A comparison of imaging modalities and decoding methods for detecting semantic information in the brain

Saskia Lauren Frisby
Gonville & Caius College
University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy

September 2024

Preface

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Chapter 2 has been published as:

Frisby, S. L., Halai, A. D., Cox, C. R., Lambon Ralph, M. A., & Rogers, T. T. (2023). Decoding semantic representations in mind and brain. *Trends in Cognitive Sciences, 27*(3), 258-281.

Chapters 3, 4, and 5 are manuscripts in preparation:

Frisby, S. L., Halai, A. D., Cox, C. R., Clarke, A., Shimotake, A., Kikuchi, T., Kuneida, T., Miyamoto, S., Takahashi, R., Matsumoto, R., Ikeda, A., Rogers, T. T., & Lambon Ralph, M. A. (in prep.). All spectral frequencies of neural activity reveal semantic representation in the human anterior ventral temporal cortex.

Frisby, S. L., Correia, M. M., Zhang, M., Rodgers, C. T., Rogers, T. T., Lambon Ralph, M. A., & Halai, A. D. (in prep.) Optimising 7T-fMRI for imaging the anterior temporal lobe.

Frisby, S. L., Cox, C. R., Halai, A. D., Lambon Ralph, M. A., & Rogers, T. T. (in prep.). Decoding semantics with 7T-fMRI: Convergent evidence and divergent discovery.

Summary

Representation of semantic information enables us to engage with the world in a meaningful way – to comprehend and produce language, identify and use objects, and understand and participate in events that involve us. Our understanding of where in the brain semantic information is represented has progressed much more rapidly than our understanding of *how* semantic information is represented – that is, how activity in the brain enables semantic knowledge to be stored, ready for later deployment. This thesis aimed (1) to develop a theoretical framework with which to describe the nature of neural representations, including semantic representations; (2) to assess and compare the capacities of electrocorticography (ECoG) and 7 tesla functional magnetic resonance imaging (7T-fMRI) to detect semantic representations; and (3) to evaluate the strengths and limitations of multivariate analysis methods, in particular methods based on regularised regression, for revealing properties of semantic representations.

In **Chapter 2** I propose a new theoretical framework for describing representations. By posing six questions about the computational and neural characteristics of the representations that different theorists posit, I situate each contemporary theory relative to the others and bring the inseparable relationship between theory and analysis into focus – different multivariate methods encapsulate different assumptions (not always made explicit) about how the brain represents information. In **Chapter 3** I investigate the temporal dynamics of semantic representations, specifically time-frequency power and phase. By analysing ECoG data recorded from grid electrodes on the ventral temporal cortical surface, I established that semantic information could be decoded from multiple frequency bands. However, only when classifiers were trained on power from all frequencies between 4 and 200 Hz did the "distributed, dynamic" properties observed in voltage data and in a computational model of semantic cognition emerge, suggesting that semantic information is represented in the ventrolateral anterior temporal lobe (vATL) in a "transfrequency" fashion. In **Chapter 4** I lay the foundations for studying semantic representations with 7T-fMRI – I optimised a 7T-fMRI acquisition sequence that improved sensitivity in the vATL while maintaining sensitivity across the rest of the brain. I demonstrated that a multi-echo, multiband sequence achieves these aims. **In Chapter 5** I used the acquisition sequence optimised in Chapter 4, plus four different multivariate decoding methods, to ask why semantic information is so rarely detected in the vATL with fMRI despite the body of evidence for its presence and to ask whether and where semantic information is represented elsewhere in

the brain. Having found evidence for dynamic, graded, multidimensional representations in the vATL, I concluded that my use of a distortion-corrected acquisition sequence and my choice of analysis methods are the most likely reasons for the difference between my findings and previous work. I also found evidence of graded, multidimensional semantic structure in posterior temporal cortex.

To conclude, this thesis (1) developed a unifying theoretical framework in which to situate theories about, and methods for discovering, semantic representations in the brain; (2) established that both ECoG and fMRI can provide insight into the properties of semantic representations; and (3) demonstrated that decoding methods that incorporate neurally-inspired regularisation penalties can be beneficial for decoding, but argued that the best decoding methods for future studies are those that are carefully selected to complement the research question.

To Papa - my first collaborator

Acknowledgements

During my PhD I have been surrounded by a community of wonderful people. They have laughed with me, kept me sane, and moulded me into the scientist that I have become.

First, I thank my supervisors, Matt Lambon Ralph, Ajay Halai, and Tim Rogers (the "three wise men"!), for journeying through these four years with me. Doing challenging, fulfilling, exuberant science with them has been an utter joy. I thank Matt for making sure I am never without a jar of honey, for teaching me diplomacy, for providing stir fry recipes, for getting me into reggae, and for showing me that it is always possible to aspire to a higher standard of empathy, wisdom, and gentleness. I thank Ajay for being there for me countless times when the PhD seemed too much, for all the silly "would you rather" questions, for keeping Orange Club biscuits in his office, and for demonstrating how quiet acts of caring can make an enormous difference – topping up the teapot for everyone at teatime, chatting to the person who is sitting alone, having lunch and laughter with the lab. I thank Tim for teaching me to shuck corn, for referring to semantic representations as "those little guys" (which made them feel infinitely more approachable), for deciding that meetings after 4pm go better with cocktails(!), and for telling me that, even on days when I feel like a speck of dust, I am the speck of dust around which a whole crystal of amazing things is formed.

I am also privileged to have a fantastic collaborative network. I thank Chris Cox for his infinite enthusiasm and generosity with his time and for persuading me to try a deep-fried cheese curd. I thank Marta Correia for her kindness and patience during hours in the control room grappling with acquisition parameters. I thank Tiger Zhang and Chris Rodgers for helping me tame the 7T scanner, Riki Matsumoto and Akihiro Shimotake for sharing their invaluable ECoG data with us (and thus making this PhD possible), Alex Clarke for the endless discussions about how best to preprocess ECoG, and Rik Henson for tolerating all my questions about the theory of decoding. I thank Martyn Blairs for transcribing many hours of participants' speech. I also thank the Wolfson Brain Imaging Centre radiographers, who were beacons of peace and sanity when everything that could go wrong with scanning did go wrong, and the CBU technical staff, especially Gary Chandler, Mark Townsend, and Anthony Parry-Jones, each of whom devoted many hours to the success of the projects described in this thesis.

This research would not have been possible without the 71 patients and participants who selflessly gave up their time to take part. I thank each and every one of them. I also thank the

Medical Research Council and the Stanley Elmore Fund for funding this PhD, the Experimental Psychology Society for funding my study visit to the Wisconsin Institute for Discovery, and Gonville & Caius College for enabling my conference travel.

The MRC Cognition and Brain Sciences Unit has been a marvellous place to work. Those who contribute to its warmth are too many to name, but I thank Shaz Henderson, Rachel Knight, Tim Sandhu, Ashley Zhou, and Alicia Smith for making Office 82 feel like home (Gang Business forever). I also thank Kate Baker, Amy Orben, and all members of the Working Group on Research Culture for their efforts to make the Unit culture even more supportive, and Liz Simmonds for her help on this quest. I thank Siddharth Suresh, Kushin Mukherjee, Ivette Colón and Sean Yun-Shiuan Chuang for making me feel so welcome during my time in Madison.

Finally, I thank those who loved me before the start of this PhD and, through all its ups and downs, continue to do so. I thank Benjamin Weaver for the constant reminders that I still don't know what momentum is, for being the best person to accidentally explore the Hebrides on public transport with, and for only being a phone call away no matter what. I thank Kate, Bruce, Angus and Philip Gentles for being my home away from home, enveloping me in warmth and support whenever I have needed it. I also thank Pandora Gentles for her unfailing optimism and her assistance with communicating this research to a wider audience. I thank Armaiti Bedford for her infectious smiles, for always having something wise to share, and for "just getting it". I thank the Latour Smith family – Keith, Anne-Laure, Merlin and Oscar – for their generosity, spontaneity and *joie de vivre*. I thank Nigel Cooper and Bridget Newns-Cooper for sharing the peace and delight that is to be found both in God and in nature. I thank Dad, Tara, Kira and Fred for all the "strolls" that turned out to be summits, the sing-alongs, the wild swimming – in fact, for every adventure. I thank Papa and Granjo, Nanny and Pops for their constant fascination with my projects and for reminding me to keep marvelling at science. Finally, I thank Maman for being there for me at any hour of the day or night, for making me cry with laughter, and for never letting me doubt that I am truly loved.

'And to love life through labour is to be intimate with life's inmost secret.'

Khalil Gibran, *The Prophet*

Contents

List of Figures

List of Tables

List of Abbreviations

3T-fMRI	3 tesla functional magnetic resonance imaging
7T-fMRI	7 tesla functional magnetic resonance imaging
AFNI	Analysis of Functional NeuroImages
aMTG	Anterior middle temporal gyrus
ANOVA	Analysis of variance
ANTs	Advanced Normalisation Tools
ATL	Anterior temporal lobe
AUC	Area under the curve
BERT	Bidirectional encoder representations from transformers
BIDS	Brain Imaging Data Structure
BOLD	Blood-oxygen-level-dependent
CAT	Computational Anatomy Toolbox
DCNN	Deep convolutional neural network
DVARS	Temporal Derivative of root mean square VARiance over voxelS
ECoG	Electrocorticography
EEG	Electroencephalography
EOG	Electrooculogram
EPI	Echo-planar imaging
FAS	Focal aware seizure
FBTCS	Focal to bilateral tonic-clonic seizure
FCD	Focal cortical dysplasia
fCNR	Functional contrast-to-noise ratio
fMRI	Functional magnetic resonance imaging
FOV	Field of view
FSL	FMRIB (Functional Magnetic Resonance Imaging of the Brain) Software Library
FWHM	Full width at half maximum
GLM	General linear model
GPT-3	Generative pretrained transformer 3

GRAPES	Grounding representations in action, perception, and emotion systems
GRAPPA	GeneRalised Autocalibrating Partial Parallel Acquisition
grOWL	Group-ordered-weighted LASSO (least absolute shrinkage and selection operator)
HA	Hippocampal atrophy
HS	Hippocampal sclerosis
ICA	Independent component analysis
iPAT	Integrated Parallel Acquisition Techniques
IPL	Intraparietal lobule
ITG	Inferior temporal gyrus
LASSO	Least absolute shrinkage and selection operator
LFP	Left frontal pole
LIFGpt	Left inferior temporal gyrus pars triangularis
LmMTG	Left medial middle temporal gyrus
LOC	Lateral occipital complex
LpMTG	Left posterior middle temporal gyrus
LSA	Latent semantic analysis
LTP	Left temporal pole
LvATL	Left ventral anterior temporal lobe
MB	Multiband
MBodd	Odd-numbered volumes of multiband
MDS	Multidimensional scaling
ME	Multi-echo
MEG	Magnetoencephalography
ME-ICA	Multi-echo independent component analysis
MEMB	Multi-echo multiband
MEMBdn	Multi-echo multiband, denoised with multi-echo independent component analysis
MEMBodd	Odd-numbered volumes of multi-echo multiband
MESB	Multi-echo single band
MESBdn	Multi-echo single band, denoised with multi-echo independent component analysis

mITG	Medial inferior temporal gyrus
MNI	Montreal Neurological Institute
MP2RAGE	Magnetisation Prepared 2 Rapid Acquisition Gradient Echoes
MPRAGE	Magnetisation Prepared Rapid Gradient Echo
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MVPA	Multivariate pattern analysis
MVPC	Multivariate pattern classification
NHS	National Health Service
NIFTI	Neuroimaging Informatics Technology Initiative
NNSE	Non-negative sparse embeddings
NSM	Neural similarity matrix
PC	Personal computer
PCA	Principal component analysis
PET	Positron emission tomography
PHG	Parahippocampal gyrus
PIQ	Performance IQ (intelligence quotient)
pMTG	Posterior middle temporal gyrus
PR	Perirhinal cortex
pTx	Parallel transmit
RITG	Right inferior temporal gyrus
ROI	Region of interest
RSA	Representational Similarity Analysis
RSL	Representational Similarity Learning
RSM	Representational similarity matrix
SASICA	SemiAutomatic Selection of Independent Components for Artifact correction in the electroencephalogram
SE	Single-echo
sEEG	Stereoelectroencephalography
SEMB	Single-echo multiband
SEMBodd	Odd-numbered volumes of multi-echo multiband
SESB	Single-echo single band

SMG	Supramarginal Gyrus
SOSLASSO	Sparse-overlapping-sets LASSO (least absolute shrinkage and selection operator)
SPM	Statistical Parametric Mapping
T1w	T1-weighted
TE	Echo time
TIQ	Full-scale IQ (intelligence quotient)
TMS	Transcranial magnetic stimulation
TP	Temporal pole
TR	Repetition time
tSNR	Temporal signal-to-noise ratio
vATL	Ventrolateral anterior temporal lobe
VERSE	Variable-rate selective excitation
VIQ	Verbal IQ (intelligence quotient)
WAB	Western Aphasia Battery
WAIS-III	Wechsler Adult Intelligence Scale (1997)
WAIS-R	Wechsler Adult Intelligence Scale (1991)
WISC MVPA	Whole-brain Imaging with Sparse Correlations MVPA (multivariate pattern analysis)
WMS-R	Wechsler Memory Scale (1987)

Chapter 1

General Introduction

Semantic cognition is the faculty by which we engage with the world in a meaningful way – the faculty that enables us to comprehend and produce language, identify and use objects, and understand and participate in events that involve us. Our understanding of where in the brain semantic information is represented has progressed much more rapidly than our understanding of *how* semantic information is represented – that is, how activity in the brain enables semantic knowledge to be stored, ready for later deployment. Therefore, this thesis aimed (1) to develop a theoretical framework with which to describe the nature of neural representations, including semantic representations; (2) to assess and compare the capacities of electrocorticography (ECoG) and 7 tesla functional magnetic resonance imaging (7T-fMRI) to detect semantic representations; and (3) to evaluate the strengths and limitations of multivariate analysis methods, in particular methods based on regularised regression, for revealing properties of semantic representations.

Chapter 2 of this thesis has already been published and Chapters 3, 4, and 5 are manuscripts in preparation. Within each of those Chapters, the theoretical and empirical background for each study is described in detail and the findings are contextualised with respect to other work in the literature. Accordingly, the purpose of this Introduction is to define key terms, to provide some general background, and to highlight themes that span this body of theoretical and experimental work.

1.1 What is a semantic representation?

Semantic knowledge spans the meaning of objects, word meanings, encyclopaedic facts and people that, unlike episodic memory, is not connected to a particular time or place (Patterson et al., 2007). Semantic representations are the means by which that knowledge is acquired, stored, expressed and used. When I see a dog, I can immediately bring semantic knowledge of the dog to mind – that it is called a "dog", that it barks, that its fur is soft to the touch, and so on. Although I have received input from only one sensory modality (vision), I can recall information pertaining

to other sensory modalities (audition, somatosensation, etc.). Semantic representations are not only *transmodal* but also *transtemporal* – i.e., reflecting the integration of information experienced over time about each concept. As a result, I know that birds fly and lay eggs even though birds never do both of those things simultaneously (Rogers & McClelland, 2004). I can also infer many properties of novel concepts. If told that a new species of cheetah has been discovered that lives at the bottom of the Mariana Trench, I may not know whether it breathes air, but I am certain that it has spots (Rogers, 2024). These examples illustrate an important feature of semantic representations – they enable retrieval and inference across modalities and across temporally separated experiences.

To continue the above example, I infer properties of the underwater cheetah by judging its similarity with concepts of which I have direct experience (e.g. ordinary land-dwelling cheetahs). Spots are a defining feature of land-dwelling cheetahs, so I know that the underwater cheetah will also have spots. Judgements like these are made on the basis of *overall* semantic similarity, not merely similarity within a particular sensory modality. For example, I know, despite the fact that an orange and a basketball look somewhat similar, and despite the fact that "banana" and "basketball" both begin with "ba", that an orange and a banana are more similar to each other than either of them is to a basketball (Devereux et al., 2018). Thus, a second important feature of semantic representations is that they express the overall semantic similarity structure that is necessary for judgements like these.

1.2 Cognitive theories of semantic representation

In cognitive science, the earliest models of semantic representation proposed that concepts (e.g. *dog*, *bird*, *animal*) and features (e.g. *fur*, *feathers*) could be viewed as nodes in a network, connected by labelled links such as *IS A* (as in "a robin is a bird") and *HAS* (as in "a bird has feathers"; Collins & Loftus, 1975; Collins & Quillian, 1969). Features were connected only to the most general category node of which they were usually true – to illustrate, *Labrador* would not be connected to *fur*, but, by virtue of the connections *Labrador IS A dog* and *dog HAS fur*, the network would contain the knowledge that a Labrador has fur. These theories were supported by behavioural evidence – participants were faster to verify that "a robin is a bird" than that "a robin is an animal", purportedly because the nodes *robin* and *animal* were connected via the node *bird* (Collins & Quillian, 1969).

The evidence, however, did not generalise well to other concepts – participants were faster to verify that " a horse is an animal" than that "a horse is a mammal" (Rips et al., 1973), implying that semantic representations were not organised in the rigidly hierarchical way that Collins & Quillian (1969; Collins & Loftus, 1975) suggested. E. E. Smith et al. (1974) therefore proposed a different model, in which categories and individual concepts were represented as lists of their features. If a concept possessed all the features that were true of a category, then the concept belonged to the category.

However, many categories do not have a clear and exhaustive list of defining features (Lambon Ralph et al., 2010; Lambon Ralph & Patterson, 2008). For example, birds usually have hollow bones and fly, while mammals are typically land-dwelling and have four limbs. Penguins (flightless, solid-boned birds) and dolphins (finned mammals) are both characterised by their swimming capabilities. Additionally, similarity structure can be found within-category as well as between-category – for example, dolphins and whales are more similar to each other than either is to a dog. Prototype theories (Rosch, 1975; Rosch et al., 1976; Rosch & Mervis, 1975) acknowledged and could account for both of these facts by proposing that categories were defined in terms of prototypes – abstract representations of the most common and typical properties of a concept. Category membership was graded rather than binary – category members that shared more features with the prototype had a greater "degree of [category] membership" (Rosch, 1975, p. 193), but there need not be any feature that was true of all category members. An alternative perspective, exemplar theories (e.g. Medin & Schaffer, 1978; Nosofsky, 1988), proposed that semantic information is not stored in an abstract way – rather, categorisation of a new object is achieved by comparing it to a number of past instances which are stored (complete with the context in which they were experienced) in memory.

Expressing semantic similarity in terms of shared features, however, poses a further problem – features vary in their importance depending on the concept to which they are applied. To illustrate, an animal that shares most features with a polar bear, but is brown, would be unlikely to be classified as a polar bear; by contrast, an appliance that shares most features with a washing machine, but is brown, would still be considered a washing machine (Lambon Ralph et al., 2010; Murphy & Medin, 1985; Rogers & McClelland, 2004; E. E. Smith & Medin, 2013). To account for this phenomenon, Rogers and colleagues (2004) proposed the hub-and-spoke computational model, developed within the parallel distributed processing framework (Rogers & McClelland, 2004; Rumelhart, McClelland, et al., 1986). This model is composed of units

organised into groups called layers. Interpretable features – names (e.g. *"bird"*), visual features (e.g. *has wings*), functional features (e.g. *can fly*) and encyclopaedic properties (e.g. *migrates*) – are each modelled as one unit within one of the model's visible layers ("the spokes"). Each unit in each visible layer is bidirectionally connected to *every* unit in a further "hidden" layer (the "hub") and these connections each receive a weight (a connection strength). The model gains semantic knowledge through learning. A visual or verbal input is provided to the model by turning on some of the features of that concept; the model must then produce the remaining features. Differences between the target output and the model's actual output are used to adjust all the weights in the model and thereby improve performance (a process called backpropagation; Rumelhart, Hinton, et al., 1986). A model with this architecture is capable of learning the relative importance of different features for different concepts, extracting semantic similarity structures and generalising information – but models which feature no hidden layer, or separate hidden layers for each sensory modalities, are not (Jackson et al., 2021; Rogers & McClelland, 2004). *Re-representation* of semantic information in a multimodal way is necessary for full semantic cognition (Lambon Ralph et al., 2010).

1.3 Semantic representation in the brain

In the 19th century, long before the development of cognitive models of semantic representation, Meynert and Wernicke (Eggert, 1977) proposed a theory of how the brain might represent semantic information. According to this view, semantic information was represented in "engrams" – sources of modality-specific semantic information that arose in secondary association cortex close to where the modality-specific input was originally processed. To retrieve multimodal semantic information about a concept, mass action of this distributed network was required. A key strength of this perspective was that it overcame what later came to be known as the symbol grounding problem (Harnad, 1990). If semantic representations are defined only by their similarity or difference to other semantic representations, there is no way to unambiguously specify the meaning of any of those representations. On the other hand, if semantic information is grounded – tied to something other than other representations of the same kind (Glenberg & Robertson, 2000), in this case to representations of the physical properties of the environment – each semantic representation can acquire an unambiguous meaning. Similar "distributed-only" views (Patterson et al., 2007) are still held today (e.g. A.

Martin, 2007, 2016).

There were two key weaknesses of distributed-only views. The first was that it was not clear how information from different modalities would be, not simply coactivated, but synthesised into a coherent concept. The second was that they predicted that unimodal semantic impairment could be observed following localised brain damage (as indeed it is in the case of associative agnosias; e.g. Riddoch & Humphreys, 2003; Simons & Lambon Ralph, 1999) but that multimodal semantic impairment would be observed only in the case of brain-wide disease. However, Pick (1892, described in Pick, 1898) and Imura (1943, described in Yamadori, 2019), identified cases of multimodal semantic impairment associated with focal neurodegeneration. Warrington (1975) established that this focal neurodegeneration could produce a pattern of multimodal impairment that affected *only* semantic tasks (sparing other cognitive abilities). Warrington's and more recent work demonstrates that, in this disorder, a variety of semantic tasks such as naming, sorting into categories, word-to-picture matching, and drawing from memory are all affected, but tasks requiring episodic memory, syntax, nonverbal reasoning, or spatial skills are unimpaired (Bozeat et al., 2000; Hodges & Patterson, 2007; Patterson et al., 2007; Rogers et al., 2004). This disease was identified as part of the spectrum of frontotemporal dementias and given the name "semantic dementia" by Snowden et al. (1989). Structural magnetic resonance imaging (MRI) and positron emission tomography (PET; Hodges et al., 1992) revealed that the region most affected by atrophy in semantic dementia was the anterior temporal lobe (ATL), bilaterally although usually asymmetrically.

This neuropsychological evidence, combined with the computational modelling discussed earlier, led to the development of the hub-and-spoke theory of semantic cognition (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004). This theory, like distributed-only views, proposes that modality-specific semantic information is represented in modality-specific areas of cortex. The ATL (specifically, its ventrolateral part; Lambon Ralph et al., 2017) functions like the hidden layer in the hub-and-spoke computational model (Rogers et al., 2004) – it synthesises information from multiple modalities and also creates additional transmodal representations of meaning (Jackson et al., 2021; Lambon Ralph et al., 2010; Lambon Ralph & Patterson, 2008; Rogers et al., 2004). Consistent with this proposal, when the model's hidden layer was damaged and it was tested on tasks that are analogous to naming, sorting into categories, word-to-picture matching, and drawing from memory, it produced the same patterns of impairment that are observed in patients with semantic dementia (Rogers et al., 2004).

Additional evidence in support of this "distributed-plus-hub" view (Patterson et al., 2007) came from neuroimaging. PET studies of semantic tasks consistently highlighted a network including the ATL (J. T. Devlin et al., 2000; Perani et al., 1999; Rogers et al., 2006; Vandenberghe et al., 1996; Visser et al., 2010). However, initial functional magnetic resonance imaging (fMRI) studies identified the same network excluding the ATL. There were three reasons for this (Visser et al., 2010). The first was that the ATLs are located close to the air-filled sinuses, which causes signal "dropout" and distortions. The second was that, since the ATLs are at the very base of the brain, studies with a restricted field of view might fail to measure signal in the ATL if the acquisition window was too narrow or the PET camera was not positioned dorsally. The third was that many studies used rest or viewing a fixation cross as a baseline. During this period, participants' minds would wander – which entails the retrieval and processing of semantic information, and so the contrast between task and baseline could be reduced (Binder et al., 1999; G. F. Humphreys et al., 2015; Visser et al., 2010). Studies that employ fMRI sequences designed to recover signal in the ATL (Binney et al., 2010; Embleton et al., 2010; Halai et al., 2014, 2015, 2024), use a large field of view, and select appropriate tasks and baseline tasks do discover activation in the ATL during semantic tasks (e.g. Binney et al., 2010).

A second source of additional evidence for the hub-and-spoke model came from brain stimulation. Transcranial magnetic stimulation (TMS) to the ATL increased participant's reaction times when naming pictures and judging similarity in word meanings but not when reading numbers or when judging similarity in number size (Binney et al., 2010; Lambon Ralph et al., 2009; Pobric et al., 2007). These transient impairments were observed when either ATL was stimulated (but not when control sites were stimulated), reinforcing the claim that the semantic hub is bilateral.

A third source of additional evidence came from human intracranial electrophysiology data, collected from electrocorticography (ECoG) electrodes placed on the ventral temporal cortical surface of patients undergoing surgery (usually for intractable epilepsy). ECoG electrodes can be viewed as a method of neuroimaging combined with a method of brain stimulation. The electrodes can be used to record changes in neural activity (which are observed during a range of semantic tasks; Matoba et al., 2024; Sato et al., 2021; Shimotake et al., 2015) and to deliver tiny electrical currents to inhibit neural activity in a procedure called cortical stimulation mapping (during which performance on a range of semantic tasks is inhibited; Lüders et al., 1991; Matoba et al., 2024; Shimotake et al., 2015). ECoG has some disadvantages –

ECoG data are scarce and, even when available, are from small samples, raising questions about statistical power. Additionally, surface grid electrodes are positioned to record from (or from the vicinity of) diseased tissue. This means that the data that those electrodes produce may be a poor representation of healthy brain activity; it also means that ECoG has very restricted cortical coverage, skewed in favour of regions that are more commonly epileptogenic (the medial temporal lobe, particularly the hippocampus, is disproportionately affected by sclerosis, and temporal, frontal, and insular regions are disproportionately affected by glioma; Duffau, 2014; Kanemoto et al., 1996). However, findings from ECoG converge with findings from neuropsychology, computational modelling, and neuroimaging (Shimotake et al., 2015).

