the two maps were precisely aligned and did not simply identify locations where large visual weights were near locations with large linguistic weights.

In this test, the shuffled component depended upon which average semantic weight vector was used in the original calculation of the metric (i.e. visual or linguistic). If the average visual semantic weight vector was used, then linguistic information from the windows was permuted. If the average linguistic semantic weight vector was used, then visual information from the window was permuted. The weights for the vertices from the chosen modality, and their respective positions along the window, were swapped with all other possible windows across cortex. Since the number of vertices within a window is variable due to differences in cortical folding, the vertices for each permuted window were selected with replacement to match the original number of vertices for that window. Then, the shift magnitude was calculated for all of these possible permutations. This resulted in a distribution of shift metrics for each window. From this distribution, we obtained a one-tailed p-value for the original shift metric. The p-values for all windows were then FDR-corrected and thresholded at 0.05. Only significant locations after FDR correction are shown in Figures 3.4, 3.5, 3.6, 3.86, and 3.S7.

3.4 Results

To test this semantic alignment hypothesis we compared semantic maps obtained in two

different experiments in the same participants: a vision experiment that used natural movies as stimuli (Huth, Nishimoto, Vu, & Gallant, 2012) and a language experiment that used naturally spoken narrative stories as stimuli (Huth, de Heer, Griffiths, Theunissen, & Gallant; 2016). We used the data from these two fMRI experiments to construct two sets of voxelwise encoding models (VMs; (Kay, Naselaris, Prenger, & Gallant, 2008; Mitchell et al., 2008; Naselaris, Kay, Nishimoto, & Gallant, 2011; Nishimoto et al., 2011)). These models predict blood-oxygen level dependent (BOLD) responses in each voxel in each individual brain, based on the semantic content of the movies and stories. To do this, we first created a 985-dimensional semantic feature space based on the co-occurrence statistics of words in a large corpus of English text (details in Methods). Then, semantic features in the movie experiment were obtained by labeling each object and action in the movies using WordNet (Miller, 1995) and projecting those labels into the 985-dimensional feature space (Huth, de Heer, Griffiths, Theunissen, & Gallant; 2016). Semantic features in the language experiment were obtained by projecting each word in the stories into the same 985-dimensional feature space. Next, for each voxel in each participant, separate visual and linguistic encoding models were estimated through regularized regression. The fit regression weights indicate how each semantic category in the movies or stories modulates BOLD signals evoked from each individual voxel and in every participant separately. Finally, we used these fit models to predict brain activity to new stimuli that were not used to train the models. From these predictions, we could then see whether the activity of each voxel was driven by visual and/or linguistic semantic information (Figure 3.S1). By examining the patterns of semantic selectivity across these two models, we were able to test the hypothesis that there is a systematic spatial correspondence between the semantic networks selective for visual versus linguistic categories.

Preliminary inspection suggests that chosen categories are semantically aligned

Before undertaking a thorough test of the semantic alignment hypothesis, we ran a preliminary test to determine if the hypothesis was plausible. To do this we examined a few semantic categories that are well-represented in anterior portions of the visual system: places, body parts, and faces. Previous studies have found that images of places such as buildings, parks, streets, cities, and mountains selectively elicit responses in the parahippocampal place area (PPA; (Epstein & Kanwisher, 1998)), occipital place area (OPA; (Nakamura et al., 2000; Hasson, Harel, Levy, & Malach, 2003; Dilks, Julian, Paunov, & Kanwisher; 2013)), and retrosplenial complex (RSC; (Aguirre, Zarahn, & D'Esposito, 1998). The semantic alignment hypothesis predicts that areas that represent linguistic descriptions of places should be found in areas neighboring these known place-selective visual regions of interest (ROIs). To identify voxels that encode semantic information related to places in either modality, we used the vision and language model weights to quantify how much place-related information in the movies and in the stories is represented in the activity of each voxel (details in Methods). Figure 3.2A indicates that the voxels that represent place information in movies are concentrated in PPA, OPA, and RSC, while voxels just anterior to these areas appear to represent place information in the stories (see Figure 3.S2 for other participants). A similar examination of semantic selectivity for body parts and faces suggests that these modality shifts from visual to linguistic semantic representations also appear along other portions of the boundary of visual cortex (Figure 3.2B-C, and Figures 3.S3 and 3.S4). For example, voxels that represent body part information in the stories appear to be located just anterior to the extrastriate body area (EBA; (Downing, Jiang, Shuman, & Kanwisher, 2001)), and voxels that represent face information in the stories appear to

be located just anterior to the fusiform face area (FFA; (Kanwisher, McDermott, & Chun, 1997)). Each of these patterns appears in all 11 participants. (See Supp. Table 1 for full evaluation of each participant per ROI.) All of these qualitative observations are consistent with the semantic alignment hypothesis.


Figure 3.2. Visual and linguistic representations of semantic concepts known to be well-represented in visual cortex.

(A) Shown here is the flattened cortex around the occipital pole for one typical participant, along with inflated hemispheres. The color of each voxel indicates the representation of placerelated information according to the legend at the right. The model weights for vision and language are shown in red and blue, respectively. White borders indicate ROIs found in separate localizer experiments. Three relevant place ROIs are labeled: PPA, OPA, and RSC. Centered on each ROI there is a modality shift gradient that runs from visual semantic categories (red) posterior to linguistic semantic categories (blue) anterior.

(B) Format same as (A) except one body ROI is labeled: EBA. EBA also shows a modality shift gradient that runs from visual to linguistic in this example participant.

(C) Format same as (A,B) except one face ROI is labeled: FFA. FFA also shows a modality shift gradient that runs from visual to linguistic in this example participant. All of these qualitative observations are consistent with the semantic alignment hypothesis.

Analysis of modality shift magnitude around entire boundary of visual cortex

We next wanted to test whether a modality shift from visual to linguistic representation of a single semantic category was a general property of the border of visual cortex. However, some portions of the border of occipital lobe are ill-defined; the only clear landmarks are the parieto-occipital sulcus and the preoccipital notch (Ono, Kubik, & Abernathy, 1990). Therefore, we manually defined the entire boundary in each participant individually. This border followed the parieto-occipital sulcus along the dorsal surface, then connected on both ends to the preoccipital notch on the ventral surface. When possible, the border followed sulcal fundi, but otherwise took shortest paths between these landmarks. Because this definition was approximate and not based on functional activations (e.g. this border does not include areas with visual representations in temporal and parietal cortex), later analyses were run on a larger area of the cortex. Thus, the analysis which searched for modality shifts was expanded to all vertices within 50mm of the drawn border. The extent of the areas selected for further analysis in each participant is shown in Figure 3.S5.

To quantify the effects that Figure 3.2 shows qualitatively, we designed a method to search exhaustively for all locations where there was a modality shift from visual to linguistic representation of a single semantic category (Figure 3.3). In our analysis, each participant's cortical surface is formed from a triangular mesh that is made up of approximately 150,000 vertices per hemisphere. We chose to search for modality shifts at vertices within the analysis region shown in Figure 3.S5. Modality shifts were calculated at all possible angles through each location using an algorithm that maps geodesic lines onto the surface of the brain (Figure 3.3A). Next, a window around each line was selected and we looked for a modality shift within each window. Each vertex in the window was given a coordinate along the geodesic line (Figure 3.3B).



Figure 3.3. Method for detecting category-specific modality shifts.

Modality shifts were estimated at thousands of locations near the boundary of the occipital lobe for each participant individually. Calculating the modality shift metric required three steps.

(A) First, for each location, possible modality gradient axes were identified by generating geodesic lines in all directions centered on the location. Here one example location is shown in red and geodesic lines are shown in black.

(B) Second, for each geodesic line, a window was centered on the line and each vertex in the window was assigned a coordinate. One example is shown here, where each vertex is colored according to its coordinate.

(C) Third, the magnitude of the modality shift along the length of each window was estimated in terms of a summary metric, S. The average semantic concept that was represented in each window was found by calculating the mean of the vision and language model weights across all vertices in the window. Then, the visual and linguistic representations of that concept were calculated at each vertex as the projection of the weights of that vertex onto the average weights. In this panel each location is plotted twice, once in red for its vision weight projection and once in blue for its language weight projection. Linear regression was then used to fit lines to these values. The intercepts (Li, Vi) and slopes (Lm, Vm) of these lines were used to derive the modality shift metric, S. Shown here is a region that is fit well by this model. This process is repeated for each location on each brain, and the full results of that analysis are shown in Figure 3.4.

(D-H) These five examples illustrate why the chosen analysis method is optimal. This method selects only regions where there is a strong shift from visual to linguistic representation, and not other related changes in semantic representation. We specifically constructed the modality shift metric, S, to be small in all five of these edge cases.

In order to understand how semantic representations changed both visually and linguistically along the length of the window, we first needed to determine the average semantic concept that was represented within the entire window for each modality. The average semantic tuning within each window was found by calculating the mean vision model weights and language model weights for all vertices, within each modality separately. For each vertex, the vision and language model weights were projected onto that average visual model weight vector. Then, to measure representational shifts along the region, a linear regression model was fit for the weight projections as a function of coordinate along the line (Figure 3.3C). This was done separately for the movie and story weight projections. Finally, we created a single metric that could describe locations where the visual representations were getting weaker and the linguistic representations were getting stronger along the length of the window, and that in fact there is a shift from one modality to another (equation in Figure 3.3C; see Methods for details). This entire process was then repeated for the average linguistic model weight vector, and the stronger of the two metrics was retained. There were some important edge cases that we specifically chose to avoid when constructing the metric. Examples of these types of windows are shown in Figures 3.3D-H, and they all have very small modality shift metrics, as intended. These examples show why this metric is superior to a simpler metric such as correlation of weights across two halves

of the window. In Figure 3.3D, the visual representation is not changing strongly over the course of the window, as indicated by the flat red line. In Figure 3.3E, the same is true for the linguistic representation, as indicated by the flat blue line. In Figure 3.3F, both the visual and linguistic representations do not change much over the course of the window, but the average weights across models are correlated. In Figure 3.3G, the analysis window seems to be approaching the boundary between visual and linguistic representations, but the regression lines for each modality do not cross within the window. In Figure 3.3H, the visual and linguistic representations are actually diverging from each other. The negative examples shown in Figures 3.3D-H indicate that we were successful in creating a modality shift metric that is specific to only the representational shift which is consistent with the semantic alignment hypothesis. These edge cases might have been incorrectly identified as modality shifts through alternative analysis methods.

To evaluate the significance of these modality shifts, we constructed a permutation test which would reveal which shifts were semantically-specific to only the concept represented in that window. A null distribution was obtained by replacing all vertices from one modality, and their respective model weights, with those of other windows across the cortical surface for each participant. In this test, the permuted modality was the one which was not used as the average semantic weight vector in the original calculation of the metric.

Because all eleven participants took part in both experiments, we were able to analyze data within each participant individually, resulting in eleven separate and independent tests of this hypothesis. Additionally, we separated our data for training and validation at two different points in the analysis pipeline. First, the data acquired from each participant were divided into separate fit and test sets for initial encoding model fitting. Then, because the data analysis procedures that we used while refining our methods were heavily exploratory, we restricted all pilot analyses to only participants 1-5. This was done to avoid overfitting and to ensure that our results would generalize to new participants. The results for all 11 participants are presented here, but it is important to note that the analysis pipeline was frozen before it was applied to participants 6-11.

In all eleven participants, we find significant (FDR-corrected, q < 0.05) modality shifts for semantic categories located along almost the entire boundary of visual cortex. These regions include the visual category-selective regions discussed above, as well as other semanticallyselective regions that have not been distinguished as visual ROIs in previous localizer studies, but which we reported earlier are indeed semantically selective (Huth, Nishimoto, Vu, & Gallant, 2012) (Figure 3.4). The only regions where we cannot yet verify this pattern are where MRI dropout obscures functional signals (see purple hatch marks in Figure 3.4). (The dropout regions are voxels with a low signal-to-fluctuation-noise-ratio (SFNR; (Friedman et al., 2006)), and these tend to occur in areas of the brain that are near tissues that are magnetically inhomogeneous, such as air sinuses (Ojemann, et al., 1997). See Methods for details on calculation of SFNR.)



Figure 3.4. Locations of category-specific modality shifts across cortex.

(A) Shown here is the flattened cortex around the occipital pole for one typical participant, along with inflated hemispheres. The modality shift metric calculated at each location near the boundary of the occipital lobe is plotted as arrows on the flattened cortex and on inflated hemispheres. The color of each arrow indicates the magnitude of the shift (a.u. = arbitrary units). The direction of each arrow indicates the shift from vision to language. Arrows are only shown at locations where the modality shift is statistically significant (p<0.05, one-sided, FDR-corrected). Areas of fMRI signal dropout are indicated with purple hatch marks. There are strong modality shifts in a clear ring around the anterior border of visual cortex.