Some have argued against the existence of a unitary semantic hub. For example, neuroimaging studies find activity in posterior medial temporal cortex when participants process tools, but activity in posterior lateral temporal cortex when they process animals (Chao et al., 1999) and in anterior temporal cortex when they process familiar faces compared to other stimuli (H. Damasio et al., 2004). Lesions to these regions produce semantic impairments that affect one category (such as living things) more severely than others (Caramazza & Mahon, 2003; H. Damasio et al., 2004). This has led some (A. R. Damasio, 1989; H. Damasio et al., 2004) to propose the existence of convergence zones for each concept (organised into category-specific convergence regions) which trigger and synchronise information represented in modality-specific cortex. A second strand of evidence concerns the angular gyrus, which is highlighted by imaging studies of semantic processes (Binder & Desai, 2011; A. R. Price et al., 2015) and which, when damaged, produces a pattern of impairment that appears to be predominantly semantic (Geschwind, 1972). This has led others (Binder & Desai, 2011) to propose the existence of multiple semantic hubs, including the ATL and the angular gyrus. However, more recent imaging evidence has demonstrated that the angular gyrus does not have an exclusively semantic function (e.g. Seghier et al., 2010) – for example, it is also recruited heavily by episodic processes (G. F. Humphreys et al., 2021). Additionally, only computational models with a single hub exhibit, when damaged, the full spectrum of impairment observed in patients (L. Chen et al., 2017; Jackson et al., 2021), which challenges both convergence-zone and multi-hub perspectives.

1.4 Multivariate approaches

Most of the neuroimaging evidence discussed so far employed univariate methods, which analyse responses at individual voxels, sensors and electrodes separately. The most common univariate approach is subtraction – activity during a task is compared to activity during a control task which recruits all but the cognitive process of interest – although there exist many more nuanced forms of univariate analysis, such as conjunction analysis (C. J. Price & Friston, 1997), psychophysiological interaction (Friston et al., 1997) and dynamic causal modelling (Friston et al., 2003). An alternative to univariate analysis is multivariate approaches, which analyse patterns of activity across multiple neural features (voxels, sensors, or electrodes), multiple stimulus features (e.g. *fur*, *feathers*) or both to characterise what information the brain is representing. These methods are promising because they can be used to test hypotheses about *how* (not only where) the brain is representing semantic information.

An early multivariate approach to neuroimaging was developed by Haxby and colleagues (2001; see also Haxby, 2012). The analysis was predicated on a simple claim – if a region represents information about category membership, multi-voxel patterns of response to items in the same category should be more similar than responses to items in different categories are. Data were therefore divided in half and correlations were calculated between response patterns from regions of interest (ROIs), both between response patterns to items in the same category and between response patterns to items in different categories. If the within-category correlation exceeded the between-category correlation, this was taken as evidence that the ROI was representing information about category. Subsequent approaches instead trained classifiers to predict category membership from patterns of neural activity (D. D. Cox & Savoy, 2003; Hanson et al., 2004; O'Toole et al., 2005). During training, classifiers receive labelled data (e.g., a pattern of neural responses and the correct category label for the stimulus that the participant was viewing) and learn a coefficient for each neural feature, which can be thought of as its importance for making the distinction of interest. During testing, the classifier is presented with previously unseen neural activity patterns and must generate the correct label; above-chance levels of performance on these held-out data indicate that neural responses encode information about those categories. This collection of approaches is called multivariate pattern analysis (MVPA; Haxby, 2012; Norman et al., 2006). Because the input is neural activity and the output is a feature of the stimulus, MVPA approaches are known as decoding approaches.

Haxby and colleagues (Hanson et al., 2004; Haxby et al., 2001; O'Toole et al., 2005)

observed that the correlation approach did not only highlight differences between categories – it also identified within-category similarity structure. For example, the biggest differences in correlations were observed when distinguishing between animate and inanimate stimuli; smaller differences were observed when distinguishing between houses and smaller inanimate objects (see also Edelman, 1998). This principle was extended and developed by Kriegeskorte and colleagues (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008) into a method called representational similarity analysis (RSA). RSA works by creating neural similarity matrices (NSMs, sometimes defined in terms of dissimilarity instead) which express the similarity in patterns of activity across neural features for each possible pair of stimuli. Similarity matrices are then compared by computing an element-wise correlation. Originally, the correlation was calculated between two NSMs – one created from human brain activity recorded with fMRI, another created from monkey brain activity recorded with electrophysiology techniques – to assess whether representations were comparable across species (Kriegeskorte, Mur, Ruff, et al., 2008). However, it is now more common to compare an NSM to a target representational similarity matrix (RSM), created to express pairwise similarity between stimuli under a hypothesis. Similarity in the target RSM can be modelled in many ways – for example, using feature norms (Dilkina & Lambon Ralph, 2012; McRae et al., 2005; Ruts et al., 2004) or activation patterns in layers of a computational model (Devereux et al., 2018).

At the same time as RSA, a third approach developed with the aim, not simply of classifying observed patterns of neural response, but of being able to predict patterns of neural response to any future stimulus. This is called encoding, or the generative approach (Kay et al., 2008; Mitchell et al., 2008). Unlike decoding approaches, encoding approaches take multiple features of the stimulus as input and it is each stimulus feature, not each neural feature, that receives a learned coefficient. Each encoding model predicts neural activity at a single voxel, sensor or electrode; a family of encoding models (one for each voxel or sensor) can be used to predict whole-brain activity.

MVPA, RSA and encoding have all been applied to the task of revealing semantic representations. For example, both classifiers and RSA have been used to decode semantic information from the ATL hub (Clarke, 2020; C. R. Cox et al., 2024; C. R. Cox & Rogers, 2021; C. B. Martin et al., 2018; Rogers et al., 2021) and from modality-specific spoke regions (C. R. Cox & Rogers, 2021; Devereux et al., 2013, 2018). However, encoding models frequently reveal a highly distributed pattern of results that implicates the entire cortex *except* the ATL in semantic

representation (e.g. Huth et al., 2016). Therefore, some multivariate methods produce results that converge with the body of evidence from neuropsychology, computational modelling, TMS and intracranial electrophysiology, while other methods produce results that diverge.

1.5 Aims, structure and themes of the thesis

1.5.1 Aims of the thesis

The aims of this thesis were:

1. To develop a theoretical framework with which to describe the nature of neural representations, including semantic representations.
2. To assess and compare the capacities of electrocorticography (ECoG) and 7 tesla functional magnetic resonance imaging (7T-fMRI) for detecting semantic representations.
3. To evaluate the strengths and limitations of multivariate analysis methods, in particular methods based on regularised regression, for revealing semantic representations.

1.5.2 Structure of the thesis

In **Chapter 2**, I review theories of semantic representation and propose a new theoretical framework for describing representations. By posing six questions about the computational and neural characteristics of the representations that different theorists posit, I situate each theory relative to the others and thereby make the large range of different approaches tractable. I also provide an overview of popular multivariate methods used to study semantic representation. This review is the first to bring the inseparable relationship between theory and analysis into focus – each method encapsulates assumptions (not always made explicit) about how the brain represents information.

While Chapter 2 considers the spatial distribution of semantic representations, **Chapter 3** investigates their temporal dynamics, specifically time-frequency power and phase. I analysed ECoG data, recorded from grid electrodes on the cortical surface of the ventrolateral ATL (vATL) of patients undergoing surgery for intractable epilepsy. The dataset was large for the field (n = 19) and this was the first time that data from all 19 patients had been analysed as a single sample (Y. Chen et al., 2016; C. R. Cox et al., 2024; Matoba et al., 2024; Rogers et al., 2021;

Shimotake et al., 2015). I also applied a bespoke, clinically-informed preprocessing pipeline to ready the data for decoding (explored further in Appendix C). Using logistic regression with LASSO regularisation (Tibshirani, 1996) – a method that encapsulates very few assumptions about the nature of the neural code – I asked (1) whether semantic information could be decoded from time-frequency power and/or phase, and, if so, from which frequency bands; and (2) whether this code exhibited the same "distributed, dynamic" properties previously observed both in ECoG and in a neural network model of semantic representation (Rogers et al., 2004, 2021). This study set itself apart from others in the field by being both grounded in a tradition of convergent evidence, especially computational modelling (cf. Rogers et al., 2021), while unprejudiced about the frequency bands in which semantic information may be represented (cf. Rupp et al., 2017).

In **Chapter 4** I lay the foundations for studies investigating semantic representations with 7T-fMRI – I optimised a 7T-fMRI acquisition sequence that improved sensitivity in the vATL while maintaining sensitivity across the rest of the brain. I compared the capabilities of parallel transmit, multi-echo and multiband sequences, using a factorial design to disentangle the effects of multi-echo and multiband. This study was distinctive because its sample size was large for the field (n = 20); because a real single-echo sequence was used as a baseline for multi-echo sequences (rather than a dataset derived from multi-echo data, which is an unfair comparison; Halai et al., 2024); and because acquisition sequences were compared on contrast during a semantic task, not simply on temporal signal-to-noise (tSNR), which ensured that the best acquisition sequence would be relevant to future functional imaging studies.

Chapter 5 represents the culmination of the three Chapters that precede it. I asked two questions: (1) why fMRI often fails to discover semantic structure in the vATL, and (2) whether other regions of the brain encode semantic structure similar to that observed in the vATL. This work was set apart from other studies in the field by its data, its analysis methods, and its convergent approach. I acquired 7T-fMRI data, using the novel whole-brain acquisition sequence developed in Chapter 4, while healthy participants named the same pictures that the ECoG patients named. I analysed the data with four different decoding methods, designed to ensure that assumptions about the nature of the neural code were minimised and well-justified. This was only the second time that two of the regularisation penalties have been used with fMRI data (C. R. Cox, 2016; C. R. Cox & Rogers, 2021), and was the first time they have been used with 7T-fMRI. Finally, although studies applying multivariate methods to fMRI data are numerous

(e.g. Connolly et al., 2012; Devereux et al., 2013, 2018; Huth et al., 2016; C. B. Martin et al., 2018; Pereira et al., 2018), this study was the first to ground its data acquisition and analysis strategy in evidence from neuropsychology, brain stimulation, computational modelling, and ECoG.

In **Chapter 6** I review the findings in Chapters 2-5, highlight the themes that permeate them, and propose directions for future research.

1.5.3 Common themes

Chapters 2, 3, 4, and 5 are presented as standalone scientific papers. Each has its own aims, which are presented in the Chapter introduction and considered in the Chapter discussion. The purpose of this section is to highlight themes that span multiple Chapters.

1.5.3.1 How is semantic information represented in the brain?

The capability to discern how the brain *does* represent semantic information depends critically on our ability to articulate hypotheses about how the brain *could* represent semantic information (and then to adjudicate between those hypotheses). The theoretical framework proposed in Chapter 2 breaks this daunting question down into six smaller and more precise questions about the computational characteristics of representations and their instantiation by one or more spatially distinct neural populations. The analytic approach described in Chapter 5 is directly informed by this theoretical work – I selected methods that could reveal semantic representations without pre-empting the answers to any of the six questions. In Chapter 3 I went beyond the scope of the framework in Chapter 2 and tested the temporal dynamics of semantic representations, but, as in Chapter 5, I did so using methods that make minimal assumptions about the computational and spatial properties of those representations. Crucially, this "assumption-light" approach enables the nature of semantic representations to be investigated rather than assumed.

1.5.3.2 Where are semantic representations?

In Chapter 3, I searched for semantic information in a well-charted territory – the vATL, the role of which has been characterised by neuropsychology (e.g. Hodges & Patterson, 2007), computational modelling (e.g. Rogers et al., 2004), noninvasive brain stimulation (e.g. Pobric et al., 2007), and neuroimaging (e.g. Binney et al., 2010; Vandenberghe et al., 1996). In Chapter 5 I employed a whole-brain approach that enables detection of semantic representation, not only

within the well-studied vATL, but elsewhere in the cortex. This is vital because both "distributed-only" and "distributed-plus-hub(s)" models of semantic information predict the representation of semantic information outside the temporal lobe (e.g. Binder & Desai, 2011; A. R. Damasio, 1989; Lambon Ralph et al., 2017); whole-brain imaging enables testing of predictions about semantic representations in these areas. This comparison is made possible by the 7T-fMRI acquisition optimised in Chapter 4, which recovers signal in the vATL while maintaining whole-brain image quality.

1.5.3.3 What sorts of evidence can be brought to bear on the nature of semantic representations?

In this thesis I used two imaging modalities to investigate semantic representations – ECoG and 7T-fMRI. ECoG data, though high in temporal resolution, are limited in availability, are from small sample sizes, have clinical characteristics that limit their generalisability (data are recorded directly from, or from the vicinity of, diseased tissue) and are limited in spatial coverage. fMRI is limited in temporal resolution but shares none of ECoG's shortcomings. In Chapter 5 I compare results from 7T-fMRI to the body of work decoding semantic information from ECoG, including Chapter 3 (see also C. R. Cox et al., 2024; Rogers et al., 2021), thereby assessing the importance of temporal resolution for detecting semantic information.

As illuminated by Chapter 2, multivariate methods encapsulate assumptions about the nature of semantic representations and these assumptions constrain the kinds of representation that each method can detect. In Chapter 5 I compare results from four different decoding approaches. Each method makes very few assumptions about the nature of semantic representations; those assumptions that are made are acknowledged explicitly and are justified thoughtfully (for example, the assumption that semantic representations are located in roughly the same place within and across individuals is justified by the knowledge that gross neuroanatomical structure is the same across individuals; C. R. Cox & Rogers, 2021). This enables me to investigate the impact that each set of assumptions has on our ability to detect semantic information in the brain.

Chapter 2

Decoding semantic representations in mind and brain: a theoretical framework and review

Foreword

In this Chapter I address the first aim of this thesis – to develop a theoretical framework with which to describe the nature of neural representations, including semantic representations.

This Chapter has been published as:

Frisby, S. L., Halai, A. D., Cox, C. R., Lambon Ralph, M. A., & Rogers, T. T. (2023). Decoding semantic representations in mind and brain. *Trends in Cognitive Sciences, 27*(3), 258-281.

I conceptualised of this review and developed it in discussion with my co-authors. I undertook the literature review summarised in Table 2.1 (and described in detail in Appendix A). I wrote the first draft of the manuscript, which was then edited collaboratively.

Highlights

- State-of-the-art brain imaging studies have recently produced a variety of sometimes contradictory conclusions about the neural systems that support human semantic memory.
- Multivariate techniques deployed in this work adopt implicit or explicit assumptions that limit the types of signal they can detect, and thus the types of hypotheses they can test.
- We lay out the space of possible cognitive and neural representations and then critically review contemporary methods to determine which analyses can test which hypotheses.

- The results account for the heterogeneity of recent findings and identify an important empirical and methodological gap that makes it difficult to connect the imaging literature to neurocomputational models of semantic processing.

Abstract

A key goal for cognitive neuroscience is to understand the neurocognitive systems that support semantic memory. Recent multivariate analyses of neuroimaging data have contributed greatly to this effort, but the rapid development of these novel approaches has made it difficult to track the diversity of findings and to understand how and why they sometimes lead to contradictory conclusions. We address this challenge by reviewing cognitive theories of semantic representation and their neural instantiation. We then consider contemporary approaches to neural decoding and assess which types of representation each can possibly detect. The analysis suggests why the results are heterogeneous and identifies crucial links between cognitive theory, data collection, and analysis that can help to better connect neuroimaging to mechanistic theories of semantic cognition.

Keywords: semantic memory, brain imaging, multivariate pattern analysis, concepts, neural decoding, cognitive neuroscience

2.1 The neurocognitive quest for semantic representations

Cognitive science has long sought to understand the mechanisms underlying human semantic memory – the storehouse of knowledge that supports our ability to comprehend and produce language, recognise and classify objects, and understand everyday events. Recently, cross-fertilisation of cognition, neuroscience, and machine learning has generated a plethora of new analysis methods to aid the discovery of neural systems that encode semantic information (C. R. Cox & Rogers, 2021; Kriegeskorte, 2015; Pereira et al., 2018; Popham et al., 2021; Visconti di Oleggio Castello et al., 2021). Although this renaissance has produced a remarkable array of new findings, the evolution of different approaches across research groups makes it difficult to track them all, understand their respective strengths and limitations, and compare results across studies. Consequently, the literature contains sometimes startlingly different conclusions about

the nature, structure, and organisation of semantic representations in the mind and brain, and the field has little recourse for understanding why the differences arise or how they might be reconciled.

We address this challenge by reviewing hypotheses about how semantic information may be encoded computationally and neurally, then critically evaluating the types of representational structure that contemporary multivariate methods can possibly discover in functional neuroimaging data. Crucially, each method encapsulates assumptions about how neural systems encode mental structure that then constrain the types of neural coding it can, and cannot, detect. Hypothesis, data collection, and analysis are therefore linked in ways that sometimes go unremarked and may explain the heterogeneity of findings in the literature. Through exposition of these points, we present an overview of the current empirical landscape with the aim of both organising current thinking about semantic representations in mind and brain, and of providing a more general field guide to contemporary multivariate methods for brain imaging.

2.2 What might semantic representations be like computationally?

Semantic representations serve at least two crucial cognitive functions. First, they express conceptual similarity structure – knowledge that items can be similar in kind even if they are distinct in appearance (e.g hummingbird and ostrich), verbal labels (e.g dog and wolf), or the action plans that engage them (e.g glue and tape). Children as young as 9 months of age detect such relationships and use them to guide reaches even when they contravene perceptual similarity (Mandler, 2006; Pauen, 2002a, 2002b). Adults can reliably judge relatedness in kind and sort items into conceptual groups on this basis (Hodges et al., 1995; López et al., 1997; Rogers et al., 2004), and both children and adults use conceptual similarity as a primary basis for generalising names and other properties (Booth & Waxman, 2008; Lin & Murphy, 2001; Waxman & Markow, 1995). Second, semantic representations support knowledge retrieval or inference – attributing to an item or event properties that are not directly observed or stated. For instance, when observing a picture of a parrot in a textbook, the student may infer that the item can fly even though the image is static; reading about a trip to the restaurant, she may infer that the diner had to pay even if this is not mentioned; observing the new neighbour's pet, a toddler may call it "doggie" even if it is an unfamiliar breed, and so on. Semantic representations thus can be

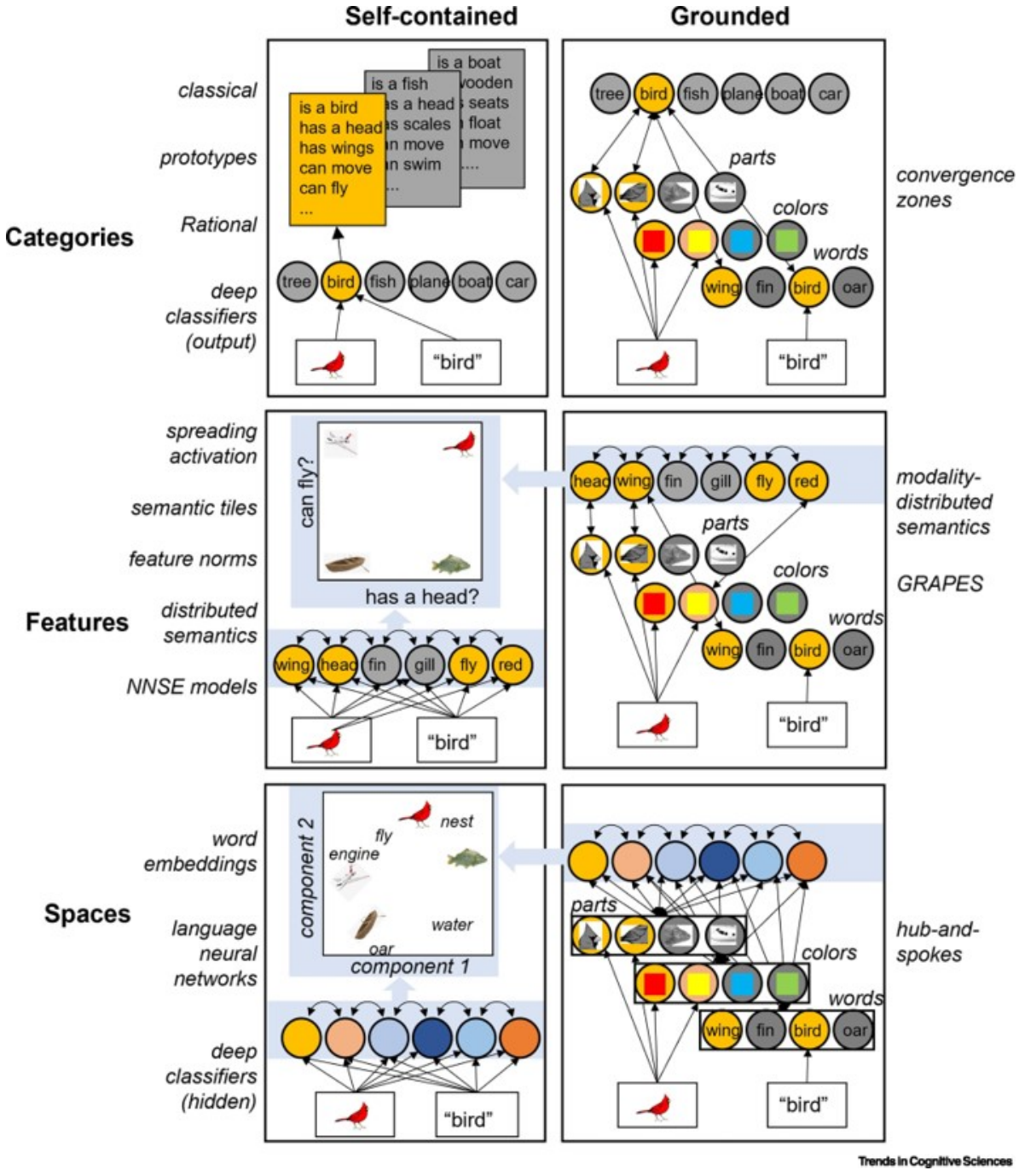
Self-contained
Grounded
Categories
classical
prototypes
Rational
deep classifiers (output)
is a boat
is a fish
is a bird
has a head
has wings
can move
can fly
...
tree
bird
fish
plane
boat
car
"bird"
parts
colors
words
wing
fin
oar
convergence zones
Features
spreading activation
semantic tiles
feature norms
distributed semantics
NNSE models
can fly?
has a head?
head
gill
fly
red
modality-distributed semantics
GRAPES
Spaces
word embeddings
language neural networks
deep classifiers (hidden)
component 2
component 1
nest
engine
water
hub-and-spokes
Trends in Cognitive Sciences

Figure 2.1

Figure 2.1: Computational hypotheses about semantic representation. There are three ways in which conceptual structure could be encoded. First, information may be encoded in discrete, independent category representations (top row). On this view, sensory inputs recruit discrete and independent category representations which either encapsulate semantic information within themselves (J. R. Anderson, 1991; Armstrong et al., 1983; Katz, 1972; Rosch, 1975; Serre et al., 2007; top left) or connect and bind modality-specific surface representations encoding characteristics of category members (A. R. Damasio, 1989; H. Damasio et al., 2004; top right). Second, semantic information may be distributed across independent and interpretable semantic feature representations, with featural overlap indicating conceptual similarity (middle). Features may independently and intrinsically encode the presence of stipulated semantic features within a concept (Cree et al., 1999; Farah & McClelland, 1991; Huth et al., 2016; Tyler et al., 2000; middle left) or gain meaning via connection to surface representations that directly encode such information (Fernandino et al., 2022; A. Martin, 2007, 2016; Popham et al., 2021; middle right). Third, semantic information may be encoded by a continuous distributed representation space that expresses conceptual similarities among items even though its dimensions are not independently interpretable (bottom). Semantic information may be self-contained by the distances encoded in such a space (J. Devlin et al., 2019; Griffiths et al., 2007; Landauer & Dumais, 1997; Mikolov et al., 2013; bottom left) or grounded via mappings from the space to modality-specific surface representations of specific properties (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004; bottom right). Black arrows illustrate how information may flow through the network given the stimuli shown. Text on either side indicates well-known perspectives in the literature that characterise each view. For feature-based and vector space representations, representational spaces are schematised on a blue background. Blue arrows point to the type of representational similarity structure encoded by the corresponding layers – note that both self-contained and grounded approaches can encode the same representational space. Abbreviations: GRAPES, grounding representations in action, perception, and emotion systems; NNSE, non-negative sparse embeddings.

defined as the cognitive and neural states that express conceptual structure and support semantic retrieval/inference. Hypotheses about the cognitive mechanisms that support these functions reside within a fairly constrained space of possibilities (Figure 2.1).

Considering conceptual structure, most approaches adopt one of three positions. The first proposes that semantic memory contains many discrete and independent **category** (see Glossary) representations, each corresponding roughly to a basic-level natural language concept such as *tree* or *boat* (J. R. Anderson, 1991; Rosch et al., 1976; Figure 2.1, top) and possibly to more general (*plant*, *vehicle*) or specific (*elm*, *yacht*) classes (Collins & Quillian, 1969; Jolicoeur et al., 1984). On this view, verbal comprehension involves discerning the category to which a word refers (Xu & Tenenbaum, 2007) whereas comprehension of visual and other sensory inputs involves correctly classifying a perceived item (G. W. Humphreys & Forde, 2001; Jolicoeur et al., 1984; Kriegeskorte, 2015; Serre et al., 2007). Category-based theories explain conceptual structure by proposing that conceptually similar items activate the same category representation

– for instance, parrots, hummingbirds, and robins are viewed as being conceptually related because they all activate the mental category *bird*.

The second view proposes that semantic representations are composed of local **features**, each independently indicating the presence/absence of a property such as *is red*, *can fly*, or *has eyes* (Figure 2.1, middle row). Each perceived item or word activates associated features, indicating properties that are likely to be true of the item (A. J. Anderson, Binder, et al., 2019; Cree et al., 1999; Farah & McClelland, 1991; A. Martin, 2007; Tyler et al., 2000). Conceptual similarity structure arises from property overlap: hummingbirds and ostriches are understood to be similar in kind because they possess many common properties (wings, feathers, etc.), but are also known to be non-identical because they possess individuating properties as well (McRae et al., 1997; Ruts et al., 2004).

Category-based approaches are often distinguished from feature-based views because of the special role that category representations play in determining conceptual similarity and supporting inference. For instance, prototype theories (Mervis & Rosch, 1981), "entry-level" (Jolicoeur et al., 1984; Mack & Palmeri, 2011) and spreading-activation views (Collins & Loftus, 1975), rational approaches (J. R. Anderson, 1991), and some neurally inspired models of object categorisation (Riesenhuber & Poggio, 1999) all propose that access to semantic information depends upon first matching a stimulus (image, word, sound, etc.) to a semantic category. Successful categorisation then provides direct access to semantic information or initiates a "search" of the semantic system, allowing retrieval of other properties. On such views, semantic categories constitute more than merely an additional feature that is attributed to a perceived item.

Nevertheless, under both approaches semantic representations can also be viewed as vectors in a high-dimensional representation space. For categorical theories, dimensions encode membership of distinct and mutually exclusive categories, and the representation of an item is a multinomial probability distribution indicating the probability that a stimulus belongs to each class. For instance, observing an item with wings, feathers, and a beak would generate a high probability density on the *bird* axis and a low density on axes corresponding to *fish*, *car*, *boat*, etc. because the probability that the item is a bird is high and the probability of it belonging to other categories is low. For feature-based theories, dimensions encode various directly interpretable properties, and the representation of an item indicates, independently on each dimension, the binomial probability that the item possesses the corresponding property. On this view, *cardinal* is

a vector with high values on dimensions such as *is red* and *can fly*, but low values on dimensions such as *has scales* and *can swim*. Moreover, some such features may directly indicate the semantic category label of an item (e.g *"bird"*, *"fish"*), although, in contrast to category-based theories, such labels have no special function beyond that of other features. In both cases, conceptual structure reflects the similarity of different points in the **vector space**.

The third proposal likewise views semantic representations as points in a high-dimensional vector space, but without assigning any directly interpretable meaning to the corresponding dimensions (Figure 2.1, bottom). Perception of a stimulus or word evokes an activation pattern across an ensemble of representation units, corresponding to a point in the space where the proximity between points expresses conceptual similarity (Landauer & Dumais, 1997; Mikolov et al., 2013; Pereira et al., 2016). Unlike feature- and category-based approaches, however, one cannot discern what information is encoded in the representation by looking at the activation of each element taken independently. Instead, what matters is the similarity of a given vector to those elicited by other items, taken across all units in the ensemble. On this view, *cardinal* is a vector with high values on some dimensions and low values on others. Examining

Box 1: Ways of estimating semantic structure

Category-based theories propose that distinct representations encode information about different semantic categories. Some have argued that different brain regions are specialised to represent categories that are important for survival over evolution, such as faces, tools, animals, foods, body parts, and shelter (Caramazza & Mahon, 2003; Caramazza & Shelton, 1998; Just et al., 2010; Kanwisher, 2010), but the general question of which categories are stored in memory and why remains controversial (Murphy, 2004; Murphy & Medin, 1985).

Feature-based theories cast semantic representations as vectors that denote the properties of a given item, such as *is red*, *can fly*, or *has blood inside* for the concept *cardinal*. Three methods have been used to construct such vectors.

1. Semantic norming studies ask participants to list the properties that are true of a given concept. Properties generated and/or verified by many participants are compiled in a matrix with rows corresponding to the tested concepts and columns corresponding to the various properties generated by the participants across all study concepts (McRae et al., 2005; Ruts et al., 2004; J. Tanaka and L. Szechter, unpublished

data).

2. Brain-inspired feature vectors identify semantic properties that, from univariate brain imaging, selectively engage different cortical areas. Participants then rate the strength of association between a given concept and each such property. The procedure produces many fewer features than norming studies, but still captures rich conceptual structure (A. J. Anderson, Binder, et al., 2019; Fernandino et al., 2022).

3. Non-negative sparse word embeddings (NNSE) estimate feature vectors from text corpora by exploiting the tendency for words with similar meanings to occur in similar contexts. Standard techniques (e.g latent semantic analysis (LSA; Landauer, 1998; Landauer et al., 1998; Landauer & Dumais, 1997) and word2vec (Mikolov et al., 2013) generate embeddings with uninterpretable dimensions, but, when embeddings are constrained to be both sparse (zeros on most dimensions) and non-negative (only positive values on the rest), the resulting elements are more interpretable and each word can be viewed as a semantic feature vector (Panigrahi et al., 2019).

Vector spaces cast semantic representations as points in a high-dimensional space where pairwise distances capture conceptual relatedness, but with uninterpretable dimensions. Two methods are used to compute such spaces.

1. Unconstrained word embeddings adopt the same corpus-based approach as non-negative sparse embeddings without sparsity or positivity constraints. The resulting spaces express comparable structure to NNSE using fewer dimensions, but the dimensions are not typically independently interpretable.

2. Deep neural networks trained on natural language and/or large image datasets learn vector space representations for photographs, words, or larger units of language. Deep image classifiers represent colour photographs with activation vectors across many serial processing layers (Krizhevsky et al., 2017; Simonyan & Zisserman, 2015); sentence-processing networks represent words, phrases, or whole passages of text as activation vectors over internal units (e.g bidirectional encoder representations from transformers (BERT; J. Devlin et al., 2019) and generative pretrained transformer 3 (GPT3; Floridi & Chiriatti, 2020).

each dimension reveals no information about the properties of the cardinal, but information can be gleaned from the fact that *cardinal* is located very close to *goldfinch*, reasonably close to *ostrich*, and far from *canoe* (Box 1).

Considering retrieval/inference, most approaches adopt one of two proposals, both compatible with the perspectives on conceptual structure outlined above. First, semantic information may be **self-contained** within the representation such that activation brings retrieval/inference along with it (Figure 2.1; left column). For categorical models, the category representation might encapsulate knowledge of properties essential to or characteristic of category members, as in classical, prototype, and rational models (Hampton, 2015; Katz, 1972; Rosch, 1978). In feature-based models, because each element of the representation vector corresponds to an explicit property, the system need only "read off" the vector elements active above some threshold to attribute the corresponding properties to the perceived/named item. Such a view is captured by semantic feature-based neural network models (Cree et al., 1999; Farah & McClelland, 1991; Tyler et al., 2000), spreading-activation models (Collins & Loftus, 1975; Kumar et al., 2022; Rotaru et al., 2018), and distributional semantic models that constrain representations to have interpretable dimensions (such as topic models and non-negative sparse embeddings; Box 1; Derby et al., 2018; Griffiths et al., 2007). For vector space models, although the dimensions of the representation space are not independently interpretable, retrieval/inference can still be self-contained by proposing that these functions rely on similarity and/or direction within the representation space (Mikolov et al., 2013). For instance, the system may infer that the cardinal can fly and breathe because the vectors for the words "fly" and "breathe" are both near to the vector for "cardinal" and are situated along a direction in the space that separates behavioural "can" properties from other property types (such as parts, names, colours, etc.). Such a perspective is captured by distributional semantic models that are not constrained to yield interpretable dimensions (e.g latent semantic analysis (Landauer & Dumais, 1997), hyperspace analogue to language (Burgess & Lund, 1997), word2vec (Mikolov et al., 2013), and language neural networks (J. Devlin et al., 2019); Box 1).

Self-contained approaches face a significant hurdle, however: retrieving the content of a representation requires a labelling scheme, without which it would be impossible to know which semantic content "goes with" which representation vectors (sometimes called the symbol grounding problem; Barsalou, 2008). The second approach to retrieval/inference (Figure 2.1, right column) addresses this problem by proposing that semantic content is **grounded** in

perception, action, and language systems that directly encode **surface representations** of the environment: shapes, colours, parts, movements, affordances, words, and so on (Barsalou, 2003; Glenberg, 2010; Glenberg & Robertson, 2000). On this view, the activation of a categorical, feature-based, or vector space representation does not in itself cause information retrieval/inference. Instead, retrieval/inference arises when these structure-encoding representations activate modality-specific representations that are identical or intimately related to those that directly mediate perception and action. Thus the categorical/featural/vector space representation of *canoe* is meaningful only in virtue of its ability to generate mental images of what a canoe looks like (including shape, colour, parts, etc.), motor actions associated with canoes (e.g paddling), words used to describe canoes ("boat", "light", "floats"), and so on.

On a grounded category-based approach, a discrete category representation connects the surface representations encoding characteristics of category members, and binds these together so that they are understood as all inhering in the same concept. For example, *bird* connects surface representations of the visual appearance of feathers, the motion of flight, the word "bird", and so on; the "convergence zone" hypothesis provides an example of this view (A. R. Damasio, 1989; H. Damasio et al., 2004). Under grounded feature-based approaches, the featural dimensions that encode the semantic representation are "labelled" by virtue of their direct/preferential connectivity to surface representations that directly encode the corresponding content – for instance, a semantic dimension encoding the colour of an object may be directly connected to colour-perception areas; a dimension encoding its associated action may be connected to motion-perception areas; and so on. Several proposals motivated by functional imaging data align with this view, including the GRAPES (grounding representations in action, perception, and emotion systems) framework (A. Martin, 2016) and the neurally inspired "experiential features" view (A. J. Anderson, Binder, et al., 2019; Fernandino et al., 2022). Finally, grounded vector-space models suggest that the representational ensemble that encodes conceptual similarity structure connects reciprocally to a variety of different surface representations such that the generation of an activity pattern across the ensemble activates surface representations that encode the specific, embodied properties associated with the corresponding item – a view consistent with the hub-and-spoke model of semantic representation (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004; Rogers & McClelland, 2004).

In sum, considering how semantic representations might serve their defining functions –

expressing conceptual structure and supporting semantic retrieval/inference – delineates a well-constrained space of hypotheses in which cognitive theories of semantic representation can be situated. The different views, and examples of theories aligning with each, are shown in Figure 2.1. Each cognitive hypothesis has implications for how neural data are best collected and analysed; for instance, adjudicating grounded versus self-contained theories may require participants to semantically process stimuli in different modalities. The next section considers how these views constrain the search for neural systems that encode semantic information.

2.3 How might semantic representations be organised in the brain?

Next, we consider how these different computational schemes might be implemented in neural systems in ways that can be measured by functional brain imaging. All such technologies can be viewed as summarising the responses of many different neural populations to a cognitive event. Different technologies such as **functional magnetic resonance imaging (fMRI)**, **electroencephalography (EEG)**, **magnetoencephalography (MEG)**, and **electrocorticography (ECoG)** yield summary estimates at different spatial and temporal granularities (e.g voxels, EEG sources, and electrodes). We will use the term "unit" to refer to the summary estimate provided by a given technology over its characteristic window of space and time. Therefore, regardless of imaging modality, the neural response to a stimulus is characterised as a pattern of activation across many units over a particular window of time. Discovering the neural underpinnings of semantic representations then requires close consideration of (1) how the representational elements proposed by a cognitive theory are encoded in unit activation patterns within and across individuals, (2) how the representational work might be divided among units participating in a representation, and (3) how signal-carrying units might be anatomically organised within and across individuals.

2.3.1 Variation of the neural code

Within an individual, the neuro-semantic code – how changes in unit activity express semantic information – can be either **homogeneous** or **heterogeneous** (Figure 2.2A). In a homogeneous code, signal-carrying units all adopt the same activation when the represented information is present – for instance, all voxels representing *cat* become more active when a cat is semantically

processed. In a heterogeneous code, different units express the same information differently – some voxels representing *cat* may be greatly activated when a cat is present, some greatly suppressed, and some only moderately active, etc. Approaches that average unit activations within participants (e.g via spatial smoothing or **region of interest (ROI)** averaging) favour the discovery of homogeneous over heterogeneous codes.

Across individuals, the neural code may be **consistent** – a given piece of information is always expressed with the same activity change in homologous units (e.g *cat* always being signalled by the same activation pattern across aligned voxels of different individuals) – or **inconsistent** (*cat* being signalled by different activation patterns across aligned voxels of different individuals; Figure 2.2A). Methods that aggregate or summarise unit activation across individuals – for instance, fitting a single model to decode all participants, computing the mean blood-oxygen-level-dependent (BOLD) response at each voxel before applying a **decoding** model, or averaging predictions of **encoding models** across participants before passing the result to further analysis – favour the discovery of consistent over inconsistent codes. Likewise, methods that align voxels across individuals on the basis of their having similar activation patterns across stimuli (e.g hyper-alignment; Guntupalli et al., 2016) implicitly assume a consistent code.

2.3.2 Independent and conjoint codes

Categorical and feature-based approaches both suggest that each unit **independently** encodes a piece of semantic information: its activity expresses the presence or absence of that information (such as category membership or a semantic feature) regardless of the states of other units. For the example shown in the left panel of Figure 2.2B, unit 1 encodes whether the stimulus is living or non-living independently of unit 2, whereas unit 2 encodes whether the stimulus can fly independently of unit 1. For any stimulus, it is possible to determine whether the item is alive solely by inspecting the state of unit 1, without needing to consider the activation of other units.

By contrast, vector space hypotheses suggest that units **conjointly** encode a representational space, and that semantic information is expressed in the activity pattern considered across multiple units such that single-unit activation may not be interpretable without consideration of other units in the ensemble. Figure 2.2B shows two examples. In the middle panel, one cannot determine whether a stimulus is living or whether it can fly solely by inspecting the activation of unit 1 (because *fish* and *plane* elicit equal activation) or unit 2 (because *boat* and *cardinal* elicit equal activation). Considering the joint activation of both units

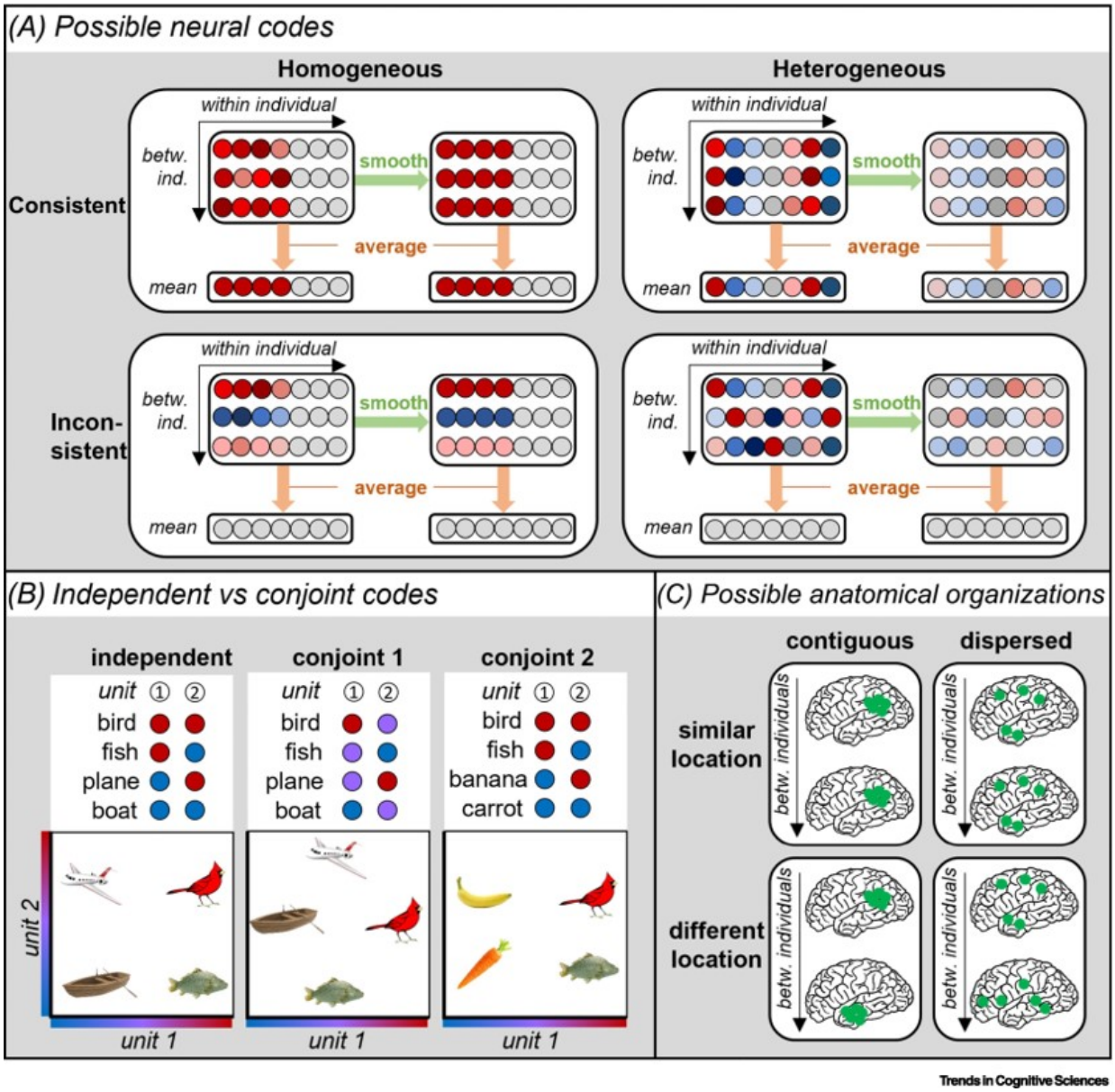

Figure 2.2: Hypotheses about the neuro-semantic code. (A) Within individuals a representation may adopt a homogeneous code (all involved units adopt the same activation change – i.e., all become more active or all become less active) or a heterogeneous code (the units involved adopt different changes to activation – i.e., some become more active than others, and/or some become more active and some less active). Across individuals the code may be consistent (the same magnitude and direction of change in all individuals) or inconsistent (different magnitudes and/or directions of change in different individuals). Spatial smoothing and cross-subject averaging can either help or hinder discovery depending on the code. (B) In the independent code shown, unit 1 activation indicates whether the item is animate, while unit 2 independently encodes whether it can fly. In the first conjoint code, the two units express the same similarity relations among the four items, but considered independently, neither unit clearly expresses either dimension. For instance, fish and plane both moderately activate unit 1, whereas bird and boat moderately activate unit 2. In the second conjoint example, unit 2 activation is difficult to interpret considered independently, but discriminates birds from fish when unit 1 is active, and fruits from vegetables when unit 1 is inactive. In both conjoint examples, understanding the neural code requires joint consideration of both units. (C) Anatomically, the units in a

Figure 2.2: representation may be localised to a contiguous region or dispersed across multiple distal areas, and the units may occupy either the same or different locations across individuals. The two brains within each white box denote two different individuals. Abbreviation: Betw. individuals, between individuals.

clearly separates living and non-living things along one diagonal, and flying from non-flying things along the other. In the right panel, unit 1 clearly encodes whether a stimulus is a plant or animal, but the behaviour of unit 2 considered independently might appear to be arbitrary (activating for *banana* and *cardinal*, but not for *carrot* or *fish*). Joint consideration of both units makes the interpretation of unit 2 clear: if unit 1 is active, it differentiates birds from fish; if inactive, it differentiates fruit from vegetables.

2.3.3 Variation of anatomical location

Within an individual, units representing a given semantic element may be anatomically **contiguous** (situated within the same brain region) or **dispersed** (residing in multiple separate regions; Figure 2.2C). Methods that analyse different areas separately (e.g analysis of different ROIs) favour the discovery of contiguous over dispersed representations. Finally, irrespective of whether units are contiguous or dispersed within an individual, signal-carrying units may be anatomically localised in the same or different areas across individuals. Averaging data across anatomically aligned brains (e.g in searchlight analyses) favours the discovery of similarly over differently localised representations, whereas techniques that align on the basis of similar responses to stimuli rather than anatomical location (e.g hyper-alignment) relax the localisation assumption.

Together these factors delineate 24 different possibilities for the organisation of the neuro-semantic code within and across individuals (Table 2.1). These are not mutually exclusive – different aspects of a representation, or representations in different conceptual domains, may be organised according to different principles. Understanding which principles best explain which aspects of representation thus requires methods capable of finding each variety of signal.

2.4 Assumptions implicit in analytic approaches

We next consider how different analytic approaches in functional brain imaging might favour the evaluation of some hypotheses over others. Such studies aim to find the units whose

Code	Within subject		Across subjects		Single voxel	Spatial blurring	ROI/SL	Average before model fitting	Average after model fitting
Type	Code	Location	Code	Location	n = 46	n = 40	n = 63	n = 45	n = 64
Independent	Homo	Contiguous	Consistent	Same	**100**	**100**	**100**	**100**	**100**
Independent	Homo	Contiguous	Consistent	Different	**100**	**100**	**100**	**100**	*10*
Independent	Homo	Contiguous	Inconsistent	Same	**100**	**100**	**100**	*62*	**62**
Independent	Homo	Contiguous	Inconsistent	Different	**100**	**100**	**100**	*62*	*9*
Independent	Homo	Dispersed	Consistent	Same	**100**	**100**	*42*	**42**	**42**
Independent	Homo	Dispersed	Consistent	Different	**100**	**100**	*42*	**42**	*9*
Independent	Homo	Dispersed	Inconsistent	Same	**100**	**100**	*42*	*23*	*23*
Independent	Homo	Dispersed	Inconsistent	Different	**100**	**100**	*42*	*23*	*8*
Independent	Hetero	Contiguous	Consistent	Same	**100**	*60*	**60**	**60**	**60**
Independent	Hetero	Contiguous	Consistent	Different	**100**	*60*	**60**	**60**	*9*
Independent	Hetero	Contiguous	Inconsistent	Same	**100**	*60*	**60**	*36*	**36**
Independent	Hetero	Contiguous	Inconsistent	Different	**100**	*60*	**60**	*36*	*8*
Independent	Hetero	Dispersed	Consistent	Same	**100**	*60*	*30*	**30**	**30**
Independent	Hetero	Dispersed	Consistent	Different	**100**	*60*	*30*	**30**	*8*
Independent	Hetero	Dispersed	Inconsistent	Same	**100**	*60*	*30*	*17*	**17**
Independent	Hetero	Dispersed	Inconsistent	Different	**100**	*60*	*30*	*17*	*7*
Conjoint	Hetero	Contiguous	Consistent	Same	*46*	*23*	**23**	**23**	**23**
Conjoint	Hetero	Contiguous	Consistent	Different	*46*	*23*	**23**	**23**	*2*
Conjoint	Hetero	Contiguous	Inconsistent	Same	*46*	*23*	**23**	*15*	**15**
Conjoint	Hetero	Contiguous	Inconsistent	Different	*46*	*23*	**23**	*15*	*2*
Conjoint	Hetero	Dispersed	Consistent	Same	*46*	*23*	*3*	**3**	**3**
Conjoint	Hetero	Dispersed	Consistent	Different	*46*	*23*	*3*	**3**	*1*
Conjoint	Hetero	Dispersed	Inconsistent	Same	*46*	*23*	*3*	*3*	**3**
Conjoint	Hetero	Dispersed	Inconsistent	Different	*46*	*23*	*3*	*3*	*1*

Table 2.1: Twenty-four hypotheses about the nature and anatomical organisation of the neuro-semantic code. Each row indicates one hypothesis and the first five columns show corresponding combinations of key factors discussed in the text (code type, within-subject homogeneity and localisation, and between-subject consistency and localisation). The remaining columns summarise a review of 100 papers using multivariate methods to uncover neuro-semantic representations. Each column represents a common analysis step that entails an implicit assumption about the neural code, including independent analysis of single voxels (assuming an independent code), spatial blurring of BOLD (assuming a homogeneous code), independent consideration of different areas via ROI or searchlight (assuming contiguous localisation within area), averaging the neural signal across subjects before model fitting (assuming a consistent code), and averaging of model fit data across subjects (assuming similar localisation). The *n* indicates how many papers adopted the corresponding step. Emphasis shows hypotheses where the associated step will benefit (bold font) or hinder (italic) discovery. The numbers indicate how many reports are capable of detecting each possible neural code considering the analysis decisions taken at each step from left to right. The final column indicates the number of reports that adopt choices capable of finding each possible code. Abbreviations: Hetero, heterogeneous; Homo, homogeneous.

measured responses to stimuli encode the representational elements specified by the cognitive theory. Because all imaging methods yield thousands of noisy measurements for each stimulus in each participant, statistical models that seek informative units must be constrained in some way. Multivariate methods vary in their approach to this problem and thus in their ability to detect different types of representations. We consider three broad approaches and their variants (Figure 2.3) with an eye to highlighting their respective strengths and limitations. Box 2 additionally considers crucial but commonly overlooked issues for collecting the data that feed these different approaches.

Multivariate pattern classification (MVPC) fits models (Gaussian naive Bayes, support vector machines, logistic/multinomial regression, etc.) to categorise stimuli from the neural activity they evoke (Norman et al., 2006; Pereira & Botvinick, 2011). During a training phase, the model receives **labelled data** consisting of the neural responses across units to each of many stimuli (e.g various images of objects) and, for each item, a label indicating the stimulus category. Training involves fitting classifier weights to output the correct label for each item in the training set. The trained model is then evaluated by assessing whether it outputs the correct category label when given neural responses for test stimuli that are not present in the training set. Where a fitted model reliably classifies held-out items, input units are interpreted as encoding information about the target categories. The approach is transparently consistent with category-based semantic representations but will also yield positive results for both feature-based and vector space representations provided that the target categories are separable in the corresponding neural activation patterns (i.e., it is possible to fit a flat hyperplane that reliably divides the target categories in the high-dimensional representation space). Because the output of a classifier depends on activation patterns across multiple units, MVPC can detect both independent and conjoint codes. Classifiers assign unique weights to each unit, and the approach can therefore detect both homogeneous and heterogeneous codes. Because separate classifiers are typically fitted for each participant, the method can potentially find inconsistent and variably localised representations as well.

A key challenge for MVPC concerns over-fitting. With more predictors (neural measurements) than datapoints (stimuli), model fitting is underdetermined without additional constraint – even with random data, an infinite set of coefficients will perfectly predict the category membership of training items (Pereira et al., 2016). MVPC variants differ in the constraints they impose to handle this issue; this has important implications for signal discovery

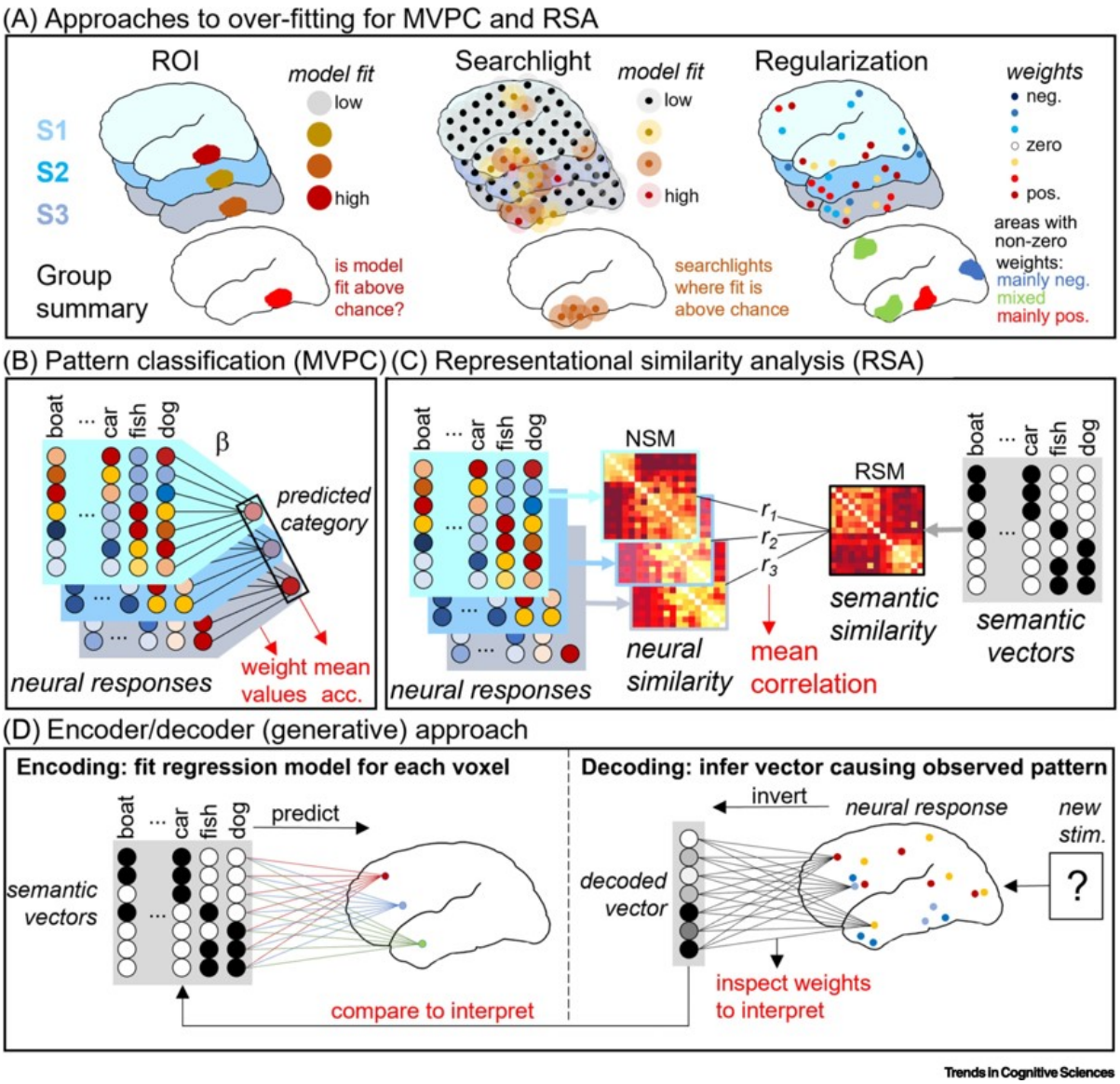

Figure 2.3: Approaches to neural decoding. (A) Different solutions to the over-fitting problem faced by multivariate pattern classification (MVPC) and representational similarity analysis (RSA) approaches. Region of interest (ROI) approaches look only at a prespecified area in each participant and evaluate whether the mean model fit (i.e., hold-out error or correlation) across participants differs reliably from chance. Searchlight methods independently evaluate model fit at many "searchlights" throughout the brain in each participant, then find areas where searchlights produce above-chance fits reliably across participants. Regularisation fits a single model in each participant using all neural features, but constrains the model to minimise prediction error jointly with an additional cost that prevents over-fitting (discussed in the main text). Non-zero coefficients in the decoding model of a subject indicate neural units that carry signal; these can be distributed across the brain and can be different for each participant. Group maps indicate areas where non-zero coefficients accumulate more than expected by chance across individuals. (B) Multivariate pattern classification fits a model to predict a stimulus category label from the neural pattern it evokes across selected neural units. Mean hold-out accuracy across participants indicates whether the selected units carry category information and classifier weights can indicate whether category membership is signalled by increased or decreased neural activation. (C) RSA computes similarity in the neural responses generated across selected units

Figure 2.3: by various stimuli, and then correlates this with a target semantic similarity matrix. Mean correlation across subjects indicates whether the selected neural units encode semantic structure. (D) Generative approaches use regression to fit models that predict the response of each neural unit to various stimuli. After fitting, the regression weights can be inspected to determine the information that each unit encodes, and novel brain responses can be "decoded" by finding the semantic vector most likely to have generated the observed neural pattern and then comparing this to known semantic vectors. Abbreviations: acc., accuracy; Neg., negative; NSM, neural similarity matrix; Pos., positive; RSM, representational similarity matrix; S1–S3, brains from three different subjects; stim., stimulus.