(B) Flattened occipital lobes for the 10 other participants. All participants show the same pattern of strong modality shifts organized into a clear ring around visual cortex.

Figure 3.4 also shows that there are a few regions of modality shifts which are directed anterior to posterior at locations that do not fall along the boundary of visual cortex, such as the right posterior superior temporal sulcus (pSTS) in Figure 3.4A. The modality shifts point in this direction because the visual semantic selectivity in pSTS is similar to the linguistic semantic selectivity of voxels just posterior to that region. However, because these modality shifts do not lie along the boundary of visual cortex, they are not within the scope of this manuscript and therefore will not be discussed further.

The modality shift metric was designed to identify locations where there is a strong shift between two unimodal representations of the same semantic category. However, the correlation of the semantic weights themselves are not directly used in the metric calculation. To directly quantify the semantic correspondence across the boundary that was found, each significant window was divided into visual and linguistic portions. This division occurred at the point where the two fit lines crossed (see variable P in Figure 3.3C). Then, the average of the visual model weights in the visual portion of the window were correlated with the average linguistic model weights in the linguistic portion of the window. The correlation value for each window is shown in Figure 3.5. The vast majority of these correlations are strongly positive, which shows that our modality shift analysis is picking up on semantic correlation across the boundary as expected.



Figure 3.5. Quantitative summary of semantic correspondence across the boundary.

(A) Shown here is the flattened cortex around the occipital pole for one typical participant, along with inflated hemispheres. Arrows indicate the correlation between semantic selectivity of vertices on each side of the visual-linguistic boundary. Red indicates positive correlations, blue indicates negative correlations, and arrow weight indicates the strength of correlation.

(B) Flattened occipital lobes for the 10 other participants. All participants show the same pattern of semantic correlations organized into a clear ring around visual cortex.

It is important to note that the results shown in Figure 3.5 alone are not sufficient support for the semantic alignment hypothesis. The modality shift analysis shown in Figure 3.4 was absolutely necessary to identify the location of the boundary. The negative examples presented in Figures 3.3D-F indicate why this is the case. The correlation value for each of those windows is significant and positive, however, none of these examples are the shifts from visual to linguistic representation that we are looking for. If only the results in Figure 3.5 were presented, we would be drawing incorrect conclusions about the semantic alignment at those cortical locations. Therefore, it is important to consider the results presented in Figure 3.5 only after having completing the modality shift analysis (Figure 3.3 and 3.4). Taken together, Figures 3.4 and 3.5 suggest that the two individual semantic maps are indeed aligned along the anterior border of visual cortex.

The analyses in Figures 3.3-3.5 strongly support the semantic alignment hypothesis. Yet it is not clear from these analyses which semantic categories are actually represented along the visual cortex boundary. To examine this, we plotted a subset of the semantic model weights onto the cortical sheet along this boundary (Figure 3.6). Figure 3.6 shows that a wide variety of semantic concepts are represented along the anterior boundary of visual cortex. Inspection reveals a clear spatial correspondence of the visual and linguistic maps. The only exception seems to be "mental" concepts (purple vertices located in dorsal region of boundary) which appear to be represented only in the stories. However, these abstract concepts were not labeled explicitly in the movies and therefore cannot be found in our visual semantic maps. This does not necessarily mean that the two semantic maps are misaligned at these locations, but future experiments and analyses would be required to resolve this matter.



Figure 3.6. Alignment of semantic selectivity along the boundary between vision and language.

(A) Semantic selectivity of vertices near the visual-linguistic boundary shown on the flattened cortex around the occipital pole for one typical participant. Vertices selective for visual semantics are shown in red, those selective for language semantics are shown in blue, and those selective for both types of information are shown in black.

(B) The flattened cortex around the occipital pole, along with inflated hemispheres, for the same participant shown in (A). Each vertex that is selective for either visual or linguistic categories is colored according to its semantic selectivity. The pattern of semantic selectivity corresponds along both sides of the visual-linguistic boundary in most locations.

(C) Semantic selectivity for the 10 remaining participants, which all show a similar pattern to the participant shown in (B).

3.5 Discussion

The results presented here support the semantic alignment hypothesis, which is that for each location along the anterior border of visual cortex that is selective for a particular visual category, there is an area immediately anterior to it that is selective for that same semantic category in language. This suggests that the border of visual cortex acts as a convergence zone where information from the modal visual semantic system enters the amodal semantic system along a set of parallel semantically-selective pathways. Given the close spatial proximity and correspondence of these modal and amodal semantic maps, we speculate that this functional arrangement likely reflects direct anatomical connections between corresponding modal and amodal semantically-selective areas (Van Essen, Anderson, & Felleman, 1992; Mohda & Singh, 2010; Ercsey-Ravasz et al., 2013). However, we cannot evaluate this possibility with the data available currently.

If these pathways provide a direct route from the visual semantic system into the amodal semantic system, bypassing the ATL, then what are we to make of the extensive evidence that ATL lesions impair semantic recognition (Ralph, Jefferies, Patterson, & Rogers, 2017; Snowden, Goulding, & Neary, 1989; Warrington, 1975; Wilkins & Moscovitch, 1978)? We suspect that the pathways that we have identified here connect modal visual experience to amodal representations, but these amodal representations alone are not sufficient to provide semantic labels to sensory experience. Instead, semantic comprehension also requires input from the memory system. This is provided by pathways that proceed through the ATL. This explanation would reconcile the large body of research supporting both the hub-and-spoke model and the theory of high level convergence zones. In other words, these two theories may merely describe different aspects of semantic comprehension.

We cannot comment on the nature of the semantic representations in the ATL in this study. This is because the fMRI pulse sequences used here were optimized for the cortex as a whole, rather than for specifically recovering signal in the ATL. Because the ATL is particularly susceptible to signal dropout, it is nearly impossible to study the ATL effectively without

specialized pulse sequences (Visser, Jefferies, & Ralph, 2009). In the future, we hope to look deeper into this topic with a targeted study of the ATL using the methods presented here.

We speculate that the precise spatial relationship that we report here between visual semantic maps and amodal language maps may also occur in other modal semantic systems. For example, the auditory semantic maps found in the temporal lobe (Lewis, Talkington, Puce, Engel, & Frum, 2011; Norman-Haignere, Kanwisher, & McDermott, 2015) may be spatially aligned with nearby amodal semantic maps, and the same may be true of unimodal somatosensory semantic maps. Performing detailed semantic mapping in every modality thus has the potential to reveal the entire network of convergence zones that feed modal sensory information into the amodal semantic network.

Our results also raise the interesting possibility that the organization of the modal and amodal semantic systems might influence one another during development. Because the largescale organization of category-selective areas in visual cortex appears to depend on geneticallyencoded gradients such as retinotopy (Levy, Hasson, Avidan, Hendler, Malach; 2001), it seems most likely that the organization of the amodal semantic system is influenced by the visual semantic system. One possible way to test this hypothesis would be to map the semantic representation of narrative language comprehension in congenitally blind participants. If their amodal semantic representations near occipital cortex are organized differently than those found in sighted participants, it would support the idea that organization of the amodal semantic system is shaped by visual semantic system. If not, it might suggest that other factors influence the organization of both systems.

3.6 Supplemental Figures and Table



Figure 3.S1. Evaluation of the visual and linguistic semantic models.

Model weights are estimated on the training dataset, then are used to predict brain activity to a held-out dataset. Prediction performance is the correlation of actual and predicted brain activity for each voxel. These performance values are presented simultaneously using a 2dimensional colormap on the flattened cortex around the occipital pole for each participant. Red voxels are locations where the visual semantic model is performing well, blue voxels are where the linguistic semantic model is performing well, and white voxels are where both models are performing equally well. These maps show where the visual and linguistic networks of the brain abut each other.



Figure 3.S2. Visual and linguistic representations of place concepts.

Identical analysis to Figure 3.2A, but for the other 10 participants. The color of each voxel indicates the representation of place-related information according to the legend at the right. The model weights for vision and language are shown in red and blue, respectively. White borders indicate ROIs found in separate localizer experiments. Three relevant place ROIs are labeled: PPA, OPA, and RSC. Centered on each ROI there is a modality shift gradient that runs from visual semantic categories (red) posterior to linguistic semantic categories (blue) anterior.



Identical analysis to Figure 3.2B, but for the other 10 participants. The color of each voxel indicates the representation of body-related information according to the legend at the right. The model weights for vision and language are shown in red and blue, respectively. White borders indicate ROIs found in separate localizer experiments. The relevant body ROI is labeled: EBA. Centered on each ROI there is a modality shift gradient that runs from visual semantic categories (red) posterior to linguistic semantic categories (blue) anterior.



Figure 3.S4. Visual and linguistic representations of face concepts.

Identical analysis to Figure 3.2C, but for the other 10 participants. The color of each voxel indicates the representation of face-related information according to the legend at the right. The model weights for vision and language are shown in red and blue, respectively. White borders indicate ROIs found in separate localizer experiments. The relevant face ROI is labeled: FFA. Centered on each ROI there is a modality shift gradient that runs from visual semantic categories (red) posterior to linguistic semantic categories (blue) anterior.



The thin yellow line indicates the estimated border of the occipital lobe of the brain in each individual participant. This was manually drawn to follow the parieto-occipital sulcus and connect to the preoccipital notch on both ends. The area of the brain which was analyzed in this study was limited to vertices within 50mm of this border, which is shown in black on each individual's brain.



Figure 3.S6. Locations of category-specific modality shifts across cortex for alternate parameter set 1.

Identical analysis to Figure 3.4, but with an ROI size of 10x25mm. Shown here is the flattened cortex around the occipital pole for one typical participant, along with inflated hemispheres. The modality shift metric calculated at each location near the boundary of the occipital lobe is plotted as an arrow. The arrow color represents the magnitude of the shift. The arrow is directed to show the shift from vision to language. Only locations where the modality shift is statistically significant are shown. Areas of fMRI signal dropout are indicated with hash marks. There are strong modality shifts in a clear ring around visual cortex in the same locations seen in Figure 3.4.



Figure 3.S7. Locations of category-specific modality shifts across cortex for alternate parameter set 2.

Identical analysis to Figure 3.4, but with an ROI size of 10x10mm. Shown here is the flattened cortex around the occipital pole for one typical participant, along with inflated hemispheres. The modality shift metric calculated at each location near the boundary of the occipital lobe is plotted as an arrow. The arrow color represents the magnitude of the shift. The arrow is directed to show the shift from vision to language. Only locations where the modality shift is statistically significant are shown. Areas of fMRI signal dropout are indicated with hash marks. There are strong modality shifts in a ring around visual cortex in the same locations seen in Figure 3.4, though the pattern is more noisy due to the shortened analysis windows.

	left RSC	right RSC	left OPA	right OPA	left PPA	right PPA	left EBA	right EBA	left FFA	right FFA
S1	X	X	X	X	X	X	Х	X	Х	X
S2	Х	X	X	X	Х	X	Х	X	Х	X
S3	Х	X	X	Х	X	Х	Х	Х	Х	X
S4	Х	X	Х		X	Х	Х	X	Х	X
S5	Х	X	X	X	X	Х	Х	Х	Х	
S6	Х	X	Х	Х	Х		Х	Х	Х	
S7	Х	X	Х	Х	X	Х	Х	Х	Х	
S8	Х	X	X	X	X	Х	Х	Х	Х	X
S9	Х	X	X	Х	X		Х	Х	Х	
S10	Х	Х	Х	Х	Х	Х	Х	Х	Х	X
S11	Х	Х	Х	Х	Х	Х	Х	Х	Х	X
Total:	11	11	11	10	11	9	11	11	11	7

Table 3.S1. Qualitative observations of modality shifts around category-specific ROIs.

Summary of results presented in Figures 3.1, 3.S2, 3.S3, and 3.S4. For each participant, we judged whether they appeared to have a visual to linguistic shift of a particular semantic category (blue = places, yellow = body parts, red = faces) centered on ROIs which were found with a separate localizer experiment.