(Figure 2.3A).

One method is to reduce the number of neural features provided as the input to the model by applying an explicit anatomical constraint. For instance, ROI-based approaches look only at the units contained in a predefined ROI – discovery therefore requires that the representation is anatomically contiguous and localised similarly across individuals, and also that a sufficient amount of the representation falls within the preselected region to drive classifier accuracy above chance. ROI selection also crucially determines how neural evidence can relate to the space of cognitive hypotheses. For instance, ROIs falling outside modality-specific areas cannot offer evidence relevant to testing grounded theories of representation, whereas those falling solely within a given modality-specific region cannot evaluate self-contained hypotheses.

Relatedly, searchlight approaches fit a separate classifier at each spatial location in each participant (e.g each voxel, source, or electrode), including as predictors all units within a prespecified anatomical radius ("searchlight"; Kriegeskorte et al., 2006; Norman et al., 2006). Thus, different brain regions are analysed separately. Typically cross-participant univariate statistics at each location assess where in the brain the classifier hold-out accuracy is reliably better than chance; this approach therefore requires that the representation is localised similarly across individuals. If this criterion is met, the searchlight can reveal anatomically dispersed codes, but only if each searchlight independently contains sufficient information to drive classifier accuracy above chance. If accurate classification depends on joint consideration of units that fall in separate searchlights, the code will be missed. In this sense, the searchlight may fail to find dispersed, conjoint codes (C. R. Cox et al., 2015; C. R. Cox & Rogers, 2021).

Note that, in principle, classifier accuracy for searchlights and ROIs could be analysed separately in each individual, relaxing the assumption of similar localisation across participants. We are not aware of such an approach being applied to semantic decoding and we therefore focus on the more usual method of using cross-subject univariate statistics to create group-level

information maps for these approaches.

A second approach chooses classifier inputs based on a summary univariate statistic that is computed independently for each unit (such as an F-statistic that contrasts unit activation for different category members (Visconti di Oleggio Castello et al., 2021), or a correlation-based metric that assesses the stability of the response of a voxel across stimuli (Vargas & Just, 2019). This avoids the anatomical assumptions of ROI and searchlight approaches but lacks a principled rationale for setting a cut-off threshold and may fail to discover conjoint representations because each included unit must independently survive the preselection criterion.

A third strategy employs model **regularisation**: all units in the cortex provide input to the classifier, which avoids over-fitting by jointly minimising classification error and an additional loss that is itself a function of the classifier weights (C. R. Cox & Rogers, 2021). Common losses include the sum of the squared coefficients (L2-norm, also known as "ridge" regression; Hoerl & Kennard, 1970), the sum of their absolute values (L1-norm, also known as "LASSO" (least absolute shrinkage and selection operator); Tibshirani, 1996), or a weighted average of these (also known as "elastic net"; Jia & Yu, 2008). The approach makes no assumption about the anatomical location of signal-carrying units within or across participants, can detect conjoint representations (because it does not require independent preselection of classifier units), and offers a principled way to guide parameterisation via nested cross-validation of prediction error (C. R. Cox & Rogers, 2021).

Crucially, however, different regularisers impose different constraints on model fitting, leading to wildly different solutions (C. R. Cox & Rogers, 2021). Regularisation with the L1 norm zeros out as many predictors as possible while still maximising predictive accuracy, and typically "selects" (i.e., places non-zero coefficients on) a very small proportion of units. By contrast, the L2 norm spreads similar weights across correlated units and places non-zero weights on all units. The choice of regulariser thus implements an assumption about the likely nature of the true signal: that signal-carrying units are sparse and uncorrelated (L1) or that they are dense and highly redundant (L2). An alternative approach designs loss functions that explicitly incorporate prior knowledge about the likely neural and cognitive structure. For instance, the sparse-overlapping-sets LASSO (SOSLASSO) penalty encourages patterns of "structured sparsity" where selected units reside in roughly similar locations across participants, promoting loose anatomical clustering that still permits some variation in signal location across participants (Rao et al., 2013, 2016).

These differences can yield radically different views of the neuro-semantic code when applied to the same data. In Figure 2.4A , neural representations of face stimuli appear to be increasingly widely distributed and heterogeneous as analytic methods progressively relax tacit assumptions about the independence, heterogeneity, and localisation of the neural code. Standard univariate contrast (assuming a consistently localised, independent, and homogeneous code) replicates the classic finding of a right-lateralised posterior fusiform area that is more active for faces. Searchlight (assuming a similarly localised and contiguous but potentially conjoint and heterogeneous code) suggests a bilateral representation localised to posterior ventral temporal cortex. Whole-brain MVPC regularised with the L1 norm (assuming a sparse code that can be dispersed, heterogeneous, and differently localised) shows a bilateral face-to-nonface gradient in posterior ventral temporal cortex and a face-selective region in right lateral occipital cortex. Regularisation with the SOS LASSO (allowing dispersed, heterogeneous, and differently localised codes, but preferring solutions with roughly similar anatomical distributions) suggests a much more broadly distributed code encompassing anterior temporal, parietal, and prefrontal regions in both hemispheres (C. R. Cox & Rogers, 2021).

Representational similarity analysis (RSA) searches for sets of units whose responses express semantic similarities among stimuli (Kriegeskorte et al., 2006; Norman et al., 2006; Pereira et al., 2009). The analysis first computes a target representational similarity matrix (RSM; sometimes defined in terms of dissimilarity where it is called a target representational dissimilarity matrix) that expresses semantic relatedness for all pairs of stimuli (Box 1). It then estimates a neural similarity matrix (NSM; sometimes called a neural representational dissimilarity matrix) that encodes pairwise similarities in stimulus-evoked neural activity across a set of units. The correlation between RSM and NSM indicates whether the selected units encode the target structure (Figure 2.3C).

Similarly to MVPC, RSA can detect categorical, feature-based, and vector space representations provided that the NSM and semantic RSM correlate positively. Because neural similarities are computed across multiple units, the technique can detect conjoint or independent codes and heterogeneous or homogeneous codes. A central challenge concerns how neural units are selected and evaluated for significance. Most studies employ either a prespecified ROI or a searchlight technique. The correlation between RSM and NSM is computed for each ROI or searchlight individually in each participant and, if these are reliably positive across individuals, the ROI/searchlight is interpreted as encoding semantic structure. As with MVPC,

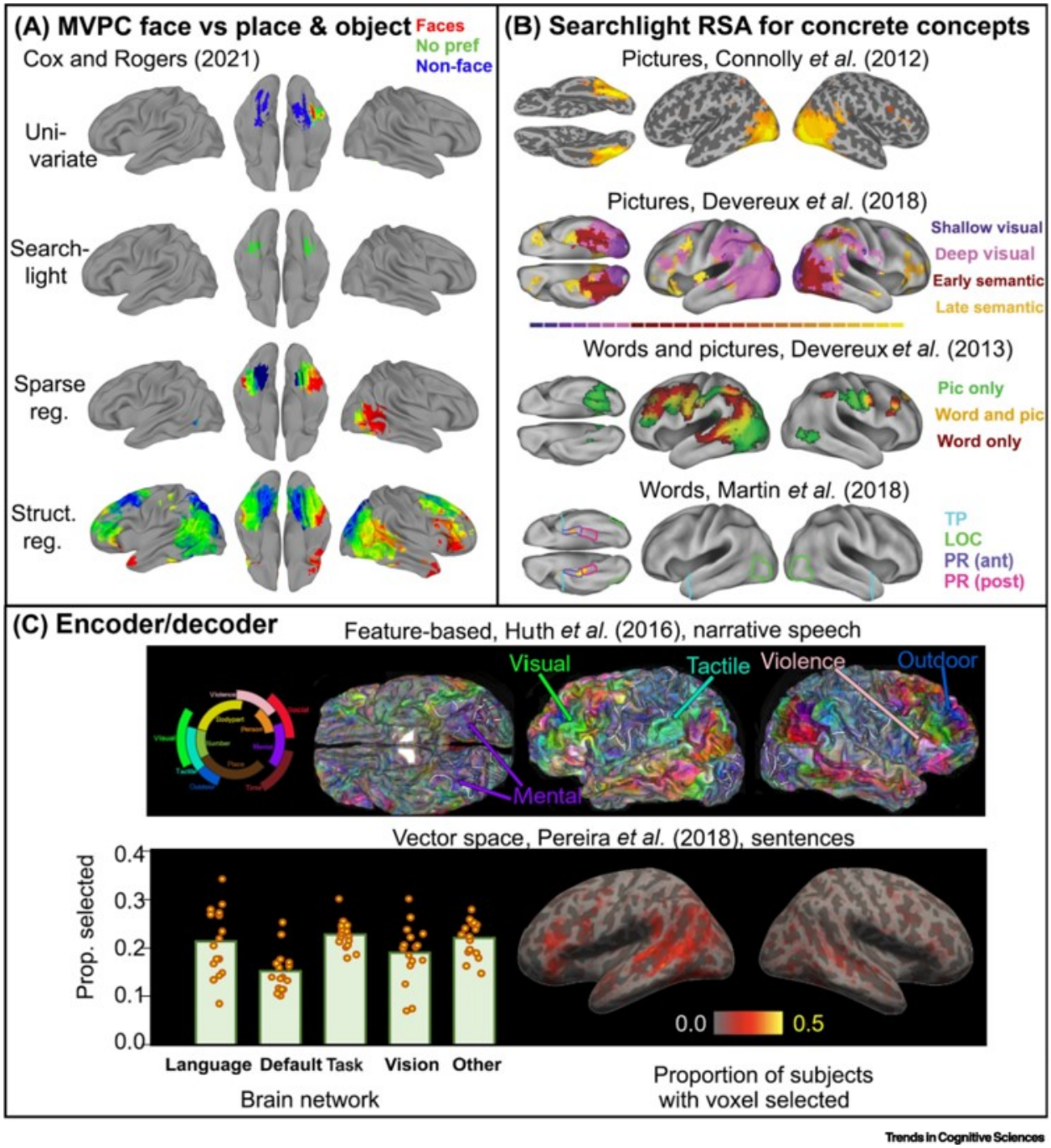

Figure 2.4: Example results from various decoding methods applied to fMRI data. (A) Four different multivariate pattern classification (MVPC) approaches applied to the same dataset. Participants made pleasantness judgments in response to images of faces, places, or objects, and each analysis sought voxel sets that differentiate face from non-face stimuli. Approaches that assume consistently localised signals (univariate and searchlight) suggest that representations are localised to posterior ventro-temporal cortex, whole-brain decoding with sparse regularisation suggests a somewhat more distributed representation, whereas decoding with structured sparsity suggests a widely distributed representation (C. R. Cox & Rogers, 2021). (B) Searchlight representational similarity analysis (RSA) decoding of semantic structure from pictures, words, or both. Results vary remarkably depending on several factors, including the representational similarity matrices (RSMs) considered (semantic similarity alone (Connolly et al., 2012) produces different results from comparing semantic versus visual similarity (Devereux et al.,

Figure 2.4: 2018); top two images) and experimental control of stimulus properties (semantic structure for words appears to be encoded in perisylvian regions when visual structure is uncontrolled (Devereux et al., 2013), but in ventral anterior temporal lobe (ATL) when controlled (C. B. Martin et al., 2018)). (C) Generative approaches for decoding semantic representations of narrative speech/sentences. When predictor vectors have semantically interpretable dimensions, and encoder weights are used to interpret the meaning of a voxel's activation, the results seem to show a mosaic of localised semantic features across cortex within each subject, but callouts show areas where the proposed semantic content is at odds with traditional understanding of function (top; images generated from online visualisation tool at https://gallantlab.org/huth2016/). Approaches that invert encoding models to decode whole-brain states (bottom) can recover sentence meanings with good accuracy, but the nature of the underlying code is difficult to discern because the approach selects thousands of voxels widely distributed across cortex in each participant (right), with approximately equal proportions residing in various pre-defined brain networks (Pereira et al., 2018; left). In both cases verbal semantic representations appear to be widely distributed across cortex and highly variable across individuals. (For references see Connolly et al., 2012; C. R. Cox & Rogers, 2021; Devereux et al., 2013, 2018; Huth et al., 2016; C. B. Martin et al., 2018; Pereira et al., 2018.) Abbreviations: ant, anterior; LOC, lateral occipital complex; Pic, picture; post, posterior; pref, preference; PR, perirhinal cortex; Prop., proportion; reg., regularisation; TP, temporal pole.

information maps could be analysed separately in each individual, but RSA as typically practiced requires that (1) representations are localised similarity across individuals, (1) information is not conjointly encoded across different searchlights or ROIs, and (3) individual searchlights contain sufficient information to drive correlations with the target matrix reliably above chance.

RSA views even small correlations as meaningful provided that they are reliably positive across participants. Because semantic structure covaries with many confounding factors, the results can be difficult to interpret. For instance, early studies using visual stimuli suggested that posterior temporo-occipital areas encode semantic structure (Connolly et al., 2012), but a recent comparative analysis found that these areas more strongly encode high-order visual structure and semantic structure was better encoded in more anterior ventro-temporal regions (Figure 2.4B, top; Devereux et al., 2018). Studies that do not control for visual similarity suggest that semantic structure for both words and pictures is encoded within a left perisylvian network (Devereux et al., 2013), but when stimuli orthogonally vary semantic and visual similarity, semantic structure for words appears to be localised to the medial-ventral anterior temporal lobe (C. B. Martin et al., 2018; Figure 2.4B, bottom). Thus, very different patterns are obtained depending upon the target RSMs, the selection of stimuli, and the input modality (Box 2).

Box 2: Implications for data acquisition

Hypotheses about the cognitive and neural systems supporting semantic cognition have crucial implications, not only for how neural data are analysed, but also for how data are collected.

Stimulus selection

Each modality of stimulus has advantages and disadvantages. Words are easily presented in the scanner, allow all concept types to be probed, and have a perceptual/orthographic structure that is unconfounded with semantic structure. However, decoding is less successful with words than with picture stimuli generally (Shinkareva et al., 2011) and written words generate a strongly asymmetric (left hemisphere) distribution of activation that contrasts with the bilateral pattern found for pictures and spoken words (Liuzzi et al., 2015).

Task selection

Tasks used to elicit semantic activation vary across studies in ways that are known to strongly impact the engagement of underlying neural systems, including their overall difficulty (Sabsevitz et al., 2005), the specificity with which an item must be identified for good performance (Rogers et al., 2006), reliance on strongly versus weakly encoded information (Noonan et al., 2013), aspects of knowledge the task foregrounds (Chiou et al., 2018; A. Martin, 2007), and the degree to which the task can be performed via alternative, non-semantic processing routes (Graves et al., 2010).

Temporal and spatial resolution

Neuroimaging methods vary in spatial and temporal resolution, limitations that may or may not affect discovery depending on the nature of the underlying code. For instance, the lag in BOLD means that successive stimuli blend into one another in fast event-related designs, which can hinder discovery if the neural code is heterogeneous. Slow event-related methods avoid temporal blending (Lewis-Peacock & Postle, 2008) but cannot be used for richer tasks such as connected speech or movie-viewing. EEG and MEG offer higher temporal resolution and thus avoid stimulus-to-stimulus blending, but at the cost of spatial blending that can compromise discovery if the neural code is heterogeneous or anatomically dispersed. ECoG offers temporal and spatial precision, but only a minority of regions are ever probed because

the sensors are placed for clinical need and only in patients who need neurosurgical intervention.

Image acquisition

The possibility that semantic representations are anatomically dispersed must be tested with whole-brain imaging, thus posing a challenge for fMRI acquisition where the signal-to-noise ratio varies substantially across the brain (T. T. Liu, 2016). Standard sequences yield especially poor signal in orbitofrontal and ventral anterior temporal regions that are thought to be crucial for semantic cognition (Binney et al., 2010). Strategies for improving the signal, including distortion-corrected spin-echo (Binney et al., 2010; Embleton et al., 2010) and multi-echo protocols (Halai et al., 2014; Kundu et al., 2017), have been available for several years but have only rarely been applied in semantic studies (Asyraff et al., 2021). Indeed, many studies have restricted the field of view to exclude ventral anterior temporal lobe (ATL) completely (Visser et al., 2010).

Finally, encoder/decoder (also known as generative) approaches use regression to fit a separate encoding model for each unit, predicting its response to a stimulus from the semantic features of the item (Just et al., 2010; Mitchell et al., 2008; Pereira et al., 2011). Successful prediction indicates that the corresponding unit independently encodes semantic information. A whole-brain response can be estimated by passing a stimulus feature vector forward through each encoder, yielding a predicted activation at every unit (Mitchell et al., 2008). Alternatively, the whole-brain response generated by a new, unknown item can be decoded by inverting the encoding models to find the semantic vector most likely to have generated the observed neural response, and then interpreting the resulting vector (Pereira et al., 2011, 2018; Figure 2.3D). Because separate models are fitted for each voxel and participant, generative approaches make no assumption about code homogeneity, cross-participant consistency, or anatomical organisation within or across individuals. However, they do face two non-trivial challenges.

First, generative approaches can fail to predict the independent activity of a unit that forms part of a conjoint code. To see this, consider the second conjoint example in Figure 2.2B right, where two units both contribute to a semantic representation. If unit 1 is active, unit 2 differentiates fish from birds; if inactive, unit 2 instead differentiates fruits from vegetables. The "meaning" of unit 2 is clear when unit 1 is taken into consideration, but might appear arbitrary

when considered independently. An encoder model might struggle to predict the independent behaviour of unit 2 from semantic features such as *can move*, *has feathers*, *is sweet*, etc., and thus might suggest that it is not involved in semantic representation.

The second challenge concerns interpretation. One strategy fits the encoders using semantic vectors whose elements are each individually interpretable (such as a semantic feature vector; Box 1), and then inspects the encoder weights for each unit to understand what content it encodes (Huth et al., 2012, 2016; Popham et al., 2021). For instance, if the activation of a voxel is reliably predicted by semantic features such as *can move*, *can grow*, and *has eyes*, these features will receive non-zero weights in the regression model for that voxel, which might then be interpreted as encoding animacy. The goal is to understand each unit as independently encoding a subset of semantic features, thereby yielding an interpretable semantic feature map of cortex that is consistent with feature-based cognitive models. Because there are many potential semantic features, however, the encoder fit must be regularised using techniques such as those described earlier for MVPC (commonly L2 norm (e.g Huth et al., 2012), although other approaches are also popular (e.g. Nunez-Elizalde et al., 2019)). As we have seen, different regularisers can produce dramatically different configurations of weights, and the interpretation of encoder weights therefore hinges crucially upon the choice of the regulariser. Perhaps for this reason, approaches adopting this strategy have yielded puzzling findings – suggesting a mosaic-like organisation of local semantic features across many cortical areas that is difficult to reconcile with the wealth of cognitive and clinical neuroscience information about the functions of these regions (Huth et al., 2016; Figure 2.4C, top).

An alternative strategy eschews the effort to identify a "meaning" for individual units and instead decodes the full activation pattern evoked across cortical units by inverting the encoder models to find the semantic vector that is most likely to have generated the whole-brain response. The recovered vector is interpreted by comparing its similarity to vectors corresponding to known words or sentences (Pereira et al., 2011, 2018). For instance, if the decoded vector is near to the known vectors for *grow*, *move*, *eat*, *eyes*, *legs*, *fur*, it will be interpreted as encoding a meaning such as *animal*. Because no effort is made to interpret each dimension, this method is consistent with vector space approaches, but can also detect category or feature-based representations. One recent study showed remarkably good decoding of sentence-level meaning using this approach (Pereira et al., 2018) – but the implications of the study for understanding neural organisation of semantics remain unclear because the results

identified thousands of voxels scattered across the cortex in each individual, with approximately equal involvement of many different brain networks and no voxels selected in more than half of the participants (Figure 2.4C, bottom).

It is worth noting that each general approach encompasses several variants – for instance, in the particular classification model adopted by MVPC (Norman et al., 2006) and the specific similarity metric used by RSA (Diedrichsen & Kriegeskorte, 2017; Haxby et al., 2014). Although a full characterisation of each is beyond the scope of this review, it seems likely that such variation further contributes to the heterogeneity of the findings reported in the literature.

2.5 Analytic implications of grounded versus self-contained theories

The issues described above arise regardless of whether neuro-semantic representations are grounded or self-contained, but this important distinction in cognitive theories carries two additional implications for the design, analysis, and interpretation of multivariate imaging studies. First, primary and secondary perceptual and motor cortices conform to localisation assumptions that are central to particular analytic choices – specifically, such areas are both contiguous and localised similarly across individuals. Grounded approaches suggest that such areas can encode semantic information about stimuli, and studies designed specifically to assess whether semantic structure arises within a given modality (Carota et al., 2017; Clarke & Tyler, 2014) therefore have good motivation to employ ROI or searchlight-based feature selection. The anatomical organisation of tertiary and association cortices is less well understood and may be more likely to vary across individuals, therefore studies seeking semantic structure outside the earlier modality-specific regions are better served by the adoption of approaches that loosen localisation, homogeneity, and consistency assumptions. Assessment of self-contained hypotheses will depend crucially on such methods because they propose that semantic representations encode information in a modality-independent manner.

Second, adjudication of grounded versus self-contained hypotheses requires studies that probe semantic information through different stimulus modalities. Self-contained views hold that the same system of semantic representation is engaged regardless of whether the stimulus is a word, picture, image, sound, etc. Such a view cannot be disconfirmed by evidence that, for instance, semantic information is decodable from visual areas when a visual stimulus appears

because such a result might also arise if the structure of purely perceptual visual representations is confounded with semantic structure (e.g Figure 2.4B). Evaluating the proposal instead requires searching for neural systems from which semantic information can be decoded across multiple different stimulus modalities. Currently, the literature contains relatively few such studies, and these have yielded mixed findings (Devereux et al., 2013; Handjaras et al., 2016; Shinkareva et al., 2011; Simanova et al., 2014; further details are given in Appendix A).

2.6 Towards best practices

To understand how the preceding issues may have shaped current thinking about semantic representation in mind and brain, we reviewed 100 papers applying multivariate techniques to the discovery of neuro-semantic representations in fMRI data (Appendix A). For each, we considered five analytic decisions, each reflecting a latent assumption about the neural code, and we evaluated which of the 24 representational possibilities the study was capable of detecting as each choice was made. The results are summarised in Table 2.1. All methods were capable of detecting neural representations that adopt an independent, homogeneous, and anatomically contiguous code that, across individuals, is consistent and similarly localised – the type of representation sought by univariate analysis. Fewer could detect other types of representational structure, and very few were capable of finding representations that are dispersed in the brain, localised differently across participants, and/or encode semantic information conjointly across units rather than independently. In this sense, methodological choices made during data analysis determine which types of neural signal can and cannot be detected – the analytic decisions effectively filter the empirical record.

A central question thus concerns how the field might best proceed given the complexity and heterogeneity of contemporary methods and the filtering that inevitably results. No analytic approach is assumption-free, and we doubt that the universal adoption of any single method will resolve the issues we have identified. Instead, we believe the field would be well served by adopting some best practices in the way that studies are designed and results are communicated.

2.6.1 Articulating explicit hypotheses about the neural code

In laying out the motivation and design of a study, it is helpful for researchers to explicitly state their working hypothesis about the nature and structure of the neuro-semantic code – what form

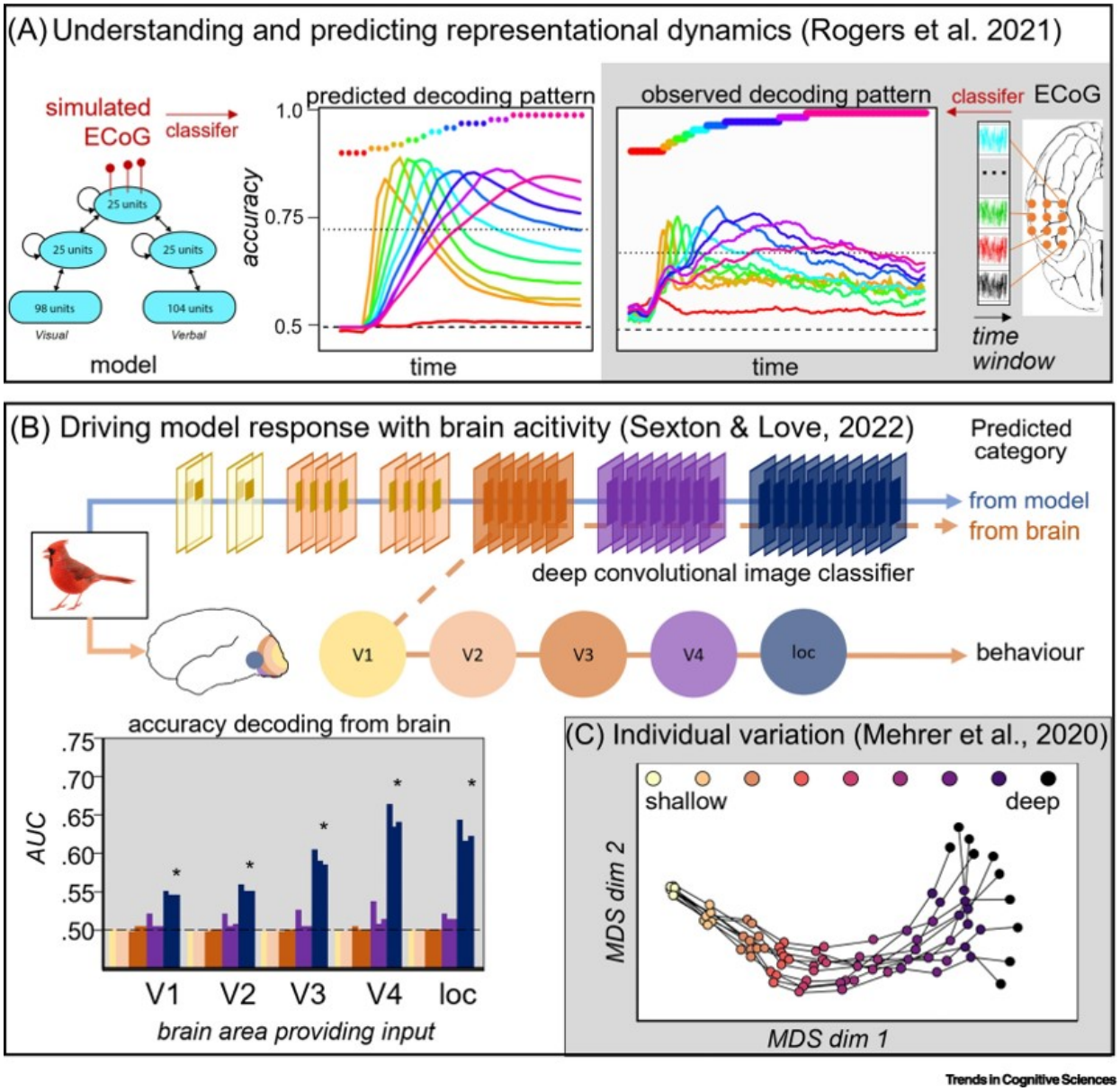

Figure 2.5: Recent examples of computational models informing neural decoding. (A) In recurrent models the activation patterns that encode semantic information change over the course of stimulus processing. In simulated electrocorticography (ECoG, left), classifiers fitted to different temporal windows (coloured dots) decode well within the same and neighboring time-windows, but poorly for more distal time-windows (coloured lines). A similar pattern arises when the same approach is used to decode ECoG from human anterior temporal cortex while participants name pictures, suggesting rapid nonlinear change in the neuro-semantic code (Rogers et al., 2021). (B) Deep convolutional neural networks (DCNNs) may provide a useful framework for understanding visual object semantics (Cadieu et al., 2014; Kriegeskorte, Mur, Ruff, et al., 2008). A recent study assessed whether a trained DCNN could classify images when activations at a given model layer were replaced by neural responses (measured by fMRI) of different visual areas (Sexton & Love, 2022). Neural patterns from each area were successfully decoded, but only when they were input to the deeper model layers (barplot) – suggesting that the richer semantic structure encoded in such layers is reflected throughout the ventral visual stream. (C) Other work uses similar models to evaluate individual differences across parts of the vision-to-semantics system (Mehrer et al., 2020). In the plot shown the authors trained several

Figure 2.5: models, measured similarity in the representational geometry acquired in each layer across models, and embedded these in two dimensions. The proximity of coloured circles indicates the similarity of the representational structure acquired by the corresponding layers. Lines connect layers in the same model. Shallower model layers (light colours) always learned relatively similar structure, whereas deeper layers – those most likely to express abstract semantic structure – learned more variable structure, suggesting that neural codes may differ more across individuals in the regions that are most likely to encode semantic structure. (For references see Mehrer et al., 2020; Rogers et al., 2021; Sexton & Love, 2022). Abbreviations: AUC, area under the curve; dim, dimension; LOC, lateral occipital complex; MDS, multidimensional scaling; V1–V4, visual cortex areas 1–4.

the cognitive representation is hypothesised to take, how its neural instantiation is reflected in the measurements taken, and how it is expected to vary within and across individuals. The cognitive and neural possibilities developed in this review provide a frame of reference for such statements, which are important because they allow the reader to understand why a given analysis method was chosen and how the observed results relate to the working hypothesis.

2.6.2 Explicit consideration of alternative hypotheses

When designing/motivating an analysis and when drawing conclusions from the results, it is helpful for researchers to consider other possible ways that the target information might be encoded in neural activity, beyond the working hypothesis. Before data collection, such habits can prompt new design or analysis ideas that allow adjudication of a richer variety of hypotheses. When drawing conclusions, explicit consideration of alternative possibilities and whether/how the current data can possibly disconfirm them can help the community to better understand seemingly heterogeneous patterns of results.

2.6.3 Connection to neurocognitive computational models

One way to make working assumptions about representation explicit is to connect the experimental design and analysis plan to a neuro-computational model of the behaviour (C. R. Cox et al., 2015; C. R. Cox & Rogers, 2021; Kriegeskorte, 2015; Richards et al., 2019; Rogers, 2020; Yang et al., 2019; Yuste, 2015). Figure 2.5 shows three recent examples. This connection serves several purposes. First, it provides a bridge between functional imaging results and explicit hypotheses about the mechanisms supporting the behaviour of interest, rendering the neural data a supporting part of a broader set of ideas about how the system works. Second, such models can offer new hypotheses about the nature of the neural code that might not otherwise

Box 3: The importance of converging evidence

The heterogeneity of imaging findings may be resolved by considering how conclusions from various studies relate to converging evidence from other methods. Some examples are given below.

Neuropsychology

Several varieties of brain damage cause semantic impairment and distinct deficits are observed depending on the neuropathology. Close consideration of these can illuminate brain imaging results. For instance, cross-modal semantic impairment can arise both from bilateral damage to the anterior temporal lobes (ATLs; Acosta-Cabronero et al., 2011; Adlam et al., 2006) and from left frontoparietal or posterior-lateral temporal stroke (Jefferies & Lambon Ralph, 2006; Rogers et al., 2015), but whereas ATL damage erodes conceptual structure, frontoparietal/posterior-lateral temporal damage instead disrupts the ability to shape semantic processing to the task context (Lambon Ralph et al., 2017). Thus, results implicating frontoparietal/posterior lateral temporal areas in semantics might best be interpreted by considering the demands on semantic control, whereas studies seeking conceptual structure in the brain should employ methods that are capable of resolving ATL signal.

Neural disruption

If imaging results suggest that a brain region selectively represents/processes a particular type of semantic information, transient disruption of the area via **transcranial magnetic stimulation (TMS)** should selectively affect retrieval of the target information. For instance, TMS applied to left or right ATL slows semantic judgments equally for animates and inanimates, but does not affect number judgments, supporting the view that bilateral ATLs encode semantic information across domains (Pobric et al., 2010b). Such studies will be especially important for testing the implications of multivariate imaging studies indicative of highly unorthodox semantic functions for various cortical areas (Huth et al., 2016).

Neural connectivity

The neural response of a given area can reflect its broader connectivity, with implications for understanding its function. For instance, medial posterior fusiform cortex responds more to

artefact than animal names – a pattern observed both in sighted and congenitally blind individuals (Mahon et al., 2009, 2010). One interpretation suggests that different brain areas natively specialise to represent distinct semantic categories (L. Chen & Rogers, 2015). However, the area of interest is functionally (Mahon et al., 2007) and structurally (L. Chen et al., 2017) connected to dorsal areas that aid in object-directed actions, suggesting that the seeming category effect may instead arise from more effective interactions between this visual area and parts of the action system (L. Chen & Rogers, 2015; Mahon et al., 2007).

Neurocognitive development

Developmental trajectories can likewise aid the understanding of mature activation patterns. For instance, the right posterior fusiform responds strongly to face images in most literate adults, perhaps suggesting an innately dedicated system for face representation (Kanwisher, 2000; Kanwisher et al., 1997). However, face perception engages the fusiform bilaterally in pre-literate children (Behrmann et al., 2016), and the left hemifield/right hemisphere advantage for face recognition emerges late in development as a child learns to read (Dundas et al., 2013). Such data suggest that the mature pattern reflects, not innate specialisation for a visual category, but experienced-based tuning of visual perception (Plaut & Behrmann, 2011).

occur to the theorist. Third, neurocomputational models can be used to better understand the strengths and weaknesses of different analytic approaches: the theorist can probe model analogues of neural signals and evaluate whether a given technique is capable of discovering information of the type captured by the model. Fourth, models allow exploration of alternative possibilities – the strengths and limitations of a given approach can be illuminated by comparing and contrasting its results when applied to models that embody different assumptions about the neural signal.

2.6.4 Simplified open data

Multivariate imaging studies pose unique challenges for the open data movement. The path from raw data to published result is often complex, software- or system-dependent, contains default parameterisations that may go unexplained, and involves many intermediate data products between raw measurements and summary results that can be exceedingly large and difficult to

document. Any single workflow can require extensive effort for outside scientists to fully understand and, because new approaches arrive with daunting frequency, it is difficult to know which bespoke pathways are worth mastering. Nevertheless, each method we have described makes use, at some level, of common data elements that are easy to understand and not too large to document and share. These include (1) the matrix that encodes, for each subject, the estimated response of each neural unit (voxel, electrode, source, etc.) to each stimulus, (2) the coordinates of the units in a standard reference frame (e.g Montreal Neurological Institute (MNI) coordinates of voxels, time and location information for ECoG, etc.), and (3) meta-information about the stimuli (e.g category labels used for decoding, semantic feature vectors used in an encoding model, the similarity matrix used for RSA, etc.). Sharing only these elements in standardised form would provide minimally sufficient information for scientists to apply a variety of different techniques to a dataset, thus promoting better understanding of how results vary with the method of analysis.

2.6.5 Convergence with other forms of evidence

Functional imaging alone will not resolve the quest for neuro-semantic representations. A fuller understanding will require relating multivariate imaging results to other diverse sources of evidence in cognitive neuroscience, including (1) the rich neuropsychology literature documenting patterns of verbal and nonverbal semantic impairment and their underlying neuropathology (Acosta-Cabronero et al., 2011; Caramazza & Mahon, 2003; L. Chen & Rogers, 2014; Jefferies & Lambon Ralph, 2006; Mesulam et al., 2013; Patterson & Hodges, 2000), (2) methods for disrupting neural processing in healthy participants, which can provide crucial evidence about causality (Lambon Ralph et al., 2009; Pobric et al., 2007, 2010b), (3) structural and functional brain connectivity (Binney et al., 2010; L. Chen et al., 2017; Mahon et al., 2007), (4) patterns of behaviour and functional activation arising over typical and atypical development (Behrmann & Plaut, 2014; Plaut & Behrmann, 2011), and (5) results of behavioural studies arising in cognitive science (Mack & Palmeri, 2011; Rogers & Patterson, 2007; Van Rullen & Thorpe, 2001). Box 3 considers how these sources of evidence can aid the interpretation of imaging data. Of course, not every paper can comprehensively review a large and complex literature – but in drawing conclusions it can be helpful for authors to explicitly consider where these cohere with results from other methodologies, where they contradict such results, and where the relevant experiments have not yet been conducted.

2.7 Concluding remarks

Our review illustrates that methodological choices in multivariate neuroimaging analysis selectively filter data to promote discovery of some types of neuro-semantic codes over others. These considerations compel a re-evaluation of the literature. Over three decades many neuroimaging studies have reported cortical areas that locally encode a particular type of semantic information in a systematic way across individuals. The preponderance and replicability of such findings suggest that some elements of neuro-semantic representation must indeed be independent, contiguous, and localised similarly across individuals. However, because this is precisely the one form of neuro-semantic code that, among many possibilities, is most robust to methodological choices, the ubiquity of such findings does not signify that these are the only, or even the most important, elements of semantic representation. On the contrary, neurocomputational models of healthy and disordered semantic cognition typically acquire internal representations that are conjoint rather than independent, are distributed across units that may be anatomically dispersed, are heterogeneous in code, and are potentially localised differently across individuals (C. R. Cox et al., 2015; C. R. Cox & Rogers, 2021; Rogers, 2020). These latter forms of semantic representation are the least likely to be revealed by most current analytical methods. The few studies capable of finding such structure often reveal a more widely distributed, heterogeneous, and variable semantic code than other studies suggest (C. R. Cox & Rogers, 2021; Huth et al., 2016; Pereira et al., 2018). Thus there exists an important lacuna in the empirical landscape that must be filled if we are to develop a mechanistic understanding of semantic cognition in the brain. We hope that this article provides a first step toward an organising framework that can bring the current heterogeneity of findings under a common explanatory umbrella (see Outstanding Questions).

Outstanding Questions

- Which cognitive hypotheses best describe semantic representations? The multivariate methods considered in this review do not indicate whether the underlying representation is categorical, feature-based, or a vector space, or is self-contained versus grounded. MVPC can produce a positive result even if neural representations are vector spaces rather than categories, and RSA can generate a positive result even if neural representations are categories and not vector spaces. How then can brain

imaging adjudicate between these views?

- When different brain areas all encode semantic structure, what data can determine whether they support the same or different functions? Semantic structure has been observed across multiple brain areas, but disruption caused by brain damage or transcranial magnetic stimulation (TMS) can produce qualitatively different patterns of impairment – suggesting that these regions serve different functions in semantic cognition.
- Can imaging data resolve which aspects of a target representational structure are, or are not, encoded within a neural system? Many studies report above-chance decoding that is nevertheless relatively weak (e.g RSA correlations as small as $r = 0.03$, binary classification accuracy of 0.55, etc.). Such effects might arise because neural data are noisy, because the neural system encodes weak confounds with the target structure, or because it encodes only part of the target structure.
- Can a combination of approaches overcome the individual limitations of each method? Each technique has strengths and limitations; perhaps the fullest picture of semantics in the brain will arise from a combination of approaches that will allow the community to evaluate the full space of representational possibilities outlined in this review.

Acknowledgements: This work was supported by an MRC Career Development Award (MR/V031481/1) to A. D. H., by a grant from the Rosetrees Trust (A1699) to A. D. H. and M. A. L. R., and by an Advanced European Research Council (ERC) award (GAP 670428-30 BRAIN2MIND_NEUROCOMP), MRC programme grant (MR/R023883/1), and intramural funding (MC_UU_00005/18) to M. A. L. R.

Competing interests: The authors declare no conflicts of interest.

Supplementary material: Supplementary materials can be found in Appendix A.

Glossary

Category

(of a representation) composed of discrete, independent units that each correspond to a concept (such as *boat*, *vehicle*, or *yacht*).

Conjoint

(of a representation) consisting of units that express different semantic information depending on the states of other units.

Consistent

(of a representation) associated with the same direction of change in activation across individuals – for example, homologous voxels in different individuals become more active when representing *cat*.

Contiguous

(of a representation) composed of units residing in the same brain region.

Decoding

predicting the stimulus (or sometimes the properties of the stimulus, or of the task) experienced by a participant using patterns of activity across multiple neural units.

Dispersed

(of a representation) composed of units residing in different brain regions.

Electrocorticography (ECoG)

a method of measuring brain activity via intracranial electrodes placed on the cortical surface.

Electroencephalography (EEG)

a method of measuring brain activity via electrodes placed on the scalp.

Encoding model

a model that predicts the activity of a single neural unit using multiple independently interpretable features of the stimulus. Multiple encoding models are used to predict activity across multiple neural units.

Feature-based

(of a representation) composed of multiple independently interpretable features (such as *is red* or *can fly*).

Functional magnetic resonance imaging (fMRI)

a method of measuring brain activity by detecting changes in blood flow.

Grounded

(of a representation) requiring the generation of modality-specific surface representations to produce retrieval/inference.

Heterogeneous

(of a representation) consisting of units that adopt different activation states when representing a concept.

Homogeneous

(of a representation) consisting of units that all adopt the same activation state when representing a concept.

Inconsistent

(of a representation) associated with different directions of change in activation across individuals – for example, homologous voxels in multiple individuals behave differently when representing *cat*, some becoming more active and others becoming less active.

Independent

(of a representation) consisting of units that express the presence or absence of the same semantic information irrespective of the states of other units.

Labelled data

a dataset specifying both input and output values for fitting an encoding or decoding model.

Magnetoencephalography (MEG)

a method of measuring brain activity by measuring magnetic fields generated by neural activity.

Multivariate pattern classification (MVPC)

the categorisation of stimuli based on the neural patterns they evoke (a form of decoding).

Region of interest (ROI)

a subset of neural units, chosen in a hypothesis-guided way, upon which an analysis is conducted.

Regularisation

a method of avoiding overfitting by finding classifier weights that jointly minimise classification error and an additional loss which is a function of the classifier weights.

Representational similarity analysis (RSA)

a method of investigating representational structure by comparing the similarity structure recorded to that hypothesised.

Self-contained

(of a representation) encapsulating semantic information within itself such that mere activation of the representation brings about retrieval/inference.

Surface representation

a sensory representation of a stimulus that is modality-specific – for example, color (specific to the visual modality) or a paddling action (specific to the motor modality).

Transcranial magnetic stimulation (TMS)

the use of magnetic fields to temporarily and reversibly disrupt brain function.

Vector space

(of a representation) composed of a pattern across representational units, the meanings of which cannot be independently interpreted.

Chapter 3

All spectral frequencies of neural activity reveal semantic representation in the human anterior ventral temporal cortex

Foreword

Chapter 2 focused primarily on the computational and spatial properties of semantic representations. In this Chapter I use ECoG to investigate their temporal dynamics. By assessing the capacity of ECoG to detect semantic representations I address the second aim of this thesis.

This Chapter is a manuscript in preparation:

Frisby, S. L., Halai, A. D., Cox, C. R., Clarke, A., Shimotake, A., Kikuchi, T., Kuneida, T., Miyamoto, S., Takahashi, R., Matsumoto, R., Ikeda, A., Rogers, T. T., & Lambon Ralph, M. A. (in prep.). All spectral frequencies of neural activity reveal semantic representation in the human anterior ventral temporal cortex.

I designed this study together with my co-authors. I developed the preprocessing and analysis pipelines with my co-authors and then conducted all preprocessing and analyses. I wrote the first draft of the manuscript, which was then edited collaboratively.

Abstract

The hub-and-spoke model of semantic representation proposes that the ventrolateral anterior temporal lobe (vATL) functions as a crucial semantic "hub". However, our understanding of *how* activity in the vATL gives rise to semantic representation is incomplete. We used regularised logistic regression to decode animacy from time-frequency power and phase extracted from

electrocorticography (ECoG) grid electrode data recorded on the surface of human ventral temporal cortex. Power in all bands – theta (4 – 7 Hz), alpha (12 – 18 Hz), beta (13 – 30 Hz), gamma (30 – 60 Hz) and high gamma (42 – 200 Hz) produced above-chance decoding. However, power from a wide range of frequencies (4 – 200 Hz) produced significantly higher decoding accuracy and exhibited all the same distinctive properties as a neural network model of semantic representation. This work demonstrated that a complete theory of semantic representation must account for information representation in theta, alpha, beta, gamma and high gamma power.

Keywords: electrocorticography, ECoG, intracranial electrophysiology, time-frequency analysis, semantic representation, decoding, multivariate pattern analysis

3.1 Introduction

Semantic cognition is our ability to understand and respond in a meaningful way to objects that we see, language that we hear, events that we experience, and any other stimuli in our environment. The brain regions underpinning this ability are well-defined (Lambon Ralph et al., 2017); however, a complete account of *how* activity in those regions gives rise to representation of semantic information is still lacking. One pivotal and unanswered question relates to the role that the time-frequency power and phase of neural activity may play in representing semantic information. We address this lacuna by using (1) a rare and informative resource – electrocorticography (ECoG) grid electrode data, recorded from the cortical surface of 19 patients undergoing surgery for intractable seizures while they named pictures; and (2) a multivariate method, regularised regression, designed to reveal information almost irrespective of how the brain might encode that information (Frisby et al., Chapter 2).

The hub-and-spoke model of semantic representation posits that information specific to each sensorimotor modality is represented in modality-specific areas of association cortex. The ventrolateral anterior temporal lobe (vATL) binds this information, accumulated over time, into transmodal, transtemporal, generalisable representations of meaning (Lambon Ralph et al., 2010, 2017; Patterson et al., 2007; Rogers et al., 2004). This model is supported by convergent evidence

from neuropsychology (Hodges & Patterson, 2007), positron emission tomography (PET; J. T. Devlin et al., 2000), distortion-corrected functional magnetic resonance imaging (fMRI; Binney et al., 2010; Halai et al., 2014; Visser et al., 2010) magnetoencephalography (MEG; Mollo et al., 2017), transcranial magnetic stimulation (TMS; Pobric et al., 2007, 2010a, 2010b), invasive cortical stimulation mapping (Lüders et al., 1991; Matoba et al., 2024) and ECoG (Halgren et al., 2006; Nobre & McCarthy, 1995; Sato et al., 2021; Shimotake et al., 2015). Recently, multivariate approaches have been applied to ECoG data in an attempt to characterise, not simply where and when task-related changes in neuronal activity occur, but what information those changes represent. Most of these analyses have focused on voltage (Y. Chen et al., 2016; C. R. Cox et al., 2024; H. Liu et al., 2009; Nagata et al., 2022; Rogers et al., 2021) and/or spikes (Kraskov et al., 2007; Reber et al., 2019).

The role that particular neural frequencies play in semantic representation, however, is still unclear (cf. Clarke, 2020; Rupp et al., 2017; Wang et al., 2011). Frequency bands have been associated with a plethora of different functions, including functions that could be relevant to semantic representation. One pervasive perspective, stemming from animal intracranial electrophysiology, is that activity in the gamma band (>30 Hz) is the primary means by which information of any kind is represented in the brain (Başar-Eroglu et al., 1996; Bressler, 1990). Others propose more specific functions for gamma activity in semantic representation, such as binding of different features of an input, thought, or action into a whole (Merker, 2013; Sauseng & Klimesch, 2008; Stryker, 1989; Tallon-Baudry & Bertrand, 1999; von der Malsburg, 1995) and/or retrieval of information irrespective of sensory modality (Sauseng & Klimesch, 2008). Consistent with these hypotheses, changes in gamma activity have been observed with ECoG during naming (Arya, 2019; Cervenka et al., 2013; Edwards et al., 2010; Forseth et al., 2018; Kojima et al., 2013; Nakai et al., 2017, 2019; Snyder et al., 2023; Tanji, 2005) and with intracranial depth electrodes during semantic property verification (Chan et al., 2011). It is debated whether activity in the "high gamma" band (> 60 Hz; Crone et al., 1998, 2011) behaves sufficiently differently from activity in the low gamma range (30 – 60 Hz) to be considered a separate process – some evidence suggests that high and low gamma covary (Crone et al., 2001), while others find that high gamma activity dissociates from low gamma, for example during language tasks (Crone & Hao, 2002; Wang et al., 2011).

However, gamma is not the only frequency that could be important for semantic representation. Proposed roles for theta activity (4 – 7 Hz) in semantic representation include

abstraction of transtemporal concepts from multiple episodic experiences (Sauseng & Klimesch, 2008), encoding semantic similarity in a manner analogous to the encoding of physical space (Solomon et al., 2019; Spiers, 2020), linking spatially distributed components of representations, or retrieving and integrating semantic information in a way that is appropriate for the context (Halgren et al., 2015; Jackson, 2021; Marko et al., 2019). Additionally, under many circumstances, the power of gamma and/or high gamma is locked to the phase of theta activity (Aru et al., 2015; Canolty et al., 2006; Canolty & Knight, 2010; Lisman & Jensen, 2013; Sederberg et al., 2003). ECoG has been used to show that the dynamics of this "cross-frequency coupling" (Benítez-Burraco & Murphy, 2019) vary during language tasks (Hermes et al., 2014), although it is unclear what mechanistic role this phenomenon may play. Between theta and gamma, changes in activity in the alpha (8 – 12 Hz) and beta (13 – 30 Hz) bands are also associated with semantic tasks in both ECoG (Abel et al., 2015; Sato et al., 2021) and MEG studies (Clarke et al., 2018; Mollo et al., 2017). Proposed roles for alpha in semantic representation include organisation of incoming visual input (Clarke et al., 2018) and "semantic orientation" – the capacity to synthesise, and orient oneself within, the meaning of the objects in one's environment (Klimesch, 2012). Amid this dizzying array of band-specific hypotheses, some researchers have proposed that cortical activity simply produces "broadband" changes that encompass all frequencies (e.g. Miller, 2010) or that activity in different bands represents transmission over variable cortical distances (intrinsic resonance frequencies are expected to be inversely related to the connection distance; Canolty & Knight, 2010; Lachaux et al., 2012), which might be important for linking spatially distributed components of representation as found in the hub-and-spoke theory (Lambon Ralph et al., 2017).

The current study sought to clarify the role of frequency information in semantic representation in the vATL and to relate this to voltage findings (Y. Chen et al., 2016; C. R. Cox et al., 2024; H. Liu et al., 2009; Nagata et al., 2022; Rogers et al., 2021) and to other sources of evidence (Lambon Ralph et al., 2017). ECoG offers simultaneously excellent spatial resolution and excellent temporal resolution and is capable of revealing high gamma activity that is blocked by the scalp when using noninvasive methods such as EEG or MEG (Lachaux et al., 2012; Parvizi & Kastner, 2018). The dataset contained a large number of patients (n = 19) and recording from the cortical surface using grid electrodes offers much greater coverage than can be obtained with cortical depth electrodes, which is critically important for detecting semantic representations that may be distributed across space (Merker, 2013; Rogers et al., 2021). We use regularised

logistic regression to decode information from time-frequency power and phase – a method that, unlike decoding methods that have previously been applied to ECoG time-frequency data (Clarke, 2020; Rupp et al., 2017; Wang et al., 2011) makes very few assumptions about the properties of the underlying code. Regularised regression is able to detect representations whether individual populations encode pieces of information separately, or whether the relationship between information and neural activity can be understood only by considering multiple populations simultaneously (Frisby et al., Chapter 2). Since it relies on parameter fitting rather than simple correlations, it is resistant to false positives and false negatives associated with purely correlational approaches (C. R. Cox et al., 2024). We asked two questions. First, in comparison to voltage, we tested whether semantic information (specifically animacy) can be decoded from time-frequency power and/or phase, and, if so, from which frequency ranges. Second, to characterise the time-frequency code further, we explored whether it exhibits the deep, distributed, dynamic properties identified by Rogers et al. (2021) in both the vATL and the deep layers of a computational model of the hub-and-spoke theory of semantic representation. These four properties are:

1. *Constant decodability*. Neural activity predicts stimulus category at every time point once activation reaches the vATL "hub".

2. *Local temporal generalisation*. Classifiers generalise best to time windows close to the window on which they were trained and more poorly to time windows further away from the training time.