CHAPTER 4

Learning and task demands alter semantic representations in the human brain

4.1 Abstract

Prior work on semantic representations have mostly focused on passive perception tasks. That work has shown that semantic information from language is represented similarly in the human brain, regardless of whether it is read or heard, but it is unclear how much these findings translate to settings where there is a more demanding cognitive task. We used fMRI to record brain activity in response to the same set of stimuli under three different task conditions. In this experiment, participants first listened to a set of songs with lyrics, then, after months of memorization, they internally sang the lyrics to those songs, and finally they listened to this set of songs again after this intensive learning process. We then created voxelwise encoding models to characterize the semantic selectivity of each voxel in each participant at each stage of the experiment. By comparing the tuning across the different stages of the experiment, we were able to determine that memorization of the song lyrics resulted in brain responses well before stimuli onset, but that the brain areas with this kind of activity varied depending on task demands. For the internal singing task, it was the inferior temporo-parietal junction (iTPJ), precuneus (PrCu), and subregions of prefrontal cortex (PFC) that showed these anticipatory responses, but for the post-learning listening task, it was superior temporal gyrus (STG), angular gyrus (AG), and other potions of lateral PFC that showed these anticipatory responses. Furthermore, it was this latter set of areas (STG, AG, and lateral PFC) that were most stable in their semantic tuning across the pre- and post-learning stages of the experiment. We analyzed the shifts in semantic tuning across these stages of the experiment and found that there were some shifts that were shared across all participants, which were likely related to the memorization process itself. Other shifts in semantic tuning were not conserved across participants, but we found preliminary evidence that these shifts may have been towards the content of the song lyrics that participants were least familiar with prior to the start of the experiment. Overall, we found that the amodal semantic network for language was much less stable as a result of memorization and task demands when compared with results for passive tasks.

4.2 Introduction

Over the past several years, lexical semantic representations in the human brain have become better understood through the use of naturalistic research paradigms (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). One of these studies showed that semantic information during narrative language comprehension is represented in a rich mosaic of category-specific functional areas located in temporal, parietal, and frontal cortex (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). A follow-up study showed that the semantic selectivity for most of this mosaic is actually the same for both listening and reading (Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). We have also recently collected a small dataset that shows that an overt language production task also produces similar semantic maps of the cortex (unpublished data).

However, all of these studies of lexical semantics have focused on unfamiliar stimuli. To our knowledge, there is a dearth of neuroscience research on the effects of familiarity on linguistic semantic representations. Studies on the familiarity of speech have instead looked at differences in the brain's responses to familiar vs. unfamiliar voices (Bethmann, Scheich, & Brechmann, 2012) or names of loved ones vs. unknown people (Vila et al., 2019), but we are unaware of any studies that contained broader stimulus sets. From work in the visual domain, we know that there can be strong effects on semantic representation and sensitivity when the stimuli are familiar to the participants. This has been shown for specific visual categories that are known to be well-represented in the human brain. Previous work has shown that the parahippocampal place area (PPA), occipital place area (OPA) and the retrosplenial cortex (RSC) respond significantly more to images of familiar scenes than unfamiliar scenes . Furthermore, past work on face representations has shown that responses to familiar or famous faces illicit stronger responses than unfamiliar faces, not only in classical face-responsive areas such as the fusiform face area (FFA), the occipital face area (OFA), and the posterior superior temporal sulcus (pSTS), but also other regions of the cortex such as the posterior cingulate cortex (PC) and medial frontal cortex (Natu & O'Toole, 2011). In addition, we know that cueing participants to attend to specific targets from a variety of semantic categories can strongly influence semantic tuning at the level of individual neurons (David, Hayden, Mazer, & Gallant, 2008), as well as at the voxel level when measured through fMRI (Çukur, Nishimoto, Huth, & Gallant, 2013).

Given these results from visual neuroscience, we wanted to understand how memorization of linguistic stimuli would modify the semantic maps that underpin lexical semantic representations. In order to recover these detailed semantic maps, we needed each participant in our experiment to complete a very complex task, in which they would memorize over an hour of linguistic content. By looking at their semantic representations before and after memorization, we can analyze how these high-dimensional representations shift with increased familiarity with the stimuli.

To address these questions, we used fMRI to record blood-oxygen-level-dependent (BOLD) activity in human participants during a 3-stage experiment that used songs as stimuli. We then used voxelwise modeling (VM) combined with banded ridge regression to characterize the semantic selectivity of each voxel in each individual participant at each stage of the experiment. Next, we interrogated the temporal response profile of each voxel for each individual participant across the different stages of the experiment. This allowed us to understand differences in representations due to learning and also task demands. Finally, we compared the semantic tuning of each voxel for each individual participant at the first and last stages of the experiment. Comparisons of these model weights provided a clear picture of semantic shifts due to learning. This comparison allowed us to identify the network of brain regions that are most altered by learning as well as the directions in the semantic space that they shifted towards.

4.3 Materials and Methods

MRI Data Collection

MRI data were collected on a 3T Siemens TIM Trio scanner at the UC Berkeley Brain

Imaging Center using a 32-channel Siemens volume coil. Functional scans were collected using gradient echo EPI with repetition time (TR) = 2.0045 s, echo time (TE) = 31ms, flip angle = 70°, voxel size = $2.24 \times 2.24 \times 4.1$ mm (slice thickness = 3.5mm with 18% slice gap), matrix size = 100×100 , and field of view = 224×224 mm. Thirty axial slices were prescribed to cover the entire cortex and were scanned in interleaved order. A custom-modified bipolar water excitation radiofrequency (RF) pulse was used to avoid signal from fat. Anatomical data were collected using a T1-weighted multi-echo MP-RAGE sequence on the same 3T scanner.

Participants

Functional data were collected on four participants (two female, one male, one nonbinary), between the ages of 25-28. All participants were healthy and had normal or corrected-tonormal vision.

Data for all three stages of the experiment were collected on participants 1-3. Data for only the internal singing and post-learning listening were collected for participant 4. Data for this participant was considered pilot data and was collected prior to determining that the pre-learning listening data were needed to complete all analyses. Unfortunately, because of the learning structure of this experiment, we could not go back and collect data on this participant for the first stage of the experiment once we realized this need.

All results are shown for participants 1-3, and results are provided for participant 4 when possible.

<u>Stimuli</u>

The versions of the songs with both music and lyrics were downloaded from YouTube as mp3 files. The instrumental midi versions of the songs were obtained from websites where free midi arrangements are available, including:

https://freemidi.org https://miditune.com https://bitmidi.com https://midiworld.com

These midi files were then converted to mp3 format in GarageBand.

The 3 versions of each song (midi music, original song, and parody song) were manually aligned to each other in Audacity. The parody version of the song was left at its normal speed, and the other versions were sped up or slowed down to match. This was done so that the stimulus features for all versions of related songs would be aligned during model estimation and validation.

The list of songs used as stimuli in this experiment can be found in Table 1. All 22 songs were used in all stages of the experiment.

Song Pair	Original Song Title	Original Song Artist	Parody Song Title	Parody Song Artist Weird Al Yankovic	
1	I Want It That Way	The Backstreet Boys	eBay		
2	Mmm Mmm Mmm Mmm	The Crash Test Dummies	Headline News	Weird Al Yankovi	
3	Piano Man	Billy Joel	Ode to a Superhero	Weird Al Yankovi	
4	Radioactive	Imagine Dragons	Inactive	Weird Al Yankovi	
5	Complicated	Avril Lavigne	A Complicated Song	Weird Al Yankovi	
6	Party in the USA	Miley Cyrus	Party in the CIA	Weird Al Yankovi	
7	Waterfalls	TLC	Phony Calls	Weird Al Yankovi	
8	The Eye of the Tiger	Survivor	The Rye or the Kaiser	Weird Al Yankovi	
9	You Belong with Me	Taylor Swift	TMZ	Weird Al Yankovi	
10	American Pie	Don McLean	The Saga Begins	Weird Al Yankov	
11	Like a Virgin	Madonna	Like a Surgeon	Weird Al Yankov	

Table 4.1. Stimuli used in all stages of the experiment.

In experiment stages 1 and 3, song pairs 1-10 were used as model estimation data and song pair 11 was used as model validation data. Model estimation data was repeated once, and model validation data was repeated 4 to 6 times.

In experiment stage 2, song pairs 1-9 were used as model estimation data and song pairs 10-11 were used as model validation data. Model estimation data was repeated twice, and model validation data was repeated 6 times. A greater number of repeats for the model estimation data was used in this experiment stage because pilot data revealed that the overall SNR of this data was lower than listening data. This is likely due to differences in the difficulty of the task. In order to increase the SNR of our models, we could have chosen to either repeat the model estimation data or include additional songs in our stimuli. We determined it would be much less burdensome on our participants to repeat trials of the songs they had already learned, rather than having them learn entirely new sets of songs. Total time spent on learning was not formally tracked for this experiment, but one participant self-reported spending about 35-40 hours learning.

During each fMRI scan, two songs were played. The length of each scan was tailored to the songs, and included 10 s of silence at the beginning of the scan, the end of the scan, and between the two songs. The data for experiment stages 1 and 3 were collected during two 3-h scanning sessions, which were performed on different days. The data for experiment stage 2 were collected during four 1.5-h scanning sessions for participants S1-S3. (Data were collected during two 3-h scanning sessions for participant S4.) During experiment stage 2, there was an

auditory cue prior to each scan that told the participant which two songs they would be internally singing on that particular trial.

Song order was counter-balanced across participants for experiment stage 2. Participants S1 and S2 learned the original versions of the songs first and participant S3 learned the parody versions of the songs first. Each participant learned 5 or 6 songs per session, and the songs in each session were different for each participant. (Participant S4 learned all 11 parody songs for their first session and then all 11 original versions of the songs for their second session.)

Songs were played over Sensimetrics S14 in-ear piezoelectric headphones. A Behringer Ultra-Curve Pro hardware parametric equalizer was used to flatten the frequency response of the headphones based on calibration data provided by Sensimetrics. All stimuli were played at 44.1 kHz using the pygame library in Python. All stimuli were normalized to have peak loudness of -3 dB relative to maximum. However, the songs were not uniformly mastered, so some differences in total loudness remain.

Behavioral Measures

Prior to the experiment, participants rated each of the songs in the experiment on a scale of 1-10 of how familiar they were with each of the songs prior to the experiment. They were told to rate each song according to the following scale: 1 = I have never heard this song before, 4 = familiar/knew some lyrics, 7 = knew most lyrics but not memorized, 10 = I could sing this completely with no karaoke prompts. After the pre-learning listening session, participants rated each song on how well they were able to focus on the "language/narrative structure" of the lyrics (1 = lowest, 10 = highest), as well as how easy they found it to hear all of the words in each song's lyrics (1 = lowest, 10 = highest). After the post-learning listening session, participants again rated each song on how well they were able to focus on the "language/narrative structure" of the lyrics in each song on how well they were able to focus on the "language/narrative structure" of the lyrics (1 = lowest, 10 = highest), as well as how easy they found it to hear all of the words in each song's lyrics (1 = lowest, 10 = highest), as well as how much they enjoyed the song (1 = hate, 10 = love).

Song Transcription and Preprocessing

The Penn Phonetics Lab Forced Aligner (P2FA33) was used to automatically align the audio of each song to a transcript of its lyrics. The forced aligner uses a phonetic hidden Markov model to find the temporal onset and offset of each word and phoneme. The Carnegie Mellon University (CMU) pronouncing dictionary was used to guess the pronunciation of each word. When necessary, words and word fragments that appeared in the transcript but not in the dictionary were manually added. After automatic alignment was complete, Praat34 was used to check and correct each aligned transcript manually.

The aligned transcripts were then converted into separate word and phoneme representations. The phoneme representation of each story is a list of pairs (p,t), where p is a phoneme and t is the time from the beginning of the story to the middle of the phoneme (that is, halfway between the start and end of the phoneme) in seconds. Similarly the word representation of each story is a list of pairs (w,t), where w is a word.

Fitting Encoding Models

Spectral features used to fit the encoding models were extracted from WaveNet (van den Oord et al., 2016), which is an autoencoder used to generate synthetic audio signals. To extract low-level spectral features from our song files, we passed those files through WaveNet to obtain the activations at the middle layer of the autoencoder. This resulted in a 16-dimensional vector for each 32ms segment of audio. These vectors were then concatenated to match the sampling rate of the fMRI data, which resulted in a 512-dimensional spectral feature vector for each 2.0045 second TR.

Semantic features used to fit the encoding models were derived from a 985-dimensional word co-occurrence space. To create this, we first constructed a 10,470-word lexicon from the union of the set of all words appearing in stories used in a previous experiment (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016) and the 10,000 most common words in the large text corpus. We then selected 985 basis words from Wikipedia's List of 1000 Basic Words (contrary to the title, this list contained only 985 unique words at the time it was accessed). This basis set was selected because it consists of common words that span a very broad range of topics. The text corpus used to construct this feature space includes the transcripts of 13 Moth stories, 604 popular books, 2,405,569 Wikipedia pages, and 36,333,459 user comments scraped from reddit.com. In total, the 10,470 words in our lexicon appeared 1,548,774,960 times in this corpus.