3. *Widening generalisation window*. The temporal window over which classifiers generalise grows wider over time.

4. *Change in code direction*. Increased or decreased neural activity can signify different semantic information at different points in time.

3.2 Results

3.2.1 Relative decoding accuracy using spectral frequency power or phase vs. voltage

The ECoG data were collected from 6 – 52 subdural grid electrodes implanted over the ventral anterior temporal lobe (vATL) in 18 patients. The patients named line drawings of animals and inanimate objects. Time-frequency power and phase were extracted using complex Morlet wavelet convolution (see Methods). Classifiers were fitted using L1 (LASSO) regularisation (Tibshirani, 1996) and assessed using ten-fold cross-validation. The classifiers were trained on either frequency features (vectors of power or phase values extracted from multiple frequencies from multiple electrodes at a single timepoint) or on voltage features (vectors of voltage values extracted from multiple electrodes from a 50 ms window centred on the timepoint of interest). Separate classifiers were trained for timepoints between 0 and 1650 ms in 10 ms time steps (0 ms, 10 ms, 20 ms, …). Each classifier was tested on every possible time window.

We first tested whether it was possible to decode the animacy of the stimuli using time-frequency power and/or phase and whether decoding performance was comparable to voltage (Rogers et al., 2021). We compared classifiers trained on all frequency features of power or phase at 60 frequencies, logarithmically spaced between 4 and 200 Hz, to classifiers trained on voltage features. Figure 3.1A shows the results. Time-frequency power showed an almost identical decoding profile to voltage – hold-out accuracy rose to around 0.7 at 200 ms after stimulus onset and remained significantly above chance (0.5) throughout the time window. At no point was it possible to decode animacy from classifiers trained on phase.

Next, we trained classifiers on frequency features composed only of power or phase values from one frequency range – theta (4 – 7 Hz), alpha (12 – 18 Hz), beta (13 – 30 Hz), gamma (30 – 60 Hz) and high gamma (42 – 200 Hz). Figure 3.1B shows the decoding profile in each frequency range. For power, the patterns were almost indistinguishable, showing the same rise to about 0.6 and remaining above chance for most of the time window of interest. This demonstrated that power in every range was sufficient, and no range was necessary, for above-chance decoding. For phase, decoding was rarely above chance, with the exception of theta in the window 0 – 500 ms. Accordingly, phase data were not analysed further.

Figure 3.1C shows hold-out accuracy for classifiers trained on power within a single range compared to accuracy for classifiers trained on power at all 60 frequencies. For each range there

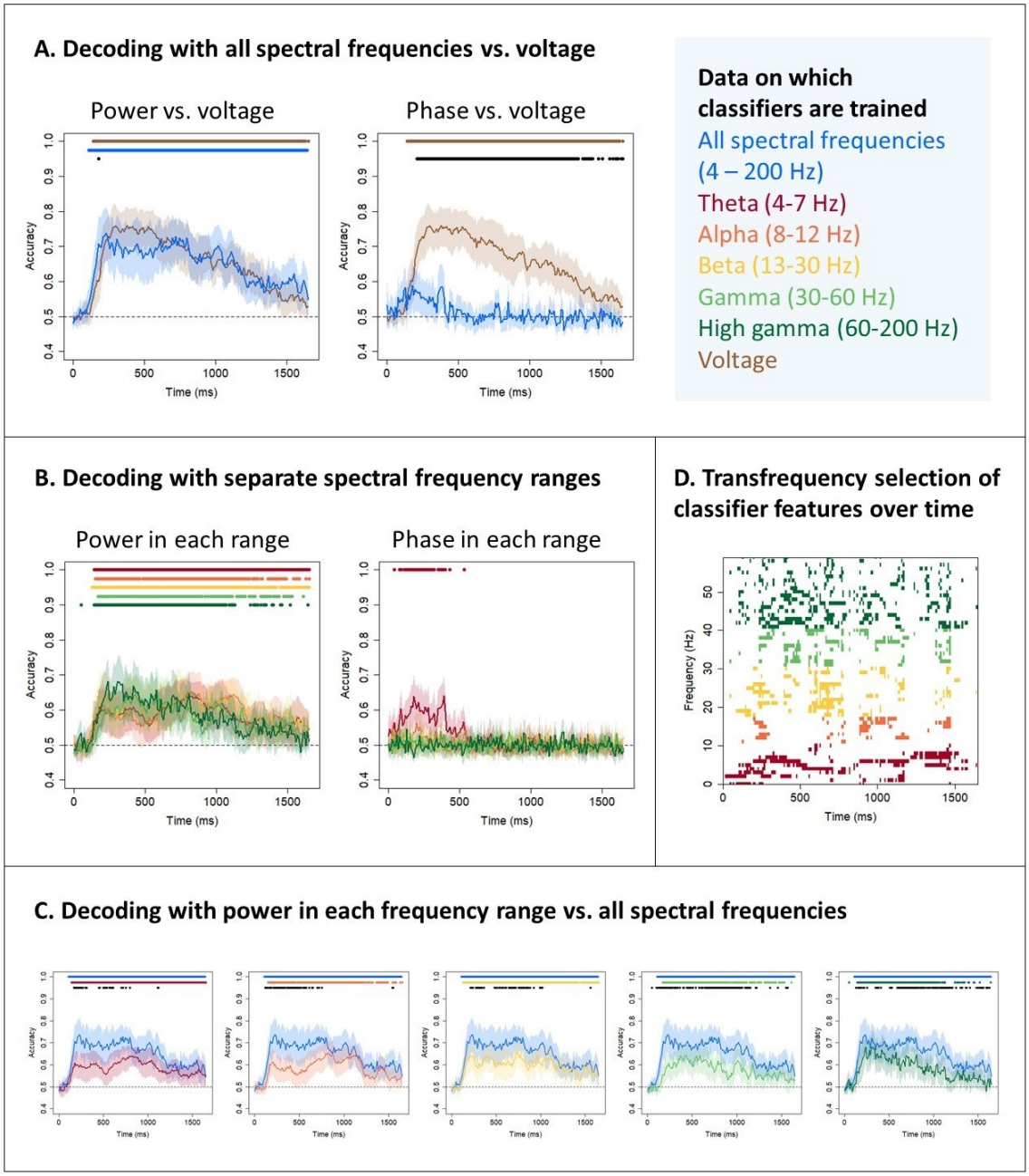

Figure 3.1: Semantic decoding with time-frequency power and phase. (A) Mean and 95 % confidence interval of the hold-out accuracy for classifiers trained on power or phase frequency features for all frequencies between 4 and 200 Hz (blue) or on voltage features (brown). Coloured dots indicate a significant difference between classifier accuracy and chance (0.5, one-sample *t*-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Black dots indicate a significant difference between accuracies at a given timepoint (paired *t*-tests with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (B) Mean and 95 % confidence interval of the hold-out accuracy for classifiers trained on frequency features composed only of power or phase values from a given range – theta (4 – 7 Hz, red), alpha (12 – 18 Hz, orange), beta (13 – 30 Hz, yellow), gamma (30 – 60 Hz, light green) and high gamma (42 – 200 Hz, dark green). Coloured dots indicate a significant difference between classifier accuracy and chance (0.5, one-sample *t*-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (C) Mean and 95 % confidence interval of the hold-out accuracy for classifiers

Figure 3.1: trained on frequency features composed only of power or phase values from a given range compared to classifiers trained on frequency features including all frequencies between 4 and 200 Hz. Coloured dots indicate a significant difference between classifier accuracy and chance; black dots indicate a significant difference between accuracies. (D) Feature selection of each frequency at each timepoint (averaged across electrodes) for classifiers trained on frequency feature vectors including all frequencies between 4 and 200 Hz for a single patient. Features with a nonzero coefficient are shown in colour.

was a time window during which decoding on all frequencies reached its peak and decoding on a single range performed significantly less well. This indicated that no individual range contains sufficient information to enable decoding accuracy comparable to accuracy based on all frequencies. To investigate which combination of frequencies enabled such successful decoding, we inspected the coefficients on each frequency. Figure 3.1D shows coefficients on each frequency, averaged across electrodes, for one sample patient (similar plots for all patients are shown in Appendix B, Figure B.1; plots showing coefficients on individual electrodes for one sample patient are shown in Appendix B, Figure B.2). Throughout the time window, frequencies across the whole range were selected by the classifier and thus contributed to decoding performance. This is evidence that the vATL represents semantic information via a transfrequency code: i.e., frequencies within theta, alpha, beta, gamma and high gamma all contribute to information representation.

3.2.2 Does the time-frequency semantic code exhibit deep, distributed, dynamic properties?

To characterise the time-frequency power code for animacy in more detail, we investigated whether time-frequency power exhibited the same deep, distributed, dynamic properties identified by Rogers et al. (2021) in both the voltage code and the deep layers of a computational model of the hub-and-spoke theory of semantic representation (see Introduction). Figure 3.1A shows that the first property, constant decodability, was found to be a property of classifiers trained on all frequencies and classifiers trained on individual frequency ranges, at least until the average onset of naming (1190 ms in the subset of patients analysed by Rogers et al., 2021).

The second property we investigated was local temporal generalisation (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; King & Dehaene, 2014; Rogers et al., 2021). Figure 3.2A shows the generalisation profile of classifiers trained on power at all frequencies compared to classifiers trained on voltage. The same properties were evident in both – after the initial rise

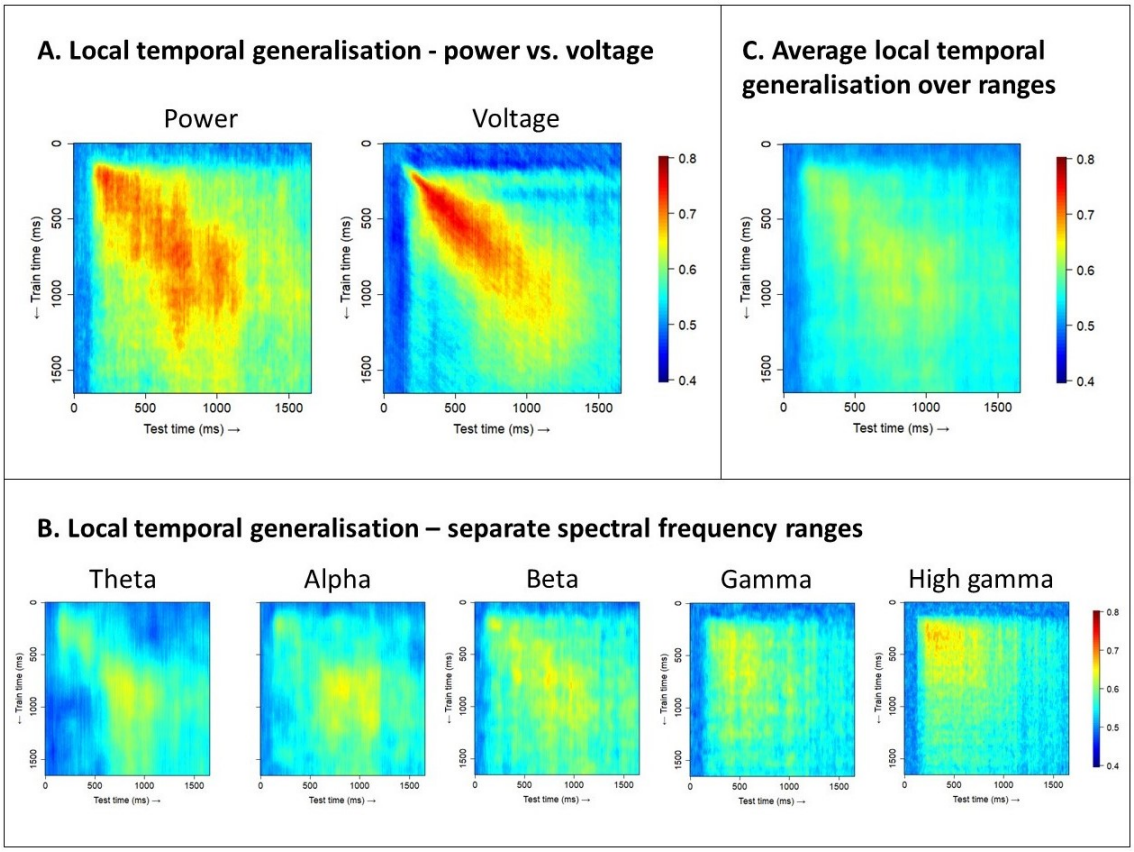

Figure 3.2: Local temporal generalisation. Plots show mean accuracy across patients for classifiers trained at each timepoint (rows) tested at every possible timepoint (columns). Warmer colours indicate higher classification accuracy (see colourbar). (A) Mean hold-out accuracy for classifiers trained on power frequency features for all frequencies between 4 and 200 Hz, or on voltage features. (B) Mean hold-out accuracy for classifiers trained on power frequency features from a single range – theta (4 – 7 Hz), alpha (12 – 18 Hz), beta (13 – 30 Hz), gamma (30 – 60 Hz) and high gamma (42 – 200 Hz). (C) Mean accuracy over frequency ranges (element-wise mean of the five matrices shown in B).

in decoding accuracy, classifiers performed above chance at almost any time window but generalised best to the time windows close to when they were trained. Classifiers trained on power within a single frequency range (Figure 3.2B) showed only one of these characteristics – after the timepoint with the best decoding accuracy, classifiers generalised to almost any other time window, but there was little evidence that classifiers generalised best to time windows close to when they were trained (statistics supporting this claim are reported in Appendix B, Figure B.3). Figure 3.2C shows the generalisation profile averaged across ranges (i.e., the element-wise average of the generalisation matrices in 3.2B). This generalisation profile resembled the generalisation profile of individual frequency ranges – classifiers tested close to the training window generalised little better than those tested further away. Linear combination of the

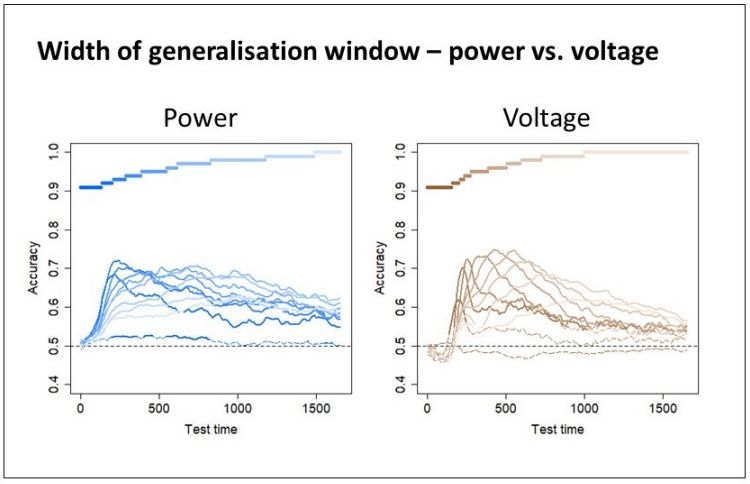

Figure 3.3: Width of generalisation window. Classifiers trained on power frequency features for all frequencies between 4 and 200 Hz (shades of blue) or on voltage features (shades of brown) were grouped into ten clusters via agglomerative hierarchical clustering (see Methods). The "timecourses" show the mean hold-out accuracy for classifiers within each cluster at each timepoint. Lines are solid where there was a significant difference between classifier accuracy and chance (0.5, one-sample *t*-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$) and dashed where there is no significant difference. Coloured bars show the grouped timepoints in each cluster.

accuracy achieved by each range was not sufficient for local temporal generalisation – rather, this property emerged when classifiers were able to assign different coefficients to different frequencies during training. This result illustrates that the code for animacy appears to be truly transfrequency – only when a classifier is able to assign coefficients to frequencies in multiple frequency ranges simultaneously do the deep, distributed, dynamic properties identified by Rogers et al. (2021) become apparent.

The third property we tested for was a widening generalisation window. We clustered the classifiers, trained on power frequency features for all frequencies or on voltage features, into ten clusters based on similar decoding profiles (see Methods). Figure 3.3 shows the performance of each cluster over time. Classifiers trained earlier showed a sharper rise and fall in accuracy, whereas classifiers trained later exhibited a wider window of generalisation (similar plots for classifiers trained on separate frequency ranges are shown in Appendix B, Figure B.4; statistics supporting this claim are reported in Appendix B, Figure B.5).

The fourth and final property tested was the change in code direction. We plotted the classifier coefficients at each timepoint onto the cortical surface. Figure 3.4 shows the sign of the coefficients in the left hemisphere for classifiers trained on power frequency features for all frequencies or on voltage features. We also animated the coefficients – videos for classifiers

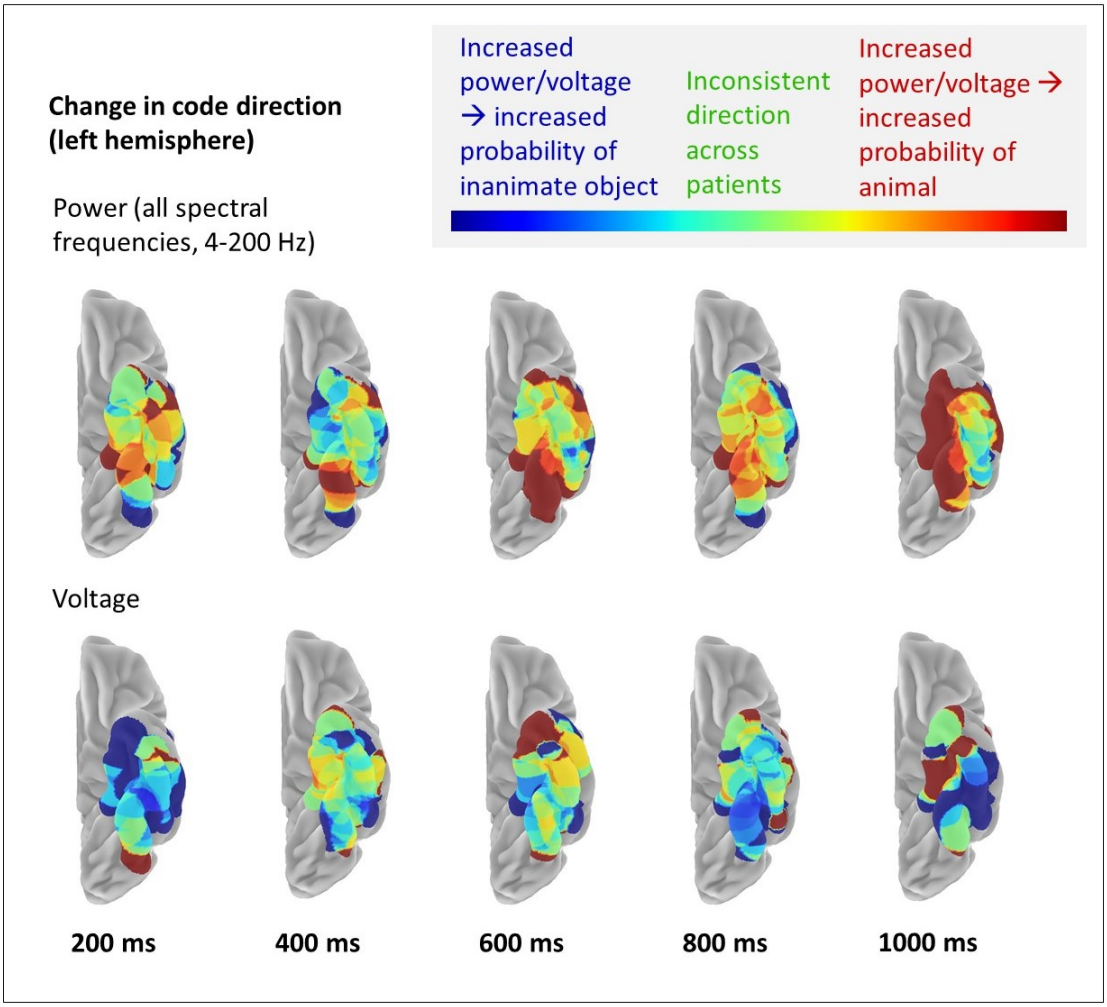

Figure 3.4: Change in code direction. Coefficients for classifiers trained on power frequency features for all frequencies between 4 and 200 Hz, or on voltage features, are projected onto the ventral surface of the left hemisphere. Time is shown relative to stimulus onset. Colours indicate the proportion of classifier coefficients at that vertex with the same sign (see Methods). Warm colours indicate that coefficients are mostly negative across patients (since animals were coded as 0 and inanimate objects as 1, a negative coefficient indicates that an increase in power is associated with increased probability that the stimulus is animate); cool colours indicate that coefficients are mostly positive across patients (an increase in power is associated with increased probability that the stimulus is inanimate); green shades indicate that the region was selected in multiple patients but that the sign of the coefficient is not consistent across patients.

trained on all frequencies, individual frequency ranges and voltage are available at https://github.com/slfrisby/ECoG_LASSO/. Changes in code direction were observed – it was evident that a given location might receive a negative coefficient (since animals were coded as 0 and inanimate objects as 1, a negative coefficient indicates that an increase in power is associated with increased probability that the stimulus is animate) at one timepoint and a positive coefficient (indicating that an increase in power is associated with increased probability that the

stimulus is inanimate) soon afterwards. Nonzero coefficients were distributed in space across the vATL surface and change dynamically over time.

3.3 Discussion

Our understanding of *how* the brain represents semantic information lags behind our understanding of *where* that information is represented (Lambon Ralph et al., 2017). In this work we address a crucial component of this account for the first time – how semantic information may be represented in the time-frequency power and/or phase of neural activity. A large task-based ECoG grid electrode dataset provided us with a rare opportunity to explore this; by applying regularised regression to these data we revealed a transfrequency code for semantic information (in this case animacy) in the vATL.

Our first question was whether it is possible to decode information from time-frequency power and/or phase and, if so, from which frequency ranges. We found that the best decoding performance was achieved when time-frequency power from all frequency ranges – theta, alpha, beta, gamma and high gamma – was provided as input to the classifier simultaneously (Rupp et al., 2017). In this case, performance was comparable to decoding voltage (Rogers et al., 2021). It was possible to decode animacy from power in every frequency range (Lam et al., 2016), but accuracy was not as high as it was when decoding from all frequency ranges. It has been proposed that different frequency bands may play different roles in semantic representation – abstraction from multiple episodic experiences (Sauseng & Klimesch, 2008), binding multiple features of an item (Merker, 2013; Sauseng & Klimesch, 2008; Stryker, 1989; Tallon-Baudry & Bertrand, 1999; von der Malsburg, 1995), binding features that may be distributed across the cortex (Marko et al., 2019), encoding semantic similarity (Solomon et al., 2019; Spiers, 2020), and retrieval independent of sensory modality (Sauseng & Klimesch, 2008) in a way that is appropriate for the context (Halgren et al., 2015; Jackson, 2021; Klimesch, 2012; Marko et al., 2019). The temporal decoding profiles for each frequency range were nearly identical, which does not support the claim that different ranges encode different components of a representation (frequency "multiplexing"; Watrous, Fell, et al., 2015). One possible explanation for the similarity between ranges is that, since the brain exhibits properties of a dynamical system (Jones et al., 2015; Yuste, 2015), oscillations may be an epiphenomenon of normal functioning or even of epileptiform activity (Drane et al., 2008; Hamberger, 2007; Henry et al., 1998). Therefore, frequency ranges

may represent information not in a "strong" sense (meaning that the brain uses time-frequency power to represent information for downstream computation) but in a "weak" sense (meaning that time-frequency power is decodable but is not used as the principal means of representation; Watrous, Fell, et al., 2015). A second possible explanation is that the same information is represented in each frequency range, but activity in different frequencies reflects the transmission of the same information over different cortical distances (Canolty & Knight, 2010; Lachaux et al., 2012) – a feature that is necessary for a hub-and-spoke semantic representation (Lambon Ralph et al., 2017). One conclusion is clear: a full theory of semantic representation must account for theta, alpha, beta, gamma and high gamma power. Focusing only on one or two ranges of interest will not enable progress towards an understanding of the semantic code.

Our second question was whether a time-frequency power code exhibits the same deep, distributed, dynamic properties observed by Rogers et al. (2021) in the sensor voltages and also in the internal activation patterns of the hub-and-spoke computational model. We found that, though classifiers trained on a single frequency range exhibit changes in code direction (Rogers et al., 2021) and generalise well to other timepoints (Cichy et al., 2014; King & Dehaene, 2014), only classifiers trained on the full frequency spectrum showed the truly *local* pattern of generalisation observed in voltage and in neural network models including the hub-and-spoke model (Rogers et al., 2021; see also Cichy et al., 2016). This pattern was not observed when results from individual ranges were linearly combined – rather, it depended on the classifier's ability to assign coefficients to a combination of different frequencies during training. Taken together with the finding that classifiers trained on multiple frequency ranges outperform those trained on a single range, an important conclusion can be drawn – semantic information (animacy) is not simply present redundantly in multiple frequencies (Rupp et al., 2017; Wang et al., 2011), but is represented in such a way that multiple frequencies must be considered together for the full set of properties to be revealed. If this conclusion is correct, then to speak of a transfrequency code and a voltage code is simply to describe the same phenomenon in two different ways (cf. Miller, 2010).

We were unsuccessful in decoding semantic information from time-frequency phase in any frequency range, apart from a brief period of significant decoding using phase in the theta range between 0 and 500 ms post stimulus onset. Any successful decoding with time-frequency phase poses questions of interpretability. If gamma and/or high gamma power and theta phase could be used to decode, simultaneously, this might suggest that theta phase "weakly" represents

information (Watrous, Fell, et al., 2015) by virtue of its coupling with gamma power. However, though we did not conduct a formal test of cross-frequency coupling, it was clear that gamma and high gamma power and theta phase produced significant decoding at different points in time (cf. Clarke et al., 2018; Kraskov et al., 2007; Mollo et al., 2017; Sato et al., 2021; Watrous, Deuker, et al., 2015). Alternatively, some may argue that this is too early for a semantic effect (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980) and suppose that this period of theta phase decoding reflects some kind of sensory process. We were careful to ensure that significant decoding could not be achieved given low-level visual properties of the stimulus alone (Rogers et al., 2021) – however, it is possible that the neural representation of visual information is correlated with animacy, even if the physical properties of the input itself are not. A third possibility is that this activity is not too early for semantic processing – semantic information representation can begin in computational models after only a few time-ticks (Rogers et al., 2021) – and that theta phase really does represent semantic information, although a mechanism by which phase might represent the complex semantic properties of a near-infinite number of concepts is not intuitively obvious.

To summarise, we have shown that it is possible to decode semantic information from every frequency range between theta and high gamma. Although individual frequency ranges contained enough information to drive decoding above chance, decoding from the whole frequency spectrum at once produced the highest decoding accuracy, equivalent to using voltage. The transfrequency semantic decoding revealed the same set of deep, distributed, dynamic properties observed in voltage and, previously found in the deeper layers of the hub-and-spoke computational model. This work demonstrated that information is represented in the vATL in a transfrequency fashion and allows us to take another important step on the journey to characterise *how* – not simply where – semantic information is represented in the brain.