Next, we constructed a word co-occurrence matrix, M, with 985 rows and 10,470 columns. Iterating through the text corpus, we added 1 to Mi,j each time word j appeared within 15 words of basis word i. A window size of 15 was selected to be large enough to suppress syntactic effects (e.g. word order) but no larger. Once the word co-occurrence matrix was complete, we log-transformed the counts, replacing Mi,j with log(1+ Mi,j). Next, each row of M was z-scored to correct for differences in basis word frequency, and then each column of M was z-scored to correct for word frequency. Each column of M is now a 985-dimensional semantic vector representing one word in the lexicon.

The matrix used for voxel-wise model estimation was then constructed from the song lyrics: for each word-time pair (w,t) in each song we selected the corresponding column of M, creating a new list of semantic vector-time pairs, (Mw,t). These vectors were then resampled at times corresponding to the fMRI acquisitions using a 3-lobe Lanczos filter with the cut-off frequency set to the Nyquist frequency of the fMRI acquisition (0.249Hz).

Prior to regression, each stimulus feature within each song run was z-scored through time. This was done to match the features to the fMRI responses, which were also z-scored through time for each functional run.

To model our data, we used a modified version of ridge regression called banded ridge (Nunez-Elizalde, Huth, & Gallant; 2019). In this framework, each voxel in the brain is assigned two different ridge parameters: one for the semantic features and one for the low-level features. These pairs of ridge parameters can vary across each voxel, and this allows the model to effectively weight the two sets of features independently across the brain.

A separate linear temporal filter with nine delays (-4, -3, -2, -1, 0, 1, 2, 3, and 4 time points) was fit for each feature. This was accomplished by concatenating feature vectors that had been shifted by -4, -3, -2, -1, 0, 1, 2, 3, and 4 time points (-8, -6, -4, -2, 0, 2, 4, 6, and 8s). Thus, in the concatenated feature space one channel represents the word rate 8s later, another 6s later, and so on, up to the word rate 8s earlier. Taking the dot product of this concatenated feature space with a set of linear weights is functionally equivalent to convolving the original stimulus vectors with linear temporal kernels that have non-zero entries for -4-. -3-, -2-, -1-, 0-, 1-, 2-, 3-, and 4-time-point delays.

As in previous publications (Huth, Nishimoto, Vu, & Gallant, 2012; Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019), data were split into a training and test set, and ridge parameters were selected through cross-validation within the training set. Model performance for both the semantic and low-level sub-models was evaluated by multiplying the fit model weights by the features for the test songs, then taking the correlation coefficient of that predicted time course per voxel with the actual brain data.

Statistical significance of model predictions was computed by comparing estimated correlations to the null distribution of correlations between two independent Gaussian random vectors of the same length. Resulting p-values were corrected for multiple comparisons within each participant using the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995).

All model fitting was performed using custom software, available as a Python package called *tikreg* (https://github.com/gallantlab/tikreg) (Nunez-Elizalde, Huth, & Gallant; 2019).

Functional Localizers

Known regions of interests (ROIs) were localized separately in each participant using standard techniques (Spiridon, Fischl, & Kanwisher, 2006; Hansen, Kay, & Gallant, 2007).

The "speech" portion of a motor localizer task was used to localize most of the languagerelated regions. Motor localizer data were collected during one 11-minute scan. The participant was cued to perform six different motor tasks in a random order in 20-second blocks. For the hand, mouth, foot, speech, and rest blocks the stimulus was simply a word at the center of the screen (e.g., "Hand"). For the saccade block, the participant was shown a pattern of saccade targets.

For the "Hand" cue, the participant was instructed to make small finger drumming movements with both hands for as long as the cue remained on the screen. Similarly for the "Foot" cue the participant was instructed to make small toe movements for the duration of the cue. For the "Mouth" cue, the participant was instructed to make small mouth movements approximating the nonsense syllables *balabalabala* for the duration of the cue—this requires movement of the lips, tongue, and jaw. For the "Speak" cue, the participant was instructed to continuously subvocalize self-generated sentences for the duration of the cue. For the saccade condition the written cue was replaced with a fixed pattern of 12 saccade targets, and the

participant was instructed to make frequent saccades between the targets. A linear model was used to find the change in BOLD response of each voxel in each condition relative to the mean BOLD response.

Broca's area (BA) was defined as the portion of Brodmann areas 44 and 45 that had positive weights for speech production responses in the linear model described above. The superior ventral premotor speech area (sPMv) was defined as the speech-responsive area just anterior to the primary motor area for the mouth. The weight map for speech production responses was also used to functionally define two other speech areas in the inferior frontal gyrus (IFG) and dorsolateral prefrontal cortex (DLPFC).

In addition, localizer data for auditory cortex (AC) were collected in one 10-minute scan. The participant listened to 10 repeats of a 1-minute auditory stimulus, which consisted of 20second segments of music (Arcade Fire), speech (Ira Glass), and natural sound (a babbling brook). To determine whether a voxel was responsive to auditory stimuli, the explainable variance of the voxel response across the 10 stimulus repeats was calculated. The resulting map was used to define AC.

Other localizers were used to define visual and additional motor ROIs, but these functional ROIs are not referenced in this paper. See previous publications (*Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019)* for details on those functional localizers.

Stimulus-Driven Brain Activity

To evaluate which voxels in the brain were responding to the instrumental music in experiment stage 2, we used only the model estimation data. Stimulus-driven brain activity was calculated as the correlation coefficient between the brain activity for original song trials with the brain activity for the parody song trials. This was done for each voxel and each individual participant separately. Statistical significance of these correlations was computed by comparing estimated correlations to the null distribution of correlations between two independent Gaussian random vectors of the same length. Resulting p-values were corrected for multiple comparisons within each participant using the false discovery rate (FDR) procedure.

Principal Components Analysis

Two versions of principal components analysis (PCA) were run in this study. The first one summarized temporal information in the model and the second summarized semantic information in the model. In each case, we first selected only the 10,000 voxels that were best predicted by the semantic model. This was done in each participant separately, for each stage of the experiment. This selection was performed to avoid including noise from poorly modeled voxels.

For temporal PCA, we averaged the semantic weights across all semantic features, leaving only the 9 temporal dimensions. Then, we applied PCA to these weights, yielding 9 principal components. For semantic PCA, we averaged the semantic weights across all temporal

features, leaving only the 985 semantic dimensions. Then, we applied PCA to these weights, yielding 985 principal components.

Cosine Similarity of Model Weights

To evaluate the semantic similarity of model weights across the experiment, a measure of cosine similarity was used. First, the semantic weights for each voxel were averaged across all temporal features, leaving only the 985 semantic dimensions. This was done for the model weights at experiment stage 1 and experiment stage 3 separately. Then, the cosine similarity of the weights for those two time points was calculated for each voxel using the following equation:

similarity
$$(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

Semantic Shifts

To evaluate the shifts in the semantic tuning across the experiment, we again averaged the semantic weights across all temporal features, leaving only the 985 semantic dimensions. Then, we calculated the difference in the semantic model weights per voxel by subtracting the pre-learning listening weights from the post-learning listening weights. To find the voxels with good performance across both of these time points, we took the minimum semantic model performance across time points T1 and T3. We then selected only the 10,000 voxels with the highest performance for each participant. This selection was performed to avoid including noise from poorly modeled voxels, as well as voxels with high performance at one time point but not the other. Finally, we applied PCA to those differences in weights, yielding 985 principal components.

4.4 Results

We sought to determine the degree to which cortical representations of semantic information would change as a result of learning. Three participants took part in a three-stage experiment that used songs and their lyrics as stimuli (Figure 3.1A). (A fourth participant took part in just stages two and three of the experiment. Results for that participant are presented when possible.) To separate brain activity evoked by covert singing from activity elicited by the instrumentation, two versions of each song were presented: one version with the original lyrics, and one version with parody lyrics. At all stages of the experiment, whole-brain BOLD activity were recorded by means of functional MRI.



Figure 4.1. Learning task structure and voxelwise modeling procedure.

(A) This is an illustration of the general task structure. At time point 1 (T1), fMRI data were collected while participants listened to the songs with music and lyrics prior to any learning. Then, they memorized the lyrics to the songs over the course of several weeks. At time point 2 (T2), fMRI data were collected while participants internally sang the lyrics to the songs while listening to the instrumental versions. Finally, at time point 3 (T3), fMRI data were collected while participants listened to the songs with music and lyrics. This experimental structure allowed us to look at differences due to learning (T1 vs. T3) as well as differences due to language modality (T2 vs. T1 and T3).

(B) Participants listened to (or internally sang) 9 pairs of songs while BOLD activity were measured using fMRI. Each word in the song lyrics was projected into a 985-dimensional word embedding space constructed using word co-occurrence statistics from a large corpus of text. Each song was also passed through WaveNet to obtain low-level auditory features. A finite impulse response (FIR) regression model was estimated individually for every voxel. The voxel-wise model weights describe how words appearing in the stories influence BOLD signals. Of particular importance in this experiment is that the music-related features are exactly repeated for the Original and Parody versions of the stimuli. This allowed us to separate brain activity related to music from brain activity related to language content in the song lyrics.

(C) Models were then tested using songs which were not included during model estimation. Model prediction performance was computed as the correlation between predicted responses to these songs and actual BOLD responses.

(D) Shown here is the flattened cortical surface for one typical participant, along with inflated hemispheres. This map shows the correlation of the brain activity for the original versions of all training songs with the brain activity for the parody versions of all training songs for each voxel at T2. High correlation values indicate that the activity of that voxel is driven by the instrumental music that the participant is hearing, since that is what is shared across those two trial types. Voxels with high correlation values are mostly located within auditory cortex (AC), the superior premotor ventral speech area (sPMv), and Broca's Area (BA). The low correlation values across much of parietal, temporal, and frontal cortex suggest that we will be able to model information related to the semantic information in the song lyrics in those brain areas, as has been seen in previous studies (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019).

In the first stage of the experiment (T1), participants listened to a series of 11 songs and their respective parody versions. Over the course of a few months, participants memorized the lyrics to that entire set of 11 songs and their respective parody versions. In the second stage of the experiment (T2), the participants internally sang the lyrics to each song while listening to instrumental-only tracks. These instrumental tracks were identical for the original and parody versions of each song. Thus, the parody songs shared the same instrumental score as the original songs but differed greatly in language content. An evaluation of the separability of music perception from internal singing is shown in Figure 4.1D. Shown there is the correlation coefficient of the brain activity for the original versions of all training songs with the brain

activity for the parody versions of all training songs of each voxel in an example participant. (Data for other participants is shown in Figure 4.S1.) Voxels with high correlation values are mostly located within auditory cortex (AC), the superior premotor ventral speech area (sPMv), and Broca's Area (BA). This indicates that these brain areas are at least partially driven by perception of the instrumental music, but there is still a chance that they are also encoding some information related to the lyrics as well. The low correlation values across much of parietal, temporal, and frontal cortex suggest that these portions of the brain are not strongly driven by music perception. Therefore, we should be able to model information related to the semantic information in the song lyrics in those brain areas, as has been seen in previous studies (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). This is in line with other work, which has shown that the language system does not support music processing (Chen et al., 2021). Finally, in the third stage of the experiment (T3), participants listened to all 11 songs and their respective parody versions again. This was effectively a repetition of the first stage of the experiment (T1). Since the stimuli were identical in T1 and T3, we were able to look at differences in representations due only to learning.

Through the use of VM and banded ridge regression, we can separate the contributions of musical and semantic information and evaluate their contributions to the activity of each voxel in the brain separately. The semantic content of the song lyrics was estimated continuously by projecting each word into a word embedding space based on word co-occurrence statistics (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). A low-level auditory feature space based on activations from WaveNet (van den Oord et al., 2016) was also used to generate additional regressors. These are considered to be nuisance regressors in this particular study because the questions we wished to answer were not related to music or auditory perception. We then used VM to estimate a set of weights for each voxel that best characterize the relationship between the features and the recorded BOLD signals separately for each stage of the experiment (Figure 4.1B). These estimated model weights were then used to predict voxel responses in a held-out validation dataset at each stage of the experiment (Figure 4.1C). This was done separately for the semantic weights and the WaveNet weights in order to disentangle the contributions of each set of features to the activity of each voxel.

Temporal Response Profiles

We first wanted to determine how much planning the participants were doing at each stage of the experiment. We accomplished this by analyzing the temporal response profiles of each voxel across the different stages of the experiment. By inspecting the fit models, we can estimate how strongly each voxel is responding to the semantic content of the song lyrics at any given point in time relative to the stimulus. To summarize the temporal differences across the surface of the cortex, we performed principal components analysis (PCA) on the semantic model weights across all participants and stages of the experiment. The first principal component (PC) explained 27.3 percent of the variance across all participants. Figure 4.2A shows the temporal response profile of the first PC. The first PC shows that many model weights are structured in a pattern similar to that of the canonical hemodynamic response function (HRF), where the BOLD response peaks about 4-6 seconds after stimulus onset (Cohen, 1997; Buxton, Uludağ, Dubowitz, & Liu, 2004). This is shown by the green plot in the right panel. The negative end of this PC

shows the opposite pattern, where instead the peak response of the voxel occurs several seconds prior to stimulus onset. This is shown in the magenta plot in the left panel. These two response profiles can then be used as basis vectors to visualize each voxel's temporal response profile.