3.4 Methods

3.4.1 Patients

19 patients participated in the study. All patients were native speakers of Japanese. Information about patients' age, sex, handedness, and clinical presentation is summarised in Table 3.1. Each patient was implanted with subdural grid electrodes for presurgical monitoring (mean 91

	Patient 01	Patient 02	Patient 03	Patient 04
Age, sex, handedness	22, M, R	29, M, R&L	17, F, R	38, F, R
WAIS-R/WAIS-III* (VIQ, PIQ, TIQ)	70, 78, 69	72, 78, 72	67, 76, 69	84, 97, 89
WMS-R (verbal, visual, general, attention, delayed recall)	99, 64, 87, 91, 82	99, 92, 97, 87, 83	51,<50,<50, 81, 56	75, 111, 83, 62, 53
WAB	95.6	96	97.2	98.5
WADA	Left	Bilateral	Left	Left
Age of seizure onset	16	10	12	29
Seizure type	FAS→FIAS, FBTCS	FAS→FIAS	FAS→FIAS	FAS→FIAS
Ictal ECoG onset	aMTG	PHG	PHG	PHG
MRI	L basal frontal cortical dysplasia, L anterior temporal arachnoid cyst	L posterior temporal cortical atrophy	L temporal tip arachnoid cyst	L HS/HA
Pathology	FCD type I	FCD type IIIa	Palmini FCD type IB	FCD type IIa
	Patient 05	Patient 06	Patient 07	Patient 08
Age, sex, handedness	55, M, R	34, M, L	41, F, R	27, F, R
WAIS-R/WAIS-III* (VIQ, PIQ, TIQ)	105, 99, 103	55, **, 44	72, 83, 75	106, **, 105
WMS-R (verbal, visual, general, attention, delayed recall)	71, 117, 84, 109, 72	52,<50,<50, 55,<50	83, 111, 89, 94, 82	112, 114, 114, 81, 100
WAB	98	88	97.3	99.6
WADA	Left	Left	Right	Left
Age of seizure onset	55	12	19	16
Seizure type	FIAS (once)	FAS→FIAS	FAS→FIAS	FAS→FIAS
Ictal ECoG onset	None	R parietal lobe/pMTG	PHG	ventral anterior temporal

MRI	L medial temporal lobe low-grade glioma	R parietal cerebral atrophy & contusion, R hippocampal sclerosis/atrophy	L HS/HA, L parieto-occipital perinatal infarction	R medial temporal cyst
Pathology	Diffuse astrocytoma	Post-traumatic change(parietal)/ scar(temporal)/ HS	FCD type I	FCD type I
	Patient 09	Patient 10	Patient 11	Patient 12
Age, sex, handedness	51, M, R	38, F, R	29, F, R	40, M, L
WAIS-R/WAIS-III* (VIQ, PIQ, TIQ)	73, 97, 83	109, 115, 112	62, 80, 67	93, 105, 98
WMS-R (verbal, visual, general, attention, delayed recall)	80, 101, 85, 91, 91	71, 79, 70, 90, 58	64, 94, 68, 79, 79	74, 94, 77, 110, 96
WAB	89.6	96.9	95.8	99
WADA	Left	Left	Left	Right
Age of seizure onset	43	28	12	6
Seizure type	FIAS	FAS→FIAS	FAS→FIAS	FAS→FIAS
Ictal ECoG onset	mITG	SMG	PHG	PHG
MRI	L temporal cavernoma	L parietal operculum tumour	L HS	R HS/HA
Pathology	Arteriovenous malformation	Oligoastrocytoma	non-neoplastic brain tissue	FCD type IIIa
	Patient 13	Patient 14	Patient 15	Patient 17
Age, sex, handedness	22, M, R	42, M, R	35, M, R	50, F, R
WAIS-R/WAIS-III* (VIQ, PIQ, TIQ)	86, 79, 81	96, 84, 90	82, 86, 82	66, 80, 70
WMS-R (verbal, visual, general, attention, delayed recall)	55, 79, 53, 90, 54	73, 85, 73, 103, 76	75, 89, 75, 92, 81	54, 100, 63, 66, 74
WAB	97.4	99.9	99.2	92.6
WADA	Left	Left	Left	Left
Age of seizure onset	14	27	20	10
Seizure type	FAS→FIAS	FAS→FIAS	FIAS	FIAS
Ictal ECoG onset	PHG	PHG	PHG	aMTG/ITG

MRI Pathology	L HS/HA FCD type IIIa	L HS FCD type IIIa	L HS/HA FCD type IIIa	L HS HS + unknown aetiology
	Patient 20	Patient 21	Patient 22	
Age, sex, handedness	23, F, R	40, M, R	28, F, R	
WAIS-R/WAIS-III* (VIQ, PIQ, TIQ)	67, 82, 71	86, 97, 90	80, 69, 72	
WMS-R (verbal, visual, general, attention, delayed recall)	78, 119, 86, 82, 64	111, 113, 113, 87, 109	77, 89, 77, 121, 73	
WAB	94.2	92.2	100	
WADA	Bilateral	Left	Left	
Age of seizure onset	15	30	12	
Seizure type	FIAS	FIAS	FIAS	
Ictal ECoG onset	IPL/SMG	aMTG	PHG	
MRI	No apparent lesion	No apparent lesion	L HS/HA	
Pathology	FCD type I	FCD type I	FCD type IIIa	

Table 3.1: Patient characteristics. Abbreviations: WAIS-R – Wechsler Adult Intelligence Scale (1991), WAIS-III – Wechsler Adult Intelligence Scale (1997), VIQ – Verbal IQ, PIQ – Performance IQ, TIQ – full-scale IQ, WMS-R – Wechsler Memory Scale (1987), FAS – focal aware seizure, FIAS – focal impaired awareness seizure, FBTCS – focal to bilateral tonic-clonic seizure, aMTG – anterior middle temporal gyrus, PHG – parahippocampal gyrus, pMTG – posterior middle temporal gyrus, mITG – medial inferior temporal gyrus, SMG – supramarginal gyrus, ITG – inferior temporal gyrus, IPL – intraparietal lobule, HS – hippocampal sclerosis, HA – hippocampal atrophy, FCD – focal cortical dysplasia. * – The WAIS-R was used to test patients 01-06 and the WAIS-III was used to test other patients. ** – missing score.

electrodes, range 56 – 128 electrodes per patient). 16 patients had electrodes implanted in the left hemisphere, of which 10 – 52 electrodes (mean 27 electrodes) covered the ventral ATL. In the remaining three patients, with electrodes implanted in the right hemisphere, 6 – 28 electrodes (mean 20 electrodes) covered the ventral ATL. The electrodes were platinum, with a recording diameter of 2.3 mm and an inter-electrode distance of 1 cm (ADTECH, WI). Three patients were also implanted with depth electrodes (sEEG), but these were not included in the analysis. All patients gave written informed consent and the study was approved by the ethics committee of the Kyoto University Graduate School of Medicine.

3.4.2 Stimuli and task

Stimuli were the same 100 line drawings used previously (Y. Chen et al., 2016; Rogers et al., 2021; Shimotake et al., 2015) – 50 animals and 50 nonliving items including buildings, tools, musical instruments and other household objects (Morrison et al., 1997; https://github.com/slfrisby/ECoG_LASSO/tree/main/data_info/stimuli/). There were no significant differences between the categories with respect to age of acquisition, visual complexity, familiarity, word frequency, name agreement and non-semantic visual structure (Barry et al., 1997; Rogers et al., 2021). MATLAB r2010a was used to display stimuli on a PC screen.

3.4.3 Data acquisition

There were four runs per patient, collected in a single session. Within each run, each stimulus was presented once in a random order. Stimuli were presented for 5 seconds each, with no break between stimuli.

Patients were instructed to name each picture as quickly and accurately as possible. Data for nine patients were recorded at 2000 Hz (with a low-pass filter of 600 Hz) and data for ten patients were recorded at 1000 Hz (with a low-pass filter of 300 Hz). Time of naming onset was measured. Responses and eye fixation were monitored via video.

3.4.4 Data analysis

3.4.4.1 Preprocessing – structural MRI

A clinical MPRAGE T_1-weighted anatomical scan was acquired before and after electrode implantation. The location of each electrode was identified on each 2D slice of the post-surgical scan. *fnirt* in FSL (Jenkinson et al., 2012; S. M. Smith et al., 2004; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) was used to coregister the electrode positions to the pre-surgical scan and then to MNI space (MNI152NLin2009cAsym). The position of each electrode was then manually adjusted to the surface.

3.4.4.2 Preprocessing – ECoG

Preprocessing was performed using in-house MATLAB (r2023b) code, composed of functions from EEGLAB (2023.1; Delorme & Makeig, 2004; https://eeglab.org) and available at

https://github.com/slfrisby/ECoG_LASSO/ . First, the CleanLine EEGlab plugin (v2.0; Delorme, 2023; Mitra & Bokil, 2007; https://github.com/sccn/cleanline/) was used to remove line noise at 60 Hz and the harmonics 120 and 180 Hz, without scanning for exact line noise frequencies and with a sliding window step size of 2 (all other parameters were set as defaults). CleanLine uses a multitaper fast Fourier transform to identify and remove sinusoidal noise and, though it removes noise less completely than traditional notch-filtering, is less disruptive to the overall frequency structure of the data. Data were then filtered using *eegfilt* (EEGLAB). The low cutoff was set to 0.5 Hz (to remove slow drifts; Delorme, 2023) and the high cutoff set to 300 Hz (for consistency across patients – some data were recorded at 1000 Hz using equipment that imposed a low-pass filter at 300 Hz). Channels that lay below the seizure onset zone (according to clinicians' reports) or with poor contact (identified by visual inspection) were removed. Data were epoched between –1000 and 3000 ms relative to stimulus onset and baseline-corrected using the mean response across trials between –200 and –1 ms. Data from the nine patients recorded at 2000 Hz was then downsampled to match the ten patients recorded at 1000 Hz by boxcar averaging pairs of neighbouring timepoints. A common average reference was applied using *reref* (EEGLAB).

Addressing artefacts in the data was a five-step process. First, the mean voltage for each channel was calculated over timepoints and trials. Any trial that contained a value more extreme than 10 standard deviations from the mean was rejected. Second, all trials were inspected individually and any that appeared to contain obvious interictal epileptiform activity, muscle activity, "electrode pop" or other artefacts were rejected. Trials that appeared to contain wicket rhythms were retained (Kang & Krauss, 2019). Since repeated presentations of the same stimulus were to be averaged it was important to check whether there was at least one good trial for each stimulus. One patient lacked any good trials for 18 stimuli (7 animate, 11 inanimate) and so was excluded from further analysis. Another lacked any good trials for three stimuli (one animate, two inanimate) – because of the small and relatively balanced number of missing stimuli, the decision was made to include this patient. For the next steps an independent component analysis (ICA) was conducted using *runica* (EEGLAB) extracting n components, where n was 75 % of the total number of good electrodes (Clarke, 2020). For the third artefact rejection step, the SASICA EEGLAB plugin (Chaumon et al., 2015; https://github.com/dnacombo/SASICA/) was used to identify ICA components with weak autocorrelation, which is characteristic of muscle activity. Components identified by SASICA were inspected and, in order to minimise rejection of signal-carrying data, rejected if the channel appeared to contain only autocorrelated noise and

no neural activity. Fourth, SASICA was used to highlight components with focal trial activity (i.e., activity in only a few trials), characteristic of muscle activity, "electrode pop" or other artefacts. Rather than rejecting the component across all trials, the atypical trials were identified via visual inspection and only those trials were removed. The ICA was then re-run and the process repeated until no components contained significant focal trial activity. Fifth, the possibility of microsaccade artefacts in the data was assessed (Clarke, 2020; Jerbi et al., 2009; Kovach et al., 2011; Yuval-Greenberg et al., 2008). The independent component activations were filtered for activity in the frequency range associated with microsaccades (20 – 190 Hz). The activations were then convolved with a saccade-related potential template and the number of microsaccades per second was calculated (Craddock et al., 2016). Since every component contained very few microsaccades (<0.000012 microsaccades per second) and no component obviously contained more microsaccades than any other, no components were rejected based on this assessment.

Time-frequency power and phase were extracted using complex Morlet wavelet convolution (Bertrand et al., 1994), implemented with *timefreq* (EEGLAB). Power values were extracted from every trial between 0 and 1650 ms in 10 ms time steps and between 4 and 200 Hz in 60 logarithmically-spaced frequency steps. A 5-cycle wavelet was used at 4 Hz, increasing to a 15-cycle wavelet at 200 Hz (Clarke, 2020). Power values were then averaged across repeated presentations of the same stimulus and any missing trials were interpolated with the median power value across all trials. Decibel normalisation was performed using the mean power across trials between –300 and –100 ms (calculated separately for each electrode and frequency) – a gap before stimulus onset was used to mitigate the effect of temporal leakage of trial activity into the baseline during wavelet convolution. Since phase values cannot be averaged (M. X. Cohen, 2014), preprocessed trials were averaged across repeated presentations of the same stimulus before phase values were extracted using the same time and frequency steps used for the extraction of power. As a comparison, preprocessed voltage was averaged over repeated presentations of the same stimulus.

The impact of preprocessing decisions on decoding results is explored in Appendix C. In summary, decoding of semantic structure (animacy) with spectral frequency features remains unchanged after this extensive and careful cleaning.

3.4.4.3 Multivariate classification

3.4.4.3.1 Decoding approach

Logistic regression classifiers were trained to discriminate animate from inanimate stimuli using *glmnet* in R (Friedman et al., 2010). For power and phase, frequency feature vectors were created for each stimulus at each timepoint (0 ms, 10 ms, 20 ms, …) by concatenating power or phase values for all electrodes for all frequencies in a range of interest. Voltage feature vectors were created for each stimulus at each timepoint by concatenating voltage values for all electrodes in a 50 ms time window centred on the timepoint of interest. The voltage feature vectors therefore reflected activity around the timepoint of interest – but note that so did frequency feature vectors, since Morlet wavelet convolution takes activity at neighbouring timepoints into account. These vectors were provided as input to the classifiers. Each classifier was trained on vectors generated at a single timepoint for a single patient.

The classifiers used LASSO (L1) regularisation (Tibshirani, 1996), which applies a penalty that scales with the sum of the absolute values of the coefficients and thus produces solutions in which many features receive coefficients of zero. This approach was selected because it can be used to assess whether the same electrodes or frequencies are used to represent information at different timepoints. If the information used by a classifier trained at time t is present at time $t \pm n$, the classifier will perform well at time $t \pm n$; other units, that may be in different states at the two timepoints, will receive coefficients of zero and therefore will not affect classifier performance (Rogers et al., 2021).

Classifier accuracy was assessed using ten-fold cross-validation. In each outer loop, ten out of 100 stimuli (five animate, five inanimate) were held out. The remaining 90 stimuli were used to search a range of 100 values (0.2 – 0.002, logarithmically spaced) for the regularisation parameter that resulted in the smallest mean squared error. This process was implemented using *cv.glmnet* (glmnet) with parallel model fitting implemented with *foreach* (https://cran.r-project.org/web/packages/foreach/index.html; all other parameters were set as defaults). A model with the best regularisation parameters was tested on the ten stimuli in the outer loop hold-out set. The same model was tested on the same ten stimuli at all other timepoints. The process was then repeated ten times with different final hold-out sets. This process yielded both a main timecourse of decoding accuracy (i.e., mean accuracy, over folds, of classifiers tested on the time window at which they were trained) and a generalisation matrix in which the y-coordinate of a cell is the time at which the classifier was trained and the

x-coordinate is the time at which the classifier was tested. Finally, a single classifier was trained at every timepoint using all 100 stimuli for training. The accuracy of these classifiers was not evaluated – they were used only for inspection of coefficients.

To assess group-level performance, the timecourses of decoding accuracy and the generalisation matrices were averaged across patients.

3.4.4.3.2 Experimental questions

We first wished to assess whether power, phase and voltage all contained sufficient information to enable decoding. We therefore created both voltage feature vectors and frequency feature vectors of power and phase data (separately) using all 60 frequencies between 4 and 200 Hz and used these as input to classifiers. We compared each group-average timecourse to chance (0.5) using one-tailed, one-sample t-tests. We compared power and voltage, and phase and voltage, using paired t-tests. Probabilities were adjusted to control the false-discovery rate at $\alpha = 0.05$ (Benjamini & Hochberg, 1995). We confirmed that near-identical patterns of results were obtained in the eight patients with left-hemisphere electrodes previously analysed by Rogers et al. (2021), the seven patients with left-hemisphere electrodes analysed for the first time in this work, and the three patients with right-hemisphere electrodes (Appendix B, Figure B.6). Having done so, we combined all patients into a single group for all subsequent analyses.

We next wished to assess whether there were differences in decoding performance between power and phase within different frequency ranges. We therefore divided the 60 frequencies into theta (4 – 7 Hz, 11 frequencies), alpha (12 – 18 Hz, 7 frequencies), beta (13 – 30 Hz, 13 frequencies), gamma (30 – 60 Hz, 10 frequencies) and high gamma (42 – 200 Hz, 19 frequencies) ranges. We constructed frequency vectors of power and phase data using only power or phase at frequencies within each range and used these as input to separate classifiers. We compared each group-average timecourse to chance and then compared each range to the timecourse of decoding using all 60 frequencies.

We then wished to test whether, if either power or phase data contain sufficient information to enable decoding, the power or phase code exhibits the same deep, distributed, dynamic properties identified by Rogers et al. (2021) in both the voltage data (for the first subset of patients) and the computational hub-and-spoke model. We tested for these properties using both frequency vectors constructed using all 60 frequencies and vectors constructed using only frequencies within each range. Constant decodability had already been assessed by one-sample t-tests of each group-average timecourse against chance.

To assess local temporal generalisation we inspected the group-average generalisation matrices generated by testing each classifier at every possible timepoint. To test whether the patterns observed were statistically significant, we identified the best-performing classifier at each timepoint and then conducted paired *t*-tests to assess whether classifiers trained at every other possible timepoint performed equally well. Probabilities were adjusted to control the false-discovery rate at $\alpha = 0.05$ (Benjamini & Hochberg, 1995).

To assess the shape of the generalisation window over time we computed the pairwise cosine distance between each row of the generalisation matrix, then applied agglomerative hierarchical clustering using *hclust* (native R) and cut the tree to create ten clusters. We selected this cluster number to make these results comparable to those of Rogers et al. (2021). We averaged the timecourses of classifiers within each cluster and inspected performance over time (again, probabilities were adjusted to control the false-discovery rate). To quantify the patterns observed we took classifiers trained every 50 ms (i.e., 0 ms, 50 ms, 100 ms …; this ensured that, for classifiers trained on voltage, consecutive 50 ms time windows did not overlap and were therefore independent). We calculated the area under the curve between each timecourse and a horizontal line at chance (0.5) and fitted a piecewise linear regression to the area values using the *segmented* library in R, using *selgmented* with the Bayesian Information Criterion to determine the number and location of breakpoints to produce the best model given the number of free parameters.

To assess and visualise changes in code direction we first took the classifiers trained on all data at each timepoint and then (separately for every electrode, every patient and every timepoint) calculated the mean classifier coefficient within each frequency range (or over all 60 frequencies, or over the 50 ms within each time window in the case of voltage). These mean classifier coefficients were converted to one NIFTI volume per patient and per timepoint, at 2.5 mm isotropic resolution in MNI space (MNI152NLin2009cAsym), using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB r2023b. One patient was excluded from this analysis because no MNI coordinates were available for their electrodes. Other patients were missing MNI coordinates from only a subset of electrodes; we included those patients in the analysis but ignored coefficients on the electrodes with missing coordinates. We projected coefficients in each volume to the pial surface (fsaverage template; Dale et al., 1999; Fischl et al., 1999) using *-volume-to-surface-mapping* with trilinear interpolation and smoothed them on the surface at 6 mm FWHM using *–metric-smoothing* in Connectome Workbench 1.5.0.

We calculated the proportion of negative coefficients that each vertex received (since animals were coded as 0 and inanimate objects as 1, a negative coefficient indicates that an increase in power is associated with increased probability that the stimulus is animate). We plotted these proportions on the pial surface using *plot_surf_stat_map* in *nilearn* implemented in Python 3.9. Finally, we animated the results – ffmpeg 5.1.6 (https://ffmpeg.org) was used to concatenate plots of coefficients at successive timepoints with a frame rate of 10 (ten times slower than real time) and motion interpolation.

Data and code availability: We are unable to share raw data for this study because the patients did not provide informed consent for us to do so. However, matrices containing power, phase and frequency features (columns) for each stimulus (rows) will be made available upon peer review and acceptance. Code is available at https://github.com/slfrisby/ECoG_LASSO/ .

Acknowledgements: This work was supported by an MRC Career Development Award (MR/V031481/1) to A. D. H., and by an Advanced European Research Council (ERC) award (GAP 670428-30 BRAIN2MIND_NEUROCOMP), MRC programme grant (MR/R023883/1), and intramural funding (MC_UU_00005/18) to M. A. L. R. We would like to thank the patients who so selflessly participated in this study.

Competing interests: A. I. belongs to the Department of Epilepsy, Movement Disorders and Physiology, an Industry-Academia Collaboration Course, supported by a grant from Eisai Corporation, Nihon Kohden Corporation, Otsuka Pharmaceutical Co., and UCB Japan Co.

Supplementary material: Supplementary results can be found in Appendix B and supplementary methods can be found in Appendix C.

Chapter 4

Optimising 7T-fMRI for imaging the anterior temporal lobe

Foreword

This Chapter lays a methodological foundation for the Chapter that follows it – an acquisition sequence that improves sensitivity in the vATL while maintaining sensitivity across the rest of the brain is an essential tool for studying semantic representations with 7T-fMRI.

This Chapter is a manuscript in preparation:

Frisby, S. L., Correia, M. M., Zhang, M., Rodgers, C. T., Rogers, T. T., Lambon Ralph, M. A., & Halai, A. D. (in prep.) Optimising 7T-fMRI for imaging the anterior temporal lobe.

I designed this study together with my co-authors. I contributed to the development and implementation of the five acquisition sequences. I recruited and scanned all participants. I developed the preprocessing and analysis pipelines with my co-authors and then conducted all preprocessing and analyses. I wrote the first draft of the manuscript, which was then edited collaboratively.

Abstract

The temporal signal-to-noise ratio (tSNR) of functional magnetic resonance imaging (fMRI) is particularly poor in the ventral anterior temporal lobes (vATLs) because of magnetic field inhomogeneity, a problem that is exacerbated at higher field strengths. In this 7T-fMRI study we compared three methods of improving sensitivity in the vATLs: parallel transmit, which uses multiple transmit elements, controlled independently, to homogenise the B_1 pulse applied to the tissue; multi-echo, which entails collection of multiple volumes at different echo times following

a single radiofrequency pulse; and multiband, in which multiple slices are acquired simultaneously. We found that parallel transmit and multi-echo improved activation magnitude (contrast betas), but only multi-echo improved activation magnitude in the vATLs. Multiband and denoising of multi-echo data with independent component analysis (ICA) both improved activation precision (contrast t-values). Exploratory results suggested that both multi-echo and ICA denoising may also benefit multivariate analyses. To summarise, a multi-echo, multiband sequence provided improvements across multiple sensitivity metrics in the vATLs while maintaining sensitivity across the whole brain and is therefore a versatile choice for studies investigating the functional roles of the vATL at 7T.

Keywords: 7T-fMRI, parallel transmit, multi-echo, multiband, semantic cognition

4.1 Introduction

The temporal signal-to-noise ratio (tSNR) of functional magnetic resonance imaging (fMRI) varies across the brain. The ventral anterior temporal lobes (vATLs), for example, are located next to the air-filled sinuses and so are affected by magnetic field inhomogeneity and the resulting signal dropout and distortions (Halai et al., 2014, 2015, 2024). This makes it challenging to use fMRI to test theories that implicate the vATLs, such as the hub-and-spoke model of semantic cognition (Lambon Ralph et al., 2017). Signal dropout and distortions are more severe at higher field strengths, implying that it may be especially difficult to measure task-related activity in the vATLs with ultra-high-field fMRI (e.g., 7T-fMRI). However, 7T-fMRI also has many advantages – in regions unaffected by magnetic susceptibility, 7T-fMRI offers improved tSNR relative to 3T-fMRI (e.g. Morris et al., 2019), which can be used to reduce voxel size and enable applications such as laminar fMRI (e.g. Koopmans et al., 2011) or to reduce acquisition times and enable shorter scan times for special populations (e.g. patients with neurodegenerative diseases; Cope et al., 2023). 7T-fMRI also benefits from improved spatial specificity (Marques & Norris, 2018). Improving signal homogeneity would allow researchers studying the whole brain (or focusing on regions affected by susceptibility artefacts) to take full advantage of these

benefits.

One possible method for counteracting signal dropout is parallel transmit (pTx), which uses multiple transmit elements, controlled independently, to homogenise the B_1 pulse applied to the tissue (Gras et al., 2019; Le Ster et al., 2019). A recent study by Ding et al. (2022) used pTx to counteract signal dropout in the vATLs. They found that pTx, compared to a standard sequence, resulted in improved tSNR across the brain, particularly in the temporal lobes, but they found no improvement in functional contrast during a semantic association task shown to recruit the vATLs (Jung et al., 2017).

A second method of recovering signal is multi-echo (ME) imaging (Kundu et al., 2017; Poser et al., 2006; Posse, 2012). T2* is known to vary across the brain (Hagberg et al., 2002); in areas of magnetic susceptibility, T2* is particularly short due to increased spin dephasing. A single echo provides sensitivity to a narrow range of T2* values (the echo time (TE) is typically selected to provide the best compromise of sensitivity across the whole brain), whereas using multiple echoes, and combining data from each, increases that range. ME has been shown to improve functional contrast (Poser & Norris, 2009) and spatial specificity (Boyacioğlu et al., 2015) at 7T and 3T (Fernandez et al., 2017; Halai et al., 2024; Kirilina et al., 2016; Lynch et al., 2020). Having multiple echoes also facilitates the separation of signal and noise – signal should decay in a well-characterised way over TEs, whereas noise would not be expected to do so. This principle underpins multi-echo independent component analysis (ME-ICA), via which ICA components that are TE-independent (and thus are likely to be noise rather than BOLD) can be removed (Dipasquale et al., 2017; Kundu et al., 2011, 2013, 2015, 2017). This method may enhance signal detection in areas of susceptibility on top of the advantage offered by ME alone (e.g. Lombardo et al., 2016). ME sequences have some potential disadvantages. For example, multi-echo can lengthen repetition time (TR). In-plane acceleration is frequently needed to achieve a sufficiently short first TE, but reduces tSNR (Yun & Shah, 2017). In turn, a short first TE, combined with hardware constraints, often limits how small voxels can be (cf. Koopmans et al., 2011). Critically, in many previous studies examining the benefits of ME sequences compared to single-echo (SE), the "single-echo" data is extracted from the ME dataset (Amemiya et al., 2019; Bhavsar et al., 2014; Caballero-Gaudes et al., 2019; A. D. Cohen et al., 2017, 2017, 2018; Dipasquale et al., 2017; Evans et al., 2015; Fernandez et al., 2017; Gilmore et al., 2022; Kovářová et al., 2022). This means that the "single-echo" data will inherit suboptimal parameters that are ME-specific, making the comparison unfair.