We visualized where the model weights for each voxel fall along the continuum between the two ends of this PC. By projecting the fit model weights from each stage of the experiment for one participant onto this PC, we obtain Figures 4.2B-D. These maps reveal that there is a shift in how the participant's brain responds across the stages of the experiment, despite the fact that the stimuli are matched across these time points. In Figure 4.2B, we see that there are many voxels that respond similarly to the canonical HRF and not many that are responding prior to stimulus onset. This is very typical of a perception-based task. However, in Figures 4.2C-D, there are many more voxels that exhibit an anticipatory response to the upcoming stimuli. These are the voxels that are colored in magenta. (Data for other participants is shown in Figure 4.S2.) Furthermore, it is apparent in Figure 4.2E that the network of brain areas that show this anticipatory response are different in the internal singing (T2) and post-learning listening (T3) tasks. Voxels colored in red exhibit preparatory activity at T2, voxels colored in blue exhibit preparatory activity at T3, and white voxels are those in which preparatory activity is occurring in both experimental settings. The apparent lack of white voxels in Figure 4.2E show that while both T2 and T3 engage preparatory activity, each task recruits different brain areas. Finally, this dissociation seems to be relatively consistent across individuals (Figure 4.S3), suggesting that task demands strongly drive which areas of the brain are involved in linguistic planning.



Figure 4.2. Differences in temporal response profiles of individual voxels across stages of the experiment.

(A) Principal components analysis of the voxel-wise semantic model weights reveals a striking temporal pattern in responses. The first principal component (PC) shows that many model weights are structured in a pattern similar to that of the canonical hemodynamic response function (HRF). This is shown by the green plot in the right panel. The negative end of this PC shows the opposite pattern, where instead the peak response of the voxel occurs several seconds prior to stimulus onset. This is shown in the magenta plot in the left panel.

(B-D) Voxel-wise model weights from each stage of the experiment were projected onto the first PC. Projections of model weights onto this PC show temporal responses to stimuli across the cortex at all stages of the experiment. This is shown on the flattened cortical surface of one participant for (B) T1: Pre-learning listening, (C) T2: Internal sining, and (D) T3: Post-learning listening. From these cortical maps, it is clear that there are many more voxels with response patterns similar to the negative end of the first PC at T2 and T3. Those are the experimental settings where preparatory activity is likely to occur.

(E) Voxels with negative projections onto the first PC were extracted from (C) and (D) in order to compare the spatial relationships across time points. Voxels colored in red exhibit preparatory activity at T2, voxels colored in blue exhibit preparatory activity at T3, and white voxels are those in which preparatory activity is occurring in both experimental settings. While both T2 and T3 engage preparatory activity, each task recruits different brain areas.

Conserved semantic representations

Next, we set out to evaluate the semantic information represented in the model weights across the different stages of the experiment. The semantic tuning of each voxel is given by a 985-dimensional vector of weights, one for each of the 985 semantic features. In a similar manner to the previous section, we were able to summarize the semantic representations across the entire cortical surface by performing PCA across all participants. The resulting semantic PCs from the group are ordered by how much variance they explain across the best-predicted voxels. The first three group-level PCs explained 31.9, 12.2, and 6.7 percent of the variance across all participants in the best-predicted voxels, respectively.

In order to visualize large-scale differences in semantic tuning across the stages of the experiment, we mapped the projections of the first three semantic PCs onto each participant's cortical surface. Each voxel was colored according to a RGB color scheme, where the color red represents the first semantic PC, the color green represents the second semantic PC, and blue represents the third semantic PC (Figure 4.3A). Inspection of the RGB maps for the different time points of the experiment reveals that the semantic tuning of many voxels across the surface of the brain are not consistent. This dissimilarity suggests that task demands and learning effects are causing the semantic tuning of the brain to shift in areas such as iTPJ, precuneus, and portions of prefrontal cortex. Figure 4.3B shows the semantic model prediction performance at each stage of the experiment. Comparing these maps with those in Figure 4.3A shows that areas that exhibit these large shifts in semantic tuning are not being found merely in noisily predicted





Figure 4.3. Shifting semantic representations due to learning.

(A) The voxel-wise semantic model weights were projected onto semantic PCs. An RGB colormap was used to color voxels based on the first three dimensions of the shared semantic space. Projections for one participant are shown for each stage of the experiment. Considerable differences can be seen in the semantic tuning across different stages of the experiment.

(B) Prediction performance for the semantic model at each stage of the experiment is shown on the cortical surface for the same participant as in (A). These maps show that some locations with large semantic tuning differences seen in (A) are still significantly predicted by the models.

(C) The cosine similarity of the semantic models weights between T1 and T3 was calculated for each voxel. This is plotted onto the flattened cortical surface for one typical participant, along with inflated hemispheres. Voxels with high values are locations at which the semantic tuning does not shift due to learning, whereas voxels with low values are locations at which the semantic tuning is shifting across the different stages of the experiment. While many areas of the semantic network in temporal, parietal, and frontal cortex have high similarity values, there are also many portions of the semantic network which exhibit large shifts with learning.

One important insight that can be drawn from these maps is where semantic tuning is stable despite learning effects. Shown in Figure 4.3C is the cosine similarity between the semantic model weights at experiment stage T1 and experiment stage T3 for each voxel. In those two experimental stages, the participants are listening to the exact same stimuli. The only difference is that stage T1 occurs before learning and stage T3 occurs after learning. Voxels with high semantic stability across these stages of the experiment are located in many portions of the semantic network in temporal, parietal, and frontal cortex. However, this is not true of all portions of the semantic network, as there are regions within iTPJ, precuneus, and portions of prefrontal cortex that are not semantically stable. (Similar results were found for participants S1 and S3, and results for them are presented in Figures 4.S4 and 4.S5.) We next set out to investigate whether there was a relationship between this semantic stability and the temporal response profiles of individual voxels.

Figure 4.4A combines information about temporal response profiles from Figure 4.2C-D with information about semantic stability from Figure 4.3C. Voxels with high semantic similarity across the experiment are shown in magenta, white, or green, depending on whether they have a negative, zero, or positive projection onto the first temporal PC, respectively. Voxels with low semantic similarity across stages of the experiment are shown with low opacity values. The panel on top shows data for experimental stage T2. This panel illustrates that the areas of the brain with stable semantic tuning across the experiment are not doing preparatory activity during the internal singing task. Conversely, the panel on the bottom is the same data for T3, and it shows that the brain areas with stable semantic tuning across the experiment are doing preparatory activity during post-learning listening. (Cortical flatmaps for participants S1 and S3 are shown in Figure 4.4B. The left panel shows that there is a significant positive correlation between
semantic stability and the temporal response profile of voxels during the internal singing task (r=0.037, p=4.0x10^-16). On the other hand, the right panel shows that there is a negative correlation between semantic stability and the temporal response profile of voxels during the post-learning listening task (r=-0.026, p= $2.9x10^{-7}$). However, this pattern is found within each participant individually only for the internal singing task, but not for the post-learning listening task. This is likely because the internal singing task requires all participants to all do the exact same behavior, whereas the post-learning listening task allows for participants to do somewhat different behaviors. Some participants are likely doing a better job of not thinking ahead in the song during the post-learning listening portion of the experiment. This would explain why we see greater variability in the number of voxels performing anticipatory activity during experiment stage T3 across participants (Figures 4.2B-D and 4.S2).



Figure 4.4. Relationship between temporal response profiles, task demands, and semantic tuning stability.

(A) These flattened cortical surfaces, along with inflated hemispheres, for one participant combine information about temporal response profiles from Figure 4.2C-D with information about semantic stability from Figure 4.3C. Voxels with low semantic similarity across stages of the experiment are shown in black. Voxels with high semantic similarity across the experiment are shown in magenta, white, or green, depending on whether they have a negative, zero, or positive projection onto the first temporal PC, respectively. The panel on the left shows data for T2, which illustrates that the areas with stable semantic tuning across the experiment are not doing preparatory activity during the internal singing task. Conversely, the panel on the left is the same data for T3, which shows that the areas with stable semantic tuning listening.

(B) Voxel data for all participants is shown in scatterplots. Only voxels with statistically significant semantic model prediction performance at the relevant stage of the experiment are included. At T2, there is a significant positive correlation between the semantic stability and the temporal response of the voxel. This means that voxels which have response patterns more like the canonical HRF during the internal singing task are more likely to be semantically stable across the experiment. However, at T3, there is a significant negative correlation between these two variables. That is, voxels which have preparatory activity response patterns during the post-learning listening phase of the experiment are more likely to be semantically stable across the experiment.

Overall, these data further explain the dissociation between voxels doing preparatory activity across different stages of the experiment. That is, we observe that voxels located in STG, AG, and potions of lateral PFC are semantically stable during learning and that they are responding typically to stimuli during internal singing, but perform preparatory activity during the post-learning listening task.

Shifts in semantic representations

In the previous section, we evaluated the response profiles of voxels that had stable semantic tuning across the learning phases of the experiment. In this final section, we analyzed the overall patterns of shifts in semantic tuning. Here, instead of looking at the similarity of the semantic model weights between experimental stages T1 and T3, we subtracted the semantic model weights at T1 from the semantic model weights at T3 for each voxel in each individual participant. This resulted in a 985-dimensional vector per voxel, which can be thought of as the direction of that voxel's tuning shift within the semantic space.

In a similar manner to previous analyses, we were able to summarize the semantic representations across the entire cortical surface by performing PCA. In this case, PCA was performed on the weight differences rather than the weights themselves. This means that the PCs found here are the dimensions within the semantic space that become more strongly or weakly represented in the human brain as a result of learning. The first three group-level PCs explained

13.4, 7.4, and 5.3 percent of the variance across all participants, respectively. Interestingly, the PCs found here were not simply recapitulations of previously found semantic PCs from the weights themselves in either this study or other previous studies of lexical semantic representations (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016) (see Figure 4.S8). If we had found PCs that were correlated with PCs from the weights themselves, this would indicate that the underlying semantic dimensions of the human brain were just being intensified as a result of learning. Finding these new semantic dimensions means that the shifts we have uncovered here are likely specific to this task or set of stimuli.

Similarly to Figure 4.3A, in order to visualize the cortical distribution of these shifts, each voxel was colored according to a RGB color scheme, where the color red represents the first semantic PC, the color green represents the second semantic PC, and blue represents the third semantic PC (Figure 4.5A). Figure 4.5A shows the map of semantic shifts for each individual participant separately. Considerable differences in the direction of the shifts across participants can be seen, as the map for each participant looks quite different when visualized in this manner. This suggests that the process of memorization is warping the semantic representations of each participant somewhat differently. One possibility is that this occurs because each participant having different portions of the stimuli that they attend to more strongly.

However, when we break down Figure 4.5A into maps of projections onto each individual PC, we do see some clear commonalities. Figure 4.5B shows the projections just onto the first PC of the semantic shift space. In this figure, we can see that for each individual, there is a positive shift along this axis for voxels along the STG, ventral TPJ, sPMv, Broca's Area, and portions of PFC. There are also shifts in the negative direction, and these occur in voxels in AG, precuneus, and other sections of lateral PFC. Since these shifts were shared across all participants, it is likely that the shifts along this PC are due to task-specific effects.

Conversely, when we look at projections onto the second and third PCs, we see much more individual variability (Figure 4.5C-D). In these panels, it is hard to pick out locations in the brain that show consistent results across all three participants. There are some shared patterns, such as shared positive projections onto the second PC along STG participants S1 and S3, but in participant S2, the projections along STG are instead negative. This suggests those two PCs are instead capturing individual participant behavior rather than something task- or stimulus-specific.



Figure 4.5. Semantic tuning shifts are partially shared across participants and partially driven by individual differences.

(A) These flattened cortical surfaces, along with inflated hemisphere for S1, show the shifts in semantic encoding from T1 to T3 for each participant individually. Semantic model weights from T1 were subtracted from T3, and then PCA was performed to summarize the distributions of these shifts. An RGB colormap was used to color voxels based on the first three dimensions of the semantic space of these shifts. Considerable differences in the direction of the tuning shifts across participants can be seen, as the map for each participant looks quite different when visualized in this manner.

(B) These flattened cortical surfaces, along with inflated hemisphere for S1, show only the semantic shifts along the first PC. While large individual differences were seen in (A), the overall changes in semantic encoding along this particular axis seem to be relatively shared across participants. The overall shifts in the positive direction for the first PC are consistent across participants and appear along the STG, ventral TPJ, sPMv, Broca's Area, and portions of PFC. It is likely that the shifts along this PC are due to task-specific effects.

(C-D) These flattened cortical surfaces show only the semantic shifts along the second (C) and third (D) PCs, respectively. These PCs highlight some of the large individual differences seen in (A), as it is hard to pick out locations in the brain where results are consistent across all three participants. These PCs may reflect differences in the content that the individual participants attended most strongly to during the learning process.

We further investigated the similarity across participants by projecting this semantic shift data into a shared brain space. We projected the individual participants model weight differences onto a standard freesurfer brain surface, and then projected those weights onto each group PC. The map of semantic shifts along the first PC within this brain space is shown in Figure 4.6. From this, we are able to verify that the first PC is shared across participants. The shifts in the positive direction along this axis for voxels along the STG, ventral TPJ, sPMv, Broca's Area, and portions of PFC. All of these regions fall within the classical language network of the brain. This analysis shows that even these specialized language areas of the brain are strongly, and consistently, influenced by learning and memorization effects.



Figure 4.6. Shared semantic tuning shifts align with the classical language network.

This flattened cortical surface along with inflated hemispheres shows only the average semantic shifts along the first PC on a standard freesurfer brain surface. After averaging across participants, we see that the overall shifts in the positive direction remain along the STG, ventral TPJ, sPMv, Broca's Area, and portions of PFC. The shifts in the negative direction remain around AG, precuneus, and portions of lateral PFC. This illustrates that the similarities observed in Figure 4.5B were common across all participants, and that learning has strongly influenced even the classical language network of the brain.

Finally, we wanted to understand the directions of the semantic shifts, and which semantic domains were the shared across participants and which were individual. We calculated the correlation of each vocabulary word in our word co-occurrence space with the vector for each semantic shift PC. For each PC, we selected the top 100 most correlated words for both the positive and negative end (Table 2). These lists were then used to manually interpret the semantic meaning of each end of each PC. The positive end of the first PC seems to be mostly comprised of "planning" words (such as "determination", "approach", "execution"), whereas the negative end of the first PC is mostly comprised of "body" words (such as "hand", "shoulder", "waist"). This suggests that overall, the tuning of the STG, ventral TPJ, sPMv, Broca's Area, and portions of PFC shifted to represent more information about "active" words in the song lyrics. On the other hand, the tuning of the AG, precuneus, and other sections of lateral PFC have shifted to represent more information about "body parts." This was consistently found across all three participants and likely reflects some similarity in the strategy that the participants used to

memorize the song lyrics, and then carried over to the post-listening learning portion of the experiment as well.

The positive end of the second PC seems to be comprised of "transactional" words (such as "provide", "cost", "funds"), whereas the negative end of the second PC is mostly comprised of "food" words (such as "bowl", "chewed", "juice"). This suggests that participants S1 and S3 were attending more to the "transactional" components of the song lyrics they learned, but that S2 was instead attending more to the "food" portions of the songs. This is reflected in the individual differences in the semantic shifts seen in Figure 4.5C.

The positive end of the third PC is made up of "location" words (such as "america", "europe", "louisiana") and the negative end of the third PC is comprised of "interaction" words (such as "hugging", "crying", "dancing"). This suggests that participant S3 were attending more to the "location" components of the song lyrics they learned, S1 was instead attending more to the "interaction" portions of the songs, and that S2 was not strongly attending to either of these concepts. This is reflected in the individual differences in the semantic shifts seen in Figure 4.5D.

Strongest Negatives	PC	Strongest Positives
hand, compartment, arm, through, padded, shoulder, carefully, floor, hands, inside, wrist, chest, metal, door, clipped, body, sliding, wall, fastened, behind, pocket, desk, pull, room, frame, gloves, use, cut, device, while, onto, down, clamped, into, straps, back, put, could, shaft, yanked, pins, holster, bent, ear, waist, wires, pencil, tube, their, front, to, off, can, used, then, head, forearm, nose, leant, chair, creaked, knotted, using, easily, stretcher, his, window, waistcoat, wire, wound, apron, quickly, cloak, out, pulled, vimes's, it, around, bending, tape, stiffened, heaved, foot, plastic, usually, gingerly, on, tucked, leg, bolts, muzzle, flung, hunched, knife, lever, wrenched, gripped, cord, cylinder, your	1	undertake, determination, search, future, pursuing, nature, critical, uncertainty, pursue, conflict, realm, intelligence, scholar, pursuit, path, notwithstanding, task, civilization, inquiry, knowledge, execution, approach, military, action, probable, seeking, subject, planning, project, conduct, motives, spiritual, mathematical, heroic, skill, interpretation, importance, nonetheless, philosophy, society, cooperation, practical, turmoil, demonstrate, relations, principle, achieve, ambitious, general, philosophical, law, implications, circumstances, nevertheless, succeed, historian, authority, learning, encouragement, plot, guide, extraordinary, supernatural, han, profession, uncertain, justice, newfound, philosopher, conquest, leadership, exceptional, resources, intellect, explore, intention, interest, ambition, skills, resolved, civil, adventure, consequence, mathematics, education, oath, consequences, resolve, tragic, conclusion, persistence, concept, concerning, ideal, circumstance, aspects, historical, teaching, difficulties, abilities
brandy, bowl, half, cups, cup, chewed, juice, tray, spoon, spilled, bottle, milk, bottles, threw, dipped, sandwiches, pint, chocolate, squeezed, salad, cream, cheese, cooked, ham, coffee, lemon, bowls, chips, butter, peel, lips, tea, kitchen, ate, candy, cans, bread, rubbed, steak, licked, soda, jug, chicken, table, breakfast, drink, fried, soup, drinks, swallowed, toast, spitting, spat, ginger, sandwich, bag, potato, screamed, poured, cakes, filling, screaming, napkins, drank, napkin, soaked, chewing, mouth, jar, brushed, whiskey, towel, lip, scrambled, twice, regular, vomit, pouring, poked, vodka, waitress, wrapped, sipping, bags, folded, teeth, barely, coughing, dried, carton, faintly, bacon, dishes, plastic, dish, mug, baked, hot, eaten, throat	2	provide, order, required, provided, purchase, lq, multiethnic, number, humaniform, obtain, cost, available, given, satel, counseled, xw, avout, iru, million, midheaven, temping, saunt, private, allow, funds, mandamus, bjorck, verdurins, exchange, grq, ekstrom, miniaturization, wennerstrom, yhu, additional, information, kdg, rq, klp, public, zdv, circuiting, zd, mantle's, khuh, comporellon, sluttiest, cheyenne's, tartabull, exw, khu, purposes, klv, individual, unexceptional, sayshell, piscean, allowed, chuck-e-cheese, kdw, letterhead, anorexically, morisette, connexion, needed, qati, solarian, janee, cloade, shillings, receive, wkh, value, branno, horribleness, costs, cutoffs, limited, brrrrrrr, permit, rxu, stogy, access, blats, rxw, tartabull's, counselor's, emergecenters, ulthar, abercormbie, appoi, providing, barfield's, khq, classifieds, spangle, existing, vdlg, dqg, narmonov
limp, throat, scream, shaking, numb, screaming, lips, laughing, kiss, chin, kissed, screams, cheek, tears, knees, breathing, asleep, breath, freaked, cheeks, forehead, laughter, neck, chest, kissing, shoulders, hip, sweat, fist, noises, eddie, choking, nose, guitar, crying, sensation, bleeding, awake, lungs, biting, stomach, touching, belly, rubbing, staring, loud, shoulder, dancing, screamed, woke, yelling, tongue, ears, fingers, knee, sore, breathe, punch, grin, shake, singing, dizzy, hits, sweaty, cry, finger, punched, jaw, scar, smiling, arm, legs, crush, pop, touches, touched, elbow, bloody, sucking, yelled, wake, ear, pounding, cried, ankles, hit, grabbed, kisses, smile, moments, rub, rocking, waking, hug, swinging, gentle, leg, spit, hair, kick	3	celebrate, sung, virgin, holiday, america, da, sweden, ole, lovers, paradise, la, singer, arizona, canada, holidays, santa, il, les, mexican, turkey, louisiana, celebrated, eve, sang, shirley, sunday, switzerland, romance, gary, mexico, country, festival, swedish, el, sometime, vermont, texas, california, spain, alabama, timothy, greatest, israel, mama, egypt, africa, cult, ann, florida, easter, born, calendar, thursday, maine, mississippi, joshua, folk, sing, officially, ka, entitled, russia, papa, sunshine, year's, georgia, neil, minnesota, ode, ba, wednesday, sings, del, liberty, lynn, tuesday, jefferson, singing, bon, australia, pa, drama, resident, tonight, le, producer, bye, nation, country's, europe, daughters, sh, dutch, lesbian, enjoyed, jan, aurora, writer, republic, saturday

Table 4.2. Words with the strongest projections onto the first three principal components of the semantic shift space.

For each PC that was found for the shifts in semantic selectivity, we selected the top 100 words that are most similar to each end of the PC. These lists were used to manually interpret the semantic meaning of each PC.

We wanted to better understand these individual differences in semantic tuning shifts, so we attempted to relate them to behavioral measures that were collected at various stages of the experiment (see Methods). For each participant, we calculated the correlation coefficient between each song's average projection onto each semantic shift PC and a given behavioral rating. The tested behavioral ratings were: how well the participant knew each song prior to the experiment, how much the participant enjoyed the song at the end of the experiment, how well they were able to track the narrative content of the song lyrics in the pre-learning listening stage of the experiment, how well they were able to track the narrative content of the song lyrics in the post-learning listening stage of the experiment, as well as the difference and the quotient of the two previous measures. None of these resulted in consistent, statistically significant measures across all participants. However, we only have 22 data points for each of these correlations, so the lack of significant results does not necessarily indicate that these measures are not related to the semantic tuning shifts, we may just be operating in an under-powered regime. In order to illustrate this, we looked at the correlation coefficient of each song's average projection onto each semantic shift PC with how well the participant knew each song prior to the experiment, aggregated across all participants. Here, we see a negative correlation between this behavioral measure and each semantic shift PC (PC1 vs. rating: r=-0.225, p=0.0687; PC2 vs. rating: r=-0.419, p=0.0005; PC3 vs. rating: r=-0.311, p=0.0109). This indicates that the semantic tuning shifts that were found may be inversely related to how well they knew each song prior to the experiment. That is, as a participant learns new semantic content of the songs they did not know well prior to the experiment, their cortical representations shift towards the semantic concepts that are most represented in those songs. It should be noted that after a Bonferroni correction (p<0.0167), the correlation between the first PC and the behavioral rating is not significant, but this actually agrees with what was seen in Figure 4.5. We speculated that the shift in the direction of the first PC (Figure 4.5B) was more likely to be some similarity in the strategy that the participants used to memorize the song lyrics. However, it may be that the individual differences seen in Figures 4.5C-D are related to the individual differences in how well each participant knew each individual song prior to the experiment. Nonetheless, we do consider this to only be preliminary evidence, as we would prefer to evaluate this effect within each individual participant in an experiment with a larger set of stimuli.

These results demonstrate that the semantic network can dramatically shift in its tuning due to learning, attention, and task demands. We believe that the semantic tuning shifts that were consistent across participants were related to the specific strategy required to accomplish this difficult, complex task. In addition, there is preliminary evidence to suggest that the semantic tuning shifts that were inconsistent across participants were instead related to differences in familiarity with stimuli prior to the experiment itself.

4.5 Discussion

In this study, we modeled brain activity evoked by songs and their lyrics over the course of a three-stage experiment to study how semantic representations in the human brain are altered as a result of learning and task demands. We found that a network of brain areas including STG, AG, and portions of lateral PFC are relatively semantically stable during learning. This stability persisted even though voxels in those areas responded well before the stimulus had actually been presented in the post-learning listening stage of the experiment. Conversely, we found that voxels in areas such as iTPJ, precuneus, and portions of prefrontal cortex exhibited the strongest semantic shifts as a result of learning. Voxels in these areas were also those that responded prior to stimulus onset during the internal singing task. Finally, we observed some semantic tuning shifts that were consistent across all participants. These were towards "planning" concepts in STG, ventral TPJ, sPMv, Broca's Area, and other language network portions of PFC, as well as towards "body" concepts in AG, precuneus, and alternate sections of lateral PFC. However, there were also other semantic tuning shifts that were not consistent across participants, and we believe that these may be related to individual differences in pre-experiment familiarity with the stimuli. These semantic tuning shifts, which moved towards more unfamiliar stimulus content, in classical language network areas suggest that even these specialized language regions of the brain are strongly influenced by learning, and therefore, familiarity effects.