A third strategy for improving acquisition is multiband (MB) imaging, also known as simultaneous multi-slice, in which multiple slices are acquired simultaneously (Barth et al., 2016; Moeller et al., 2010; Setsompop et al., 2012). This can be used to increase resolution or to multiply the number of volumes collected, thereby multiplying the effective degrees of freedom, resulting in increased power to detect a significant effect. Increasing the number of simultaneously-acquired slices also reduces TR, reducing noise aliasing and counteracting the increase in TR associated with ME (Halai et al., 2024; Puckett et al., 2018). Possible disadvantages of MB include a reduction in tSNR due to increases in *g*-factor effects (Demetriou et al., 2018; Risk et al., 2021; Setsompop et al., 2012) and leakage of signal into the simultaneously-excited slices (Todd et al., 2016).

To summarise, there is a need to evaluate 7T-fMRI sequences to determine which parameters are important for counteracting signal dropout and improving sensitivity in susceptible regions without compromising on signal quality elsewhere. We tested five possible sequences: pTx, plus a 2 × 2 factorial design varying echo and band (standard single-echo single band (SESB), single-echo multiband (SEMB), multi-echo single band (MESB), and multi-echo multiband (MEMB)). We used a semantic judgment task that is known to evoke activity across the semantic network, including areas with and without susceptibility artefacts (Jung et al., 2017). We had two univariate effects of interest – activation magnitude and activation precision (Halai et al., 2024). Activation magnitude is the magnitude of the BOLD signal change, operationalised as the 1st-level beta values extracted from each voxel. We hypothesised that both pTx and ME would recover signal and increase activation magnitude relative to a standard sequence. Activation precision is the reliability of the BOLD signal change, analogous to the functional contrast-to-noise ratio (fCNR) and operationalised as the 1st-level *t*-values extracted from each voxel. We hypothesised that MB sequences, with greater effective degrees of freedom, would increase activation precision relative to single band (SB) sequences. We had two secondary hypotheses: firstly, since ME-ICA denoising removes TE-dependent noise, greater activation precision would be observed in multi-echo denoised data (MEdn) relative to ME data without denoising; secondly, that an MB advantage would be due to the increase in the number of volumes (which we tested by temporally downsampling the MB data to match the SB data). Our study focused primarily on univariate effects. However, multivariate analysis techniques, which exploit variance and covariance between voxels and can accommodate participant-specific differences in activation patterns (Coutanche, 2013; Davis et al., 2014; Davis & Poldrack, 2013)

are rapidly gaining popularity. These methods frequently rely on the assumption of good-quality signal across the whole brain (Frisby et al., Chapter 2) and so we conducted an exploratory analysis, following the method of Haxby et al. (2001), to decode task condition. Finally, we tested for the presence of slice leakage artefacts in our MB data.

4.2 Materials and methods

4.2.1 Participants

20 healthy native speakers of British English (age range 18 – 50, mean age 33.45 years, 12 female, 8 male) participated in the study. All were right-handed, had normal or corrected-to-normal vision, and had no neurological or sensory disorders. All participants gave written informed consent and the research was approved by a local National Health Service (NHS) ethics committee.

4.2.2 Stimuli and task

All participants performed a semantic association task and a pattern matching task (hereafter the control task) adapted from a study by Jung et al. (2017). Each stimulus consisted of three pictures presented simultaneously (Figure 4.1). Some pictures were line drawings taken from the Pyramids and Palm Trees Test (Howard & Patterson, 1992) and some were colour cartoons or photographs taken from the Camel and Cactus Test (Bozeat et al., 2000). In the semantic task, participants were instructed to indicate which of the two pictures at the bottom of the screen "matched" (i.e. had the closest semantic relationship to) the picture at the top (hereafter the probe picture). In the control task, participants were instructed to indicate which of two scrambled pictures (generated from the pictures used in the semantic task) at the bottom of the screen was identical to a scrambled probe. There were 248 unique picture triplets, so some stimuli were repeated in different runs, but no stimulus was repeated within a run. E-Prime software (Psychology Software Tools Inc., Pittsburgh, USA) was used to display stimuli and record responses. Stimuli for both tasks were rear-projected onto a screen at the back of the MRI scanner, and observers viewed stimuli through a mirror mounted to the head coil directly above the eyes.

The study had a block design with three types of block – semantic, control, and rest. Each task block consisted of four trials. Each trial consisted of a fixation cross presented for 500

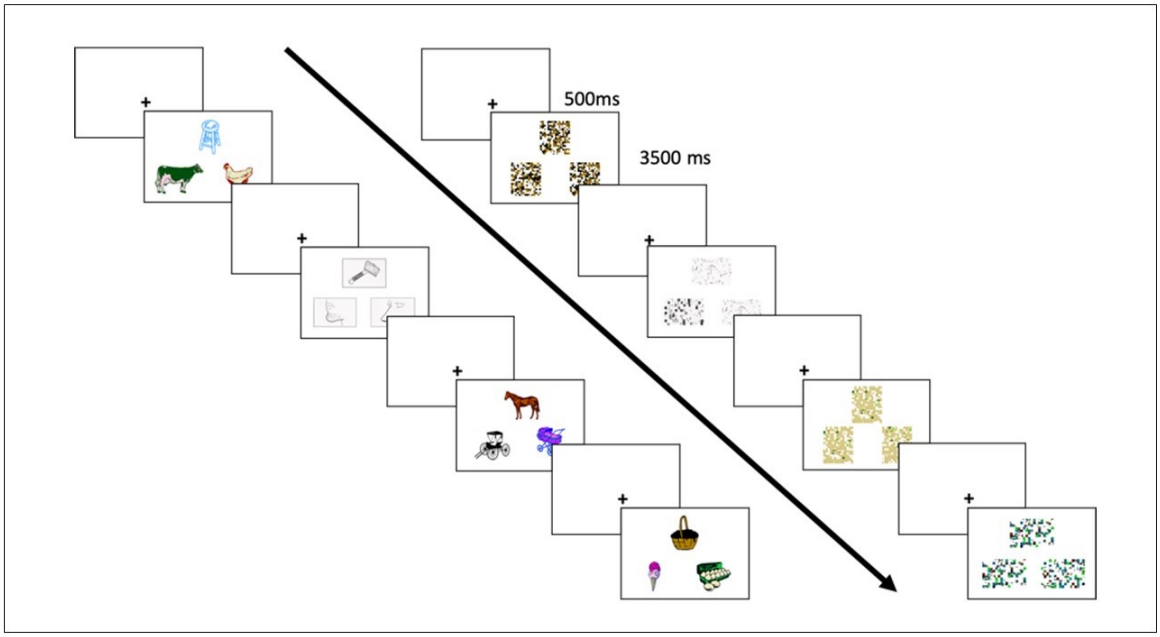

Figure 4.1: One block of the semantic association task (left of the arrow) and one block of the pattern matching (control) task (right of the arrow). Each trial consisted of a fixation cross (500 ms) followed by three pictures presented simultaneously. Participants had to select which of the two pictures at the bottom was more semantically similar to (in the semantic task) or visually identical to (in the control task) the picture at the top (the probe picture). Each block lasted 16 s in total. Reprinted with permission from Halai et al. (2024).

ms followed by a stimulus presented for 3500 ms. Each rest block consisted of a fixation cross presented for 16 s. Each run began 16 s after the start of the MR sequence and then consisted of 30 blocks presented in the order semantic, control, semantic, control, rest. There were five runs per participant, collected in a single session. Accuracy and reaction time were measured for each trial; since reaction time is not normally distributed, both metrics were compared across tasks using Wilcoxon's signed-rank tests and across sequences using Friedman's nonparametric ANOVA.

4.2.3 Image acquisition

All images were acquired on a whole-body 7T Siemens MAGNETOM Terra MRI scanner (Siemens Healthcare, Erlangen, Germany) with an 8Tx32Rx head coil (Nova Medical, USA).

An MP2RAGE anatomical scan was acquired with the following parameters: 224 sagittal slices (interleaved acquisition), FOV 240 × 225.12 × 240 mm, voxel size 0.75 × 0.75 × 0.75 mm, TR 4300 ms, inversion times 840 and 2370 ms, TE 1.99 ms, flip angles 5° and 6°, GRAPPA acceleration factor 3.

Next, manual B_0-shimming was performed over the volume to be used for echo-planar imaging (EPI). The aim was to reduce the water linewidth, defined as full width of the spectrum at half height, to below 40 Hz. However, for some participants the adjustments proved time-consuming and so adjustment time was capped at 30 minutes from the start of scanning after which the best shim parameters were adopted. The actual average water linewidth was 42.21 Hz (standard deviation = 9.02 Hz; missing data for 4 participants). Then a dummy pTx-EPI scan of one volume was acquired for the offline calculation of the variable-rate selective excitation (VERSE) pulses (Hargreaves et al., 2004) for the pTx sequence only.

There were five functional runs of EPI, one run of each sequence. The parameters of each sequence are given in Table 4.1. The order of sequences was counterbalanced across participants. The following parameters were held constant across sequences: 48 axial slices (interleaved acquisition), FOV 210 × 210 × 210 mm to cover the whole brain in most participants (visual inspection ensured that the vATL was included in the FOV for all participants and the FOV was tilted up at the nose to avoid ghosting of the eyes into the vATL), voxel size 2.5 × 2.5 × 2.5 mm (no gap), and A-P phase encoding direction. After each run, 5 further volumes were acquired

	SESB	SEMB	MESB	MEMB	pTx
No. echoes	1	1	3	3	1
Multiband factor	1	2	1	2	1
TR (ms)	3020	1510	3020	1510	3000
TE_1 (ms)	25.00	25.00	11.80	11.80	25.00
TE_2 (ms)	–	–	27.05	27.05	–
TE_3 (ms)	–	–	42.30	42.30	–
iPAT type	Off	Off	GRAPPA	GRAPPA	GRAPPA
iPAT factor	Off	Off	3	3	2
Phase partial Fourier	7/8	7/8	7/8	7/8	Off
Flip angle (°)	50	50	78	63	40
Bandwidth (Hz/Px)	2204	2204	2204	2204	1804
Number of volumes acquired	171	340	171	340	172
Number of pulses	1	1	1	1	5 (VERSE)

Table 4.1: Parameters of each sequence. SESB = single-echo single band, SEMB = single-echo multiband; MESB = multi-echo single band; MEMB = multi-echo multiband, pTx = parallel transmit.

with the phase encoding direction changed to P-A. These volumes were to be used for distortion correction during preprocessing.

4.2.4 Data analysis

All analysis code is available at https://github.com/slfrisby/7TOptimisation/.

4.2.4.1 Preprocessing

For ease of data sharing we converted all DICOMs to BIDS format (Gorgolewski et al., 2016) using *heudiconv* (v1.0.0; Halchenko et al., 2024).

Standard reproducible preprocessing pipelines designed for 3T-fMRI, such as *fMRIprep* (Esteban et al., 2019; Gorgolewski et al., 2011) perform poorly on 7T-fMRI EPI data; therefore, the analysis pipeline was split into stages using different software packages. The MP2RAGE T1w (combined) image first had its background noise removed using O'Brien regularisation (O'Brien et al., 2014) and was then submitted to CAT12 for segmentation (Gaser et al., 2023; in SPM12; https://www.fil.ion.ucl.ac.uk/spm/). The bias-corrected and global-intensity-corrected T1w (combined) image produced was provided to as input to *fMRIPrep* 21.0.1. The T1w (combined) image was corrected for intensity non-uniformity with *N4BiasFieldCorrection* (Tustison et al., 2010) and skull-stripped with a Nipype implementation of the *antsBrainExtraction.sh* workflow (ANTs). Volume-based spatial normalisation to standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with *antsRegistration.sh* (ANTs 2.3.3; Avants et al., 2009, 2011; https://github.com/ANTsX/ANTs/).

Functional preprocessing was performed using in-house code, composed of functions from AFNI (v.18.3.03; R. W. Cox, 1996; R. W. Cox & Hyde, 1997; https://afni.nimh.nih.gov), FSL (v.5.0; Andersson et al., 2003; Jenkinson et al., 2012; S. M. Smith, 2002; S. M. Smith et al., 2004; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/), tedana (v.23.0.1; DuPre et al., 2021; Kundu et al., 2011, 2013; https://tedana.readthedocs.io/en/stable/index.html) and ANTs (v. 2.2.0; Avants et al., 2009, 2011; https://github.com/ANTsX/ANTs/). EPIs were despiked using *3dDespike* (AFNI), slice-time corrected to the middle slice using *3dTshift* (AFNI), motion corrected with *3dvolreg* and *3dAllineate* (AFNI) using the first volume of each run as a reference (for ME datasets, the TE_1 was aligned and subsequent transforms were applied to the TE_2 and TE_3), and skull-stripped using *BET* (FSL) to create a participant-specific brain mask.

For ME datasets only, *tedana* was used to create two timeseries – one where the echoes

were optimally combined based on T2* weighting (Posse, 2012), and one in which the T2* optimally-combined data were denoised using ICA. *tedana* conducts denoising by decomposing data using PCA and ICA, classifying components according to whether the signal scales linearly with TE (as BOLD signal does), and reconstructing the data using only BOLD-like components. The brain mask created with *BET* was used as the mask for this stage.

Next, for all datasets, unwarping was conducted using *topup* and *applytopup* (FSL). Field displacement maps were calculated using ten volumes (five with A-P phase encoding direction, extracted from the start of each functional run, and five with P-A phase encoding direction, collected separately after each run) and the resulting correction was applied to all images. Finally, the mean EPI for each run was coregistered to the skull-stripped native structural image using a rigid-body registration with *AntsRegistrationSyN.sh* (ANTs). EPIs were then transformed into standard space (MNI152NLin2009cAsym) by combining the transforms from native EPI to native T1 and the transforms from native T1 to standard space and applying those transforms to the EPIs using *antsApplyTransforms* (ANTs). Images were smoothed with a 6 mm FWHM Gaussian filter in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) for GLM analysis.

A separate functional preprocessing pipeline was used to create images for slice leakage analysis. This pipeline differed from the main pipeline in the following ways. Despiking was omitted to avoid the removal of noise. For ME datasets only, rather than run the full *tedana* workflow, we conducted optimal combination of data from multiple echoes (but no denoising) using the *t2smap* command using all voxels in the volume (i.e. no brain mask). All coregistration steps were omitted (the images remained in native EPI space) and no smoothing was applied.

4.2.4.2 1st-level (within-participant) GLM

Data were analysed using the general linear model (GLM) approach implemented in SPM12 in MATLAB r2019a. We had 5 timeseries of primary interest – standard single-echo single band (SESB), parallel transmit (pTx), single-echo multiband (SEMB), multi-echo single band (MESB) and multi-echo multiband (MEMB). For the latter two sequences, data from all echoes was optimally combined but were not ICA-denoised. We also generated two timeseries with ME-ICA denoising (multi-echo single band with ICA denoising (MESBdn) and multi-echo multiband with ICA denoising (MEMBdn)) and two downsampled MB timeseries created by extracting odd-numbered volumes to match the number of volumes in the single band timeseries (odd-numbered volumes of single-echo multiband (SEMBodd) and odd-volumes of multi-echo

multiband (MEMBodd)).

At the individual subject level, each block of the semantic and control task was modelled as a boxcar function (resting blocks were modelled implicitly) and these boxcar functions were subsequently convolved with SPM's difference of gammas haemodynamic response function. The six motion parameters extracted during preprocessing were used as regressors of no interest. The micro-time resolution was set as the number of slices (n = 48), the micro-time onset was set as the reference slice for slice-time correction (n = 24), and the high-pass filter cutoff was 128 seconds. The same template used during preprocessing (MNI152NLin2009cAsym) was used as a mask for the analysis. The parameter estimation method was restricted maximum likelihood estimation (ReML) and serial correlations were accounted for using an autoregressive AR(1) model during estimation. For univariate analyses the contrast of interest was greater activation for the semantic task than the control task (S>C). For exploratory multivariate pattern analysis (MVPA) each block (12 semantic and 12 control) was modelled individually in order to obtain one beta image per block.

Finally, for the slice leakage analysis, the modelling was rerun on the minimally-preprocessed timeseries without any brain mask. We obtained both univariate contrasts of interest (S>C) and beta images per block for MVPA.

4.2.4.3 2nd-level (across-participant) GLM

4.2.4.3.1 Region of interest (ROI) analysis

Regions of interest (ROIs) were defined based on a large-scale (n = 69) distortion-corrected 3T-fMRI study of the semantic network (G. F. Humphreys et al., 2015). For each comparison, ROIs were included only if they overlapped by at least one voxel with the whole-brain contrast of interest (S>C) summed over all sequences in the comparison.

4.2.4.3.1.1 Univariate analysis

Activation magnitude (1st-level beta values) and activation precision (1st-level t-values) were extracted from each ROI using a publicly-available script, *roi_extract.m* (https://github.com/MRC-CBU/riksneurotools/blob/master/Util/roi_extract.m). There were two planned t-tests for activation magnitude (pTx>SESB, ME>SE) and four planned t-tests for activation precision (MB>SB, MEdn>ME, MBodd>SB, SB>MBodd). For each planned t-test, results were Bonferroni-corrected for the number of ROIs included.

4.2.4.3.1.2 Exploratory multivariate pattern analysis (MVPA)

The input to this analysis was the activation magnitude values extracted from each block individually (12 semantic and 12 control). For each block and each ROI, a vector of beta values within that ROI was created. The cosine dissimilarity between every possible pair of blocks was calculated. MVPA performance was operationalised as the mean between-task dissimilarity minus the mean within-task dissimilarity and paired *t*-tests were used to compare the metric across sequences (all planned *t*-tests described in the univariate ROI analysis were conducted; Haxby et al., 2001).

4.2.4.3.2 Whole-brain analysis

All the above contrasts (activation magnitude: pTx>SESB, ME>SE; activation precision: MB>SB, MEdn>ME, MBodd>SB, SB>MBodd) were assessed at the whole-brain level using *t*-tests (for comparing pTx and SESB) or using random-effects ANOVAs with one-sample *t*-tests on the summary statistic (for the three factorial designs: varying echo and band (SESB, SEMB, MESB and MEMB); varying denoising and band (MESBdn, MEMBdn, MESB and MEMB); and varying echo and downsampled band (SEMBodd, MEMBodd, SEMB and MEMB)). The ANOVAs were conducted using a publicly-available script, *batch_spm_anova.m* (https://github.com/MRC-CBU/riksneurotools/blob/master/SPM/batch_spm_anova.m). The group *t*-maps were assessed for significance by using a voxel-height threshold of $p<0.001$ to define clusters and then a cluster-defining family-wise-error-corrected threshold of $p<0.05$ for statistical inference.

4.2.4.3.3 Slice leakage analysis

The group-level, whole-brain contrast of interest (S>C) in standard space (MNI152NLin2009cAsym) was inspected and the coordinates of the peak *t*-values of the top 5 clusters were identified. These coordinates were back-projected to obtain 5 sets of corresponding coordinates in each participant's native EPI space (hereafter "seeds", labelled A). Next, voxels to which signal might be warped were identified as artefact locations based on phase shift (FOV/2, labelled B) and, in the MEMB data, GRAPPA (labelled Ag) and phase shift plus GRAPPA (labelled Bg). A spherical ROI, 4 voxels in radius, was defined around each seed location and possible artefact location using a modified version of the scripts developed for McNabb et al. (2020; https://github.com/DrMichaelLindner/MAP4SL/; our version available at https://github.com/slfrisby/7TOptimisation/).

Both univariate (McNabb et al., 2020) and multivariate (Halai et al., 2024) slice leakage tests were conducted. Activation magnitude was extracted from minimally-preprocessed data, and, for the multivariate analysis, the difference between mean within-task similarity and mean between-task similarity was calculated as in the ROI analysis. Paired t-tests were conducted for each seed and artefact location between each sequence of interest (SEMB and MEMB) and a control sequence. For artefact locations based on phase shift (B), the control sequence was the corresponding SB sequence (SESB for SEMB and MESB for MEMB). For artefact locations based on GRAPPA, the control sequence was the corresponding SE sequence, because SE sequences were collected without GRAPPA (SEMB for MEMB). Results were Bonferroni-corrected for the number of peaks (n = 5) and the number of possible artefact regions (n = 1 for SEMB and n = 3 for MEMB).

4.3 Results

4.3.1 Excluded participants

Two participants were excluded because of excessive head motion (this was defined by calculating, for each participant, the percentage of volumes per run with absolute translation values of over 2 mm or absolute rotation values over 1°, averaging these percentages over runs, and excluding any participant whose mean percentage was greater than 2 standard deviations above the mean percentage across participants). One participant was excluded because of technical problems during data acquisition which meant that the pTx run failed. All subsequent analyses were conducted on the remaining 17 participants.

4.3.2 Behavioural results

The 17 participants had good performance on both tasks and, importantly, there were no significant differences between the 5 sequences (SESB, pTx, SEMB, MESB, MEMB) in accuracy (Friedman's χ^2 = 5.26, p = 0.26) or reaction time (Friedman's χ^2 = 7.20, p = 0.16). Differences between the tasks did not reach significance for accuracy (Wilcoxon's z = 0; p = 0.06) or reaction time (Wilcoxon's z = 7; p = 1.00).

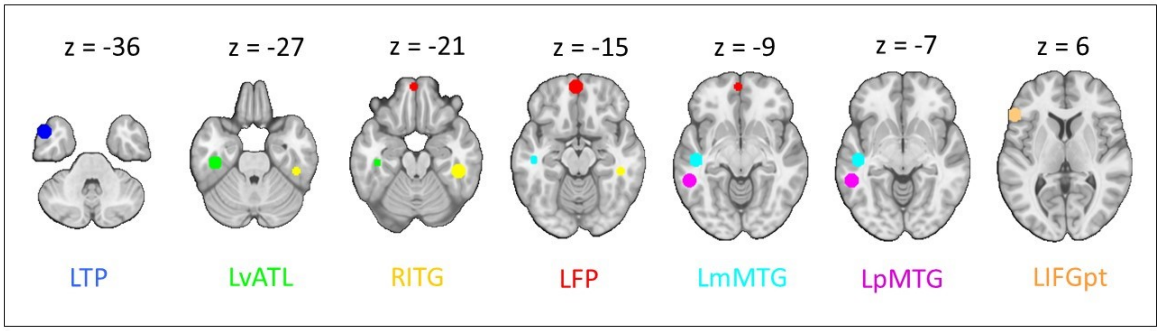

Figure 4.2: Regions of interest, taken from a meta-analysis of semantic tasks by G. F. Humphreys et al. (2015). All spheres are 8 mm in radius. Regions of interest are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym). LTP = left temporal pole; LvATL = left ventral anterior temporal lobe; RITG = right inferior temporal gyrus; LFP = left frontal pole; LmMTG = left medial middle temporal gyrus; LpMTG = left posterior middle temporal gyrus; LIFGpt = left inferior temporal gyrus pars triangularis.

4.3.3 ROI analysis

ROIs that overlapped with the contrast of interest (S>C) for at least one of the five sequences are shown in Figure 4.2.

	LTP	LvATL	RITG	LFP	LmMTG	LpMTG	LIFGpt
Activation magnitude							
pTx>SESB	–	0.2488	0.0240*	–	–	0.0042**	0.4114
ME>SE	0.0502	0.0001**	0.0333*	0.2951	0.0092*	0.0003*	0.4155
Activation precision							
MB>SB	0.4569	0.0007**	0.0019**	0.0989	0.2187	0.0266*	0.0016**
MEdn>ME	0.3117	<0.0001**	<0.0001**	0.0014**	0.1363	<0.0001**	<0.0001**
MBodd>SB	0.4225	0.1057	0.0481*	–	0.5697	0.1879	0.3300
MVPA							
pTx>SESB	–	0.8569	0.3280	–	–	0.1529	0.9179
ME>SE	0.0096*	0.0063**	0.1612	0.0655	0.0465*	0.0024**	0.2508
MB>SB	0.2455	0.1417	0.0667	0.3021	0.3949	0.2693	0.5207
MEdn>ME	0.0027**	<0.0001**	<0.0001**	0.0023**	0.0001**	0.0012**	0.0007**
SB>MB	0.7416	0.4570	0.7222	–	0.3803	0.6716	0.3242
MBodd>SB	0.2584	0.5430	0.2778	–	0.6197	0.3284	0.6758

Table 4.2: p-values for all t-tests within regions of interest (ROIs) based on the semantic network. * = $p<0.05$, ** = $p<0.05$ (Bonferroni-corrected for the number of ROIs), pTx = parallel transmit, SESB = single-echo single band, ME = multi-echo without ICA denoising, SE = single-echo, MB = multiband, SB = single band, MEdn = multi-echo with ICA denoising, MBodd = multiband downsampled, LTP = left temporal pole, LvATL = left ventral anterior temporal lobe, RITG = right inferior temporal gyrus, LFP = left frontal pole, LmMTG = left medial middle temporal gyrus, LpMTG = left posterior middle temporal gyrus, LIFGpt = left inferior temporal gyrus pars triangularis.

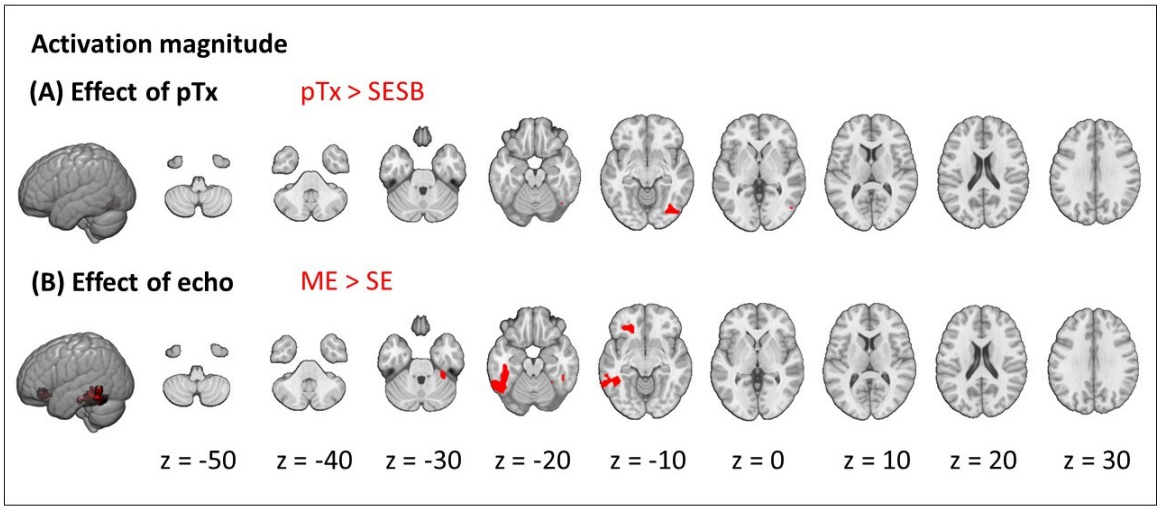

Figure 4.3: Effects on activation magnitude: (A) effect of parallel transmit (pTx>SESB, red); (B) effect of echo (ME>SE, red). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

4.3.3.1 Univariate analysis

Table 4.2 shows the *p*-values for all planned *t*-contrasts. pTx provided significantly better activation magnitude than SESB in the left posterior middle