From previous research on the visual system (Epstein, Higgins, Jablonski, & Feiler, 2007; Natu & O'Toole, 2011), we expected to see differences in response to familiar vs. unfamiliar stimuli. Our study extended these results to the language system, and showed that brain responses there also change with familiarity. Furthermore, a vast majority of previous work looked only at contrasts of familiar vs. unfamiliar stimuli, rather than the exact shifts in tuning. Through the use of VM, we are able to determine not only where shifts due to familiarity occurred, but also the direction through semantic space of those shifts. Finally, in this study, we had a greater degree of experimental control than many past studies because our familiar and unfamiliar stimuli were perfectly matched. There was one previous study in visual neuroscience with naturalistic stimuli that also had perfectly matched stimuli, and it showed that shifts in semantic representation in vision can be explained by attention (Çukur, Nishimoto, Huth, & Gallant, 2013). However, that study had designated attention targets, and that task structure would not allow for many of the exploratory analyses of semantic shifts that were shown here. Future linguistic studies that interrogate attentional targets at many points throughout the experiment would be able to test more hypothesis-driven analyses like those performed in that study.

Our results also supplement existing research on the nature of anticipatory responses during complex language tasks. Stephens et al. showed that during a two-person conversation task the listener's brain responses precede the speaker's brain responses in dIPFC, mPFC, and the striatum (Stephens, Silbert, & Hasson, 2010). We also observed some anticipatory brain responses in PFC during both our internal singing and post-learning listening tasks, but our results do not fully overlap with theirs. However, this further illustrates how much task demands can influence with regions of the brain are engaged in linguistic predictions. Further study in an array of other linguistic tasks is needed to fully characterize exactly which cognitive demands recruit each brain area to make linguistic predictions.

We envision three main avenues for future study. First, in order to better study the individual differences in semantic shifts, it would be ideal to use a larger set of stimuli. However, there was already a large burden on the participants to memorize all 22 songs in our experiment. (One participant self-reported spending about 35-40 hours learning.) If one wanted to run a similar study with a larger stimulus set, one option would be to not use entire songs. For example, a study could be run that is similar to this, but instead participants would need to learn short segments of 100 songs. In order to maximize the strength of individual differences that could be found, participants or stimuli should be selected so that the variance across participants familiarity ratings of the songs would be maximized. We would also recommend collecting a larger amount of behavioral responses before and after the experiment to be used as covariates of the semantic tuning shifts that are found.

The second direction for future study can be viewed as the opposite of the experimental suggestion in the previous paragraph. Instead of using a larger stimulus set, one could simply ensure that all participants in the experiment had no previous exposure to the chosen stimuli. This experimental design would allow us to disentangle the individual differences in semantic tuning shifts from those that are more general task-related shifts (as in Figure 4.5B). If cortical maps of semantic tuning shifts are then consistent across all participants, it would provide further evidence that the individual differences seen here were related to each participant's pre-experiment level of familiarity with the stimuli. However, if individual differences are still found in that regime, then they may instead reflect differences in how the participants have chosen to perform the task itself.

The final major direction for future study would be to study the effect of learning in language production instead of perception. In this study, the effect of learning could only be evaluated during the listening portion of the task. This is because there was no complement to the internal singing task with spontaneously produced stimuli. This could be accomplished in an overt verbal production task by having participants spontaneously speak out loud while fMRI data was collected, and then later have them complete a similar task to the singing paradigm used here, as long as semantic space coverage across those two tasks were balanced. Accomplishing something similar with covert speech production would be difficult, but not impossible. It would likely require a task that alternated between periods of overt and covert production so that ground truth labels and timing for each word could be obtained for the periods of covert production. Regardless of whether overt or covert speech is the target of future work, extending this research to language production would be incredibly exciting. This could also shed light on the mechanisms through which patients with semantic dementia are sometimes capable of producing previously memorized songs, but not spontaneous speech (Hailstone, Omar, & Warren, 2009; Omar, Hailstone, & Warren, 2012). If these strong effects of familiarity on brain responses exist for speech production as well, it may be that those patient studies are revealing something about memorized language rather than just music. This could be investigated through case studies of patients with semantic dementia who were previously actors, if they had memorized monologues or dialogues earlier in their lives. If cues could trigger them to speak those memorized lines, it would suggest that those earlier studies of song production generalize to all memorized language.

4.6 Supplemental Figures



Figure 4.S1. Stimulus-driven brain activity for additional participants.

Shown here are the flattened cortical surfaces for the additional three participants in this experiment. These maps show the correlation of the brain activity for the original versions of all training songs with the brain activity for the parody versions of all training songs for each voxel at T2, as in Figure 4.1D. High correlation values indicate that the activity of that voxel is driven by the instrumental music that the participant is hearing, since that is what is shared across those two trial types. Voxels with high correlation values are mostly located within auditory cortex (AC), the superior premotor ventral speech area (sPMv), and Broca's Area (BA). The low correlation values across much of parietal, temporal, and frontal cortex suggest that we will be able to model information related to the semantic information in the song lyrics in those brain areas, as has been seen in previous studies (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Deniz, Nunez-Elizalde, Huth, & Gallant, 2019).



Figure 4.S2. Temporal response profiles across all stages of the experiment for additional participants.

Shown here are flattened cortical surfaces for the additional three participants in this experiment. Voxel-wise model weights from each stage of the experiment were projected onto the first temporal PC, as in Figure 4.2B-D. From these cortical maps, we see that there are voxels with response patterns similar to the negative end of the first PC at T2 in all participants, which is the stage of the experiment where preparatory activity is most likely to occur. We see less consistent results across participants at T3, and this likely reflects that some participants are doing a better job of not thinking ahead in the song during the post-learning listening portion of the experiment.



Figure 4.S3. Comparison of preparatory activity for additional participants.

Shown here are the flattened cortical surfaces for the additional three participants in this experiment. Voxels with negative projections onto the first PC during T2 and T3 were extracted from Figure 4.S2 in order to compare the spatial relationships across time points, as in Figure 4.2E. Voxels colored in red exhibit preparatory activity at T2, voxels colored in blue exhibit preparatory activity at T3, and white voxels are those in which preparatory activity is occurring in both experimental settings. While both T2 and T3 engage some amount of preparatory activity, each task recruits different brain areas.



Figure 4.S4. Shifting semantic representations due to learning for participant S1.

(A) Shown here are the voxel-wise semantic model weights projected onto the group semantic PCs. An RGB colormap was used to color voxels based on the first three dimensions of the shared semantic space, as in Figure 4.3A. Considerable differences can be seen in the semantic tuning across different stages of the experiment for this participant as well.

(B) Shown here are voxels that are significantly predicted by the semantic model at each stage of the experiment.

(C) Shown here are the cosine similarity of the semantic models weights between T1 and T3 was calculated for each voxel. There are many portions of the semantic network which exhibit large shifts with learning for this participant as well.



Figure 4.S5. Shifting semantic representations due to learning for participant S3.

(A) Shown here are the voxel-wise semantic model weights projected onto the group semantic PCs. An RGB colormap was used to color voxels based on the first three dimensions of the shared semantic space, as in Figure 4.3A. Considerable differences can be seen in the semantic tuning across different stages of the experiment for this participant as well.

(B) Shown here are voxels that are significantly predicted by the semantic model at each stage of the experiment.

(C) Shown here are the cosine similarity of the semantic models weights between T1 and T3 was calculated for each voxel. There are many portions of the semantic network which exhibit large shifts with learning for this participant as well.



Figure 4.S6. Semantic model weights and model performance for participant S4.

(A) Shown here are the voxel-wise semantic model weights projected onto the group semantic *PCs. An RGB colormap was used to color voxels based on the first three dimensions of the shared semantic space, as in Figure 4.3A. Considerable differences can be seen in the semantic tuning across different stages of the experiment for this participant as well.*

(B) Shown here are voxels that are significantly predicted by the semantic model at stages T2 and T3 of the experiment.



Figure 4.S7. Relationship between temporal response profiles, task demands, and semantic tuning stability for additional participants.

These flattened cortical surfaces, along with inflated hemispheres, for participants S1 and S3 combine information about temporal response profiles from Figure 4.S2 with information about semantic stability from Figures 4.S4 and 4.S5. Voxels with low semantic similarity across stages of the experiment have low alpha values. Voxels with high semantic similarity across the experiment are shown in magenta, white, or green, depending on whether they have a negative, zero, or positive projection onto the first temporal PC, respectively. The top panels show data for T2, which illustrates that the areas with stable semantic tuning across the experiment are not doing preparatory activity during the internal singing task. The panels on the bottom are the same data for T3, which shows that the areas with stable semantic tuning across the experiment are have different temporal response profiles across individual participants. This likely reflects that some participants are doing a better job of not thinking ahead in the song during the post-learning listening portion of the experiment. For example, participant S3 has more voxels which respond similarly to the canonical HRF, suggesting that they were not thinking ahead in the song as much as participant S1, who exhibits both kinds of temporal response profiles.



Figure 4.S8. Correlation values between semantic shift PCs and other semantic PCs.

(A) Shown here are the correlations between the semantic shift PCs and the semantic model weight PCs. There are no strong correlations between these two sets of PCs, which means that these semantic shifts likely specific to this task.

(B) Shown here are the correlations between the semantic shift PCs and semantic model weight PCs from a previous study (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). There are no strong correlations between these two sets of PCs, which means that these semantic shifts are likely specific to to this task, rather than just intensifications of the underlying semantic maps of the human brain.

CHAPTER 5

5.1 Conclusions

The information provided in this dissertation illustrates some of the benefits of using voxelwise modeling (VM) to try to understand representations in the human brain. In Chapter 2, I described some of the holes in existing literature that had been filled by using this modeling technique. In Chapter 3, I showed how detailed comparisons of model weights from VM can uncover a simple, intuitive relationship between two very complex, high-dimensional semantic maps. This analysis method empowered us to provide strong evidence for one side of a decades-long debate in cognitive neuroscience. In Chapter 4, I showed that linguistic semantic maps are stable only in settings where learning does not occur. In this experiment, once the stimuli were committed to memory, we saw dramatic shifts in both the temporal and semantic properties of those cortical semantic maps.

5.2 Future Directions

The results shown in Chapter 3 provide evidence that the visual and linguistic semantic maps present in the human brain are strongly linked. However, it is still largely unknown how semantic representations in other sensory modalities relate to linguistic semantic representations. It would be incredibly exciting to obtain these semantic maps for the auditory and somatosensory systems and then determine if there is also semantic alignment where these unimodal maps abut the linguistic semantic network.

While the study in Chapter 4 gave us a better understanding of how semantic representations shift with learning, it also provided us with many further questions. One of the biggest questions that remains is the role of individual differences in learning. The individual differences in the data presented here are likely due to either individual differences in pre-task familiarity with the stimuli or individual differences in how the participants chose to perform the task generally. However, we cannot determine which of these explanations is more likely from this experiment alone, and follow up studies which control for these variables are still needed.

Finally, it needs to be stated that the research presented in Chapter 4 was actually originally intended to answer entirely different questions. Since previous research had shown that listening and reading produced very similar semantic maps (Deniz, Nunez-Elizalde, Huth, & Gallant, 2019), and unpublished data from the lab suggested that this finding also generalized to overt speech production, I was interested in seeing if we could find this for covert production as well. However, as was seen in Chapter 4, memorization of linguistic content warped the semantic maps a great deal, so we could not actually answer the question of whether natural covert speech production produced similar maps to our other data. For the future, we know that the memorization element of our experiments must be removed if we wish to investigate how semantic representations during covert language production relate to other linguistic modalities. With results from studies like that, we could work towards building better brain-computer interfaces for people who only have the ability to produce language internally.

References

- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An Area within Human Ventral Cortex Sensitive to "Building" Stimuli: Evidence and Implications. *Neuron*, 21(2), 373–383.
- Auerbach, S. H., Allard, T., Naeser, M., Alexander, M. P., & Albert, M. L. (1982). Pure deafness. *Brain*, 105(2), 271–300.
- Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, 47(6), 907–918.
- Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, 22(04), 637–660.
- Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language*, 98(2), 127–139.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 57(1), 289–300.
- Bethmann, A., Scheich, H., & Brechmann, A. (2012). The temporal lobes differentiate between the voices of famous and unknown people: an event-related fMRI study on speaker recognition. *PloS One*, *7*(10), e47626.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767–2796.
- Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., . . . Weaver, K. E. (2011). Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *NeuroImage*, 54(2), 1465–1475.
- Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17(6), 905–917.
- Bird, H., Lambon Ralph, M. A., Patterson, K., & Hodges, J. R. (2000). The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. *Brain and Language*, 73(1), 17–49.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25, 151–188.
- Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, 38(9), 1207–1215.
- Bozeat, S., Ralph, M. A. L., Patterson, K., & Hodges, J. R. (2002). The influence of personal familiarity and context on object use in semantic dementia. *Neurocase*, 8(1–2), 127–134.
- Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, 23(Suppl 1), S220–S233.

- Breedin, S. D., Saffran, E. M., & Branch Coslett, H. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology*, 11(6), 617–660.
- Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, 7(8), 354–361.
- Chan, D., Fox, N. C., Scahill, R. I., Crum, W. R., Whitwell, J. L., Leschziner, G., ... Rossor, M. N. (2001). Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Annals of Neurology*, 49(4), 433–442.
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919.
- Chen, X., Affourtit, J., Ryskin, R., Regev, T. I., Norman-Haignere, S., Jouravlev, O., . . . Fedorenko, E. (2021). The human language system does not support music processing. *bioRxiv*:2021.06.01.446439.
- Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, 6(2), 93–103.
- Çukur, T., Huth, A. G., Nishimoto, S., & Gallant, J. L. (2016). Functional subdomains within scene-selective cortex: Parahippocampal place area, retrosplenial complex, and occipital place area. *Journal of Neuroscience*, 36(40), 10257–10273.
- Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770.
- Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, *1*(1), 123–132.
- Damasio, A. R., & Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. *Large-Scale Neuronal Theories of the Brain*, 6174.
- Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, *380*(6574), 499–505.
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1–2), 179–229.
- David, S. V., Hayden, B. Y., Mazer, J. A., & Gallant, J. L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron*, 59(3), 509–521.
- Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, 15(9), 5870– 5878.
- Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *The Journal of Neuroscience*, 39(39), 7722–7736.
- Desgranges, B., Matuszewski, V., Piolino, P., Chételat, G., Mézenge, F., Landeau, B., ... Eustache, F. (2007). Anatomical and functional alterations in semantic dementia: A voxelbased MRI and PET study. *Neurobiology of Aging*, *28*(12), 1904–1913.

- Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, 33(48), 18906–18916.
- Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., Matthews, P. M., & Tyler, L. K. (2000). Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *NeuroImage*, 11(6 Pt 1), 589–600.
- Diehl, J., Grimmer, T., Drzezga, A., Riemenschneider, M., Förstl, H., & Kurz, A. (2004). Cerebral metabolic patterns at early stages of frontotemporal dementia and semantic dementia. A PET study. *Neurobiology of Aging*, 25(8), 1051–1056.
- Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of Neuroscience*, *33*(4), 1331–1336a.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293(5539), 2470–2473.
- Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology*, 97(5), 3670–3683.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.
- Ercsey-Ravasz, M., Markov, N. T., Lamy, C., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron*, 80(1), 184–197.
- Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *Journal of Neuroscience*, 33(25), 10552–10558.
- Farah, M. J. (2004). Visual agnosia. Cambridge, MA: MIT Press.
- Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia*, 27(4), 461–470.
- Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39), 16428–16433.
- Friedman, L., Glover, G. H., & Fbirn Consortium. (2006). Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *NeuroImage*, 33(2), 471–481.
- Galton, C. J., Patterson, K., Graham, K., Lambon-Ralph, M. A., Williams, G., Antoun, N., ... Hodges, J. R. (2001). Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology*, *57*(2), 216–225.
- Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: an interactive surface visualizer for fMRI. *Frontiers in Neuroinformatics*, 9, 23.
- Garrard, P., & Carroll, E. (2006). Lost in semantic space: A multi-modal, non-verbal assessment of feature knowledge in semantic dementia. *Brain*, *129*(Pt. 5), 1152–1163.
- Garrard, P., Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness,

and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174.

- Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience*, 26(24), 6523–6532.
- Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *Journal of Neuroscience*, 26(18), 4917–4921.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441.
- Hailstone, J. C., Omar, R., & Warren, J. D. (2009). Relatively preserved knowledge of music in semantic dementia. *Journal of Neurology, Neurosurgery, and Psychiatry, 80*(7), 808–809.
- Hansen, K. A., Kay, K. N., & Gallant, J. L. (2007). Topographic organization in and near human visual area V4. *The Journal of Neuroscience*, *27*(44), 11896–11911.
- Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, 37(6), 1027–1041.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.
- Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, *115*(Pt. 6), 1783–1806.
- Hodges, J. R., Patterson, K., Ward, R., Garrard, P., Bak, T., Perry, R., & Gregory, C. (1999). The differentiation of semantic dementia and frontal lobe dementia (temporal and frontal variants of frontotemporal dementia) from early Alzheimer's disease: A comparative neuropsychological study. *Neuropsychology*, 13(1), 31–40.
- Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *NeuroImage*, *36*(3), 924–932.
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6), 1210–1224.
- Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*(Pt. 8), 2132–2147.
- Jefferies, E., Patterson, K., Jones, R. W., Bateman, D., & Lambon Ralph, M. A. (2004). A category-specific advantage for numbers in verbal short-term memory: Evidence from semantic dementia. *Neuropsychologia*, 42(5), 639–660.
- Jefferies, E., Patterson, K., Jones, R. W., & Lambon Ralph, M. A. (2009). Comprehension of

concrete and abstract words in semantic dementia. Neuropsychology, 23(4), 492–499.

- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302–4311.
- Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185), 352–355.
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805–825.
- Kramer, J. H., Jurik, J., Sha, S. J., Rankin, K. P., Rosen, H. J., Johnson, J. K., & Miller, B. L. (2003). Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. *Cognitive and Behavioral Neurology*, 16(4), 211–218.
- Kussmaul, A. (1877). Word deafness and word blindness. Cyclopaedia of the Practice of Medicine. New York, NY: William Wood, 770–778.
- Laisney, M., Giffard, B., Belliard, S., de la Sayette, V., Desgranges, B., & Eustache, F. (2011). When the zebra loses its stripes: Semantic priming in early Alzheimer's disease and semantic dementia. *Cortex*, 47(1), 35–46.
- Lauro-Grotto, R., Piccini, C., & Shallice, T. (1997). Modality-specific operations in semantic dementia. *Cortex*, 33(4), 593–622.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, *4*, 533.
- Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., & Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *Journal of Cognitive Neuroscience*, 23(8), 2079–2101.
- Luzzi, S., Snowden, J. S., Neary, D., Coccia, M., Provinciali, L., & Lambon Ralph, M. A. (2007). Distinct patterns of olfactory impairment in Alzheimer's disease, semantic dementia, frontotemporal dementia, and corticobasal degeneration. *Neuropsychologia*, 45(8), 1823– 1831.
- Lynch, J. C., Mountcastle, V. B., Talbot, W. H., & Yin, T. C. (1977). Parietal lobe mechanisms for directed visual attention. *Journal of Neurophysiology*, 40(2), 362–389.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, 11(2), 194–201.
- Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, 270(5233), 102–105.
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48(7), 788–804.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), 39–41.

- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880), 1191–1195.
- Modha, D. S., & Singh, R. (2010). Network architecture of the long-distance pathways in the macaque brain. Proceedings of the National Academy of Sciences of the United States of America, 107(30), 13485–13490.
- Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology*, 47(1), 36–45.
- Mummery, C. J., Patterson, K., Wise, R. J., Vandenberghe, R., Price, C. J., & Hodges, J. R. (1999). Disrupted temporal lobe connections in semantic dementia. *Brain*, *122*(Pt. 1), 61–73.
- Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., Ito, K., Fukuda, H., Schormann, T., & Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. A PET study. *Brain*, 123(Pt. 9), 1903–1912.
- Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, 56(2), 400–410.
- Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: a review and synopsis. *British Journal of Psychology*, *102*(4), 726–747.
- Nestor, P. J., Fryer, T. D., & Hodges, J. R. (2006). Declarative memory impairments in Alzheimer's disease and semantic dementia. *NeuroImage*, *30*(3), 1010–1020.
- Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21(19), 1641–1646.
- Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, 88(6), 1281–1296.
- Nozari, N., & Thompson-Schill, S. L. (2016). Left ventrolateral prefrontal cortex in processing of words and sentences. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 569–584). San Diego: Academic Press.
- Nunez-Elizalde, A. O., Deniz, F., Gao, J. S., & Gallant, J. L. (2018). Discovering brain representations across multiple feature spaces using brain activity recorded during naturalistic viewing of short films. Presented at the Society for Neuroscience, San Diego, CA.
- Nunez-Elizalde, A. O., Huth, A. G. & Gallant, J. L. (2019). Voxelwise encoding models with non-spherical multivariate normal priors. *Neuroimage*, 197, 482–492.
- Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinstry, R. C., Raichle, M. E., & Conturo, T. E. (1997). Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *NeuroImage*, 6(3), 156–167.
- Omar, R., Hailstone, J. C., & Warren, J. D. (2012). Semantic memory for music in dementia. *Music Perception*, 29(5), 467–477.

- Ono, M., Kubik, S. & Abernathy, C. D. (1990). *Atlas of the Cerebral Sulci*. Thieme Medical Publishers, Inc.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976–987.
- Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the boundary of human visual cortex. *Nature Neuroscience (accepted)*.
- Posner, M. I., Walker, J. A., Friedrich, F. A., & Rafal, R. D. (1987). How do the parietal lobes direct covert attention? *Neuropsychologia*, 25(1A), 135–145.
- Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17(9), 458–470.
- Ralph, M. A. L., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia—what matters? *Neuropsychologia*, 36(8), 775–784.
- Ralph, M. A. L., Graham, K. S., Patterson, K., & Hodges, J. R. (1999). Is a picture worth a thousand words? Evidence from concept definitions by patients with semantic dementia. *Brain and Language*, 70(3), 309–335.
- Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18(1), 42–55.
- Ralph, M. A. L., & Patterson, K. (2008). Generalization and differentiation in semantic memory: Insights from semantic dementia. *Annals of the New York Academy of Sciences*, 1124, 61–76.
- Ralph, M. A. L., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, 107(6), 2717–2722.
- Riddoch, M. J., & Humphreys, G. W. (1987). A case of integrative visual agnosia. *Brain*, *110*(Pt. 6), 1431–1462.
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8), 1261–1269.
- Rosen, H. J., Gorno-Tempini, M. L., Goldman, W. P., Perry, R. J., Schuff, N., Weiner, M., . . . Miller, B. L. (2002). Patterns of brain atrophy in frontotemporal dementia and semantic dementia. *Neurology*, 58(2), 198–208.
- Roskies, A. L., Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (2001). Taskdependent modulation of regions in the left inferior frontal cortex during semantic processing. *Journal of Cognitive Neuroscience*, 13(6), 829–843.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, 19(4), 1835–1842.
- Schwartz, M. F., Marin, O. S. M., & Saffran, E. M. (1979). Dissociations of language function in dementia: A case study. *Brain and Language*, 7(3), 277–306.
- Silveri, M. C., Brita, A. C., Liperoti, R., Piludu, F., & Colosimo, C. (2018). What is semantic in semantic dementia? The decay of knowledge of physical entities but not of verbs, numbers

and body parts. Aphasiology, 32(9), 989-1009.

- Snowden, J. S. (2015). Semantic memory. In Wright, James D., International Encyclopedia of the Social & Behavioral Sciences (pp. 572–578). Elsevier.
- Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology*, 2(3), 167-182.
- Snowden, J. S., Harris, J. M., Thompson, J. C., Kobylecki, C., Jones, M., Richardson, A. M., & Neary, D. (2018). Semantic dementia and the left and right temporal lobes. *Cortex*, 107, 188– 203.
- Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of categoryspecific regions in human extrastriate cortex. *Human Brain Mapping*, 27(1), 77–89.
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America*, 107(32), 14425-14430.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of* the National Academy of Sciences, 94(26), 14792-14797.
- van den Oord, A., Dieleman, S., Zen, H. & Simonyan, K., Vinyals, O., Graves, A., . . . Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv*:1609.03499.
- Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, 255(5043), 419–423.
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, 1(2), 219–247.
- Vila, J., Morato, C., Lucas, I., Guerra, P., Castro-Laguardia, A. M., & Bobes, M. A. (2019). The affective processing of loved familiar faces and names: Integrating fMRI and heart rate. *PloS One*, 14(4), e0216057.
- Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, 22(6), 1083–1094.
- Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*(2), 329–338.
- Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27(4), 635–657.
- Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebral Cortex*, 21(5), 1066–1075.
- Wilkins, A., & Moscovitch, M. (1978). Selective impairment of semantic memory after temporal lobectomy. *Neuropsychologia*, 16(1), 73–79.
- Wu, M. C.-K., David, S. V., & Gallant, J. L. (2006). Complete functional characterization of sensory neurons by system identification. *Annual Review of Neuroscience*, 29, 477–505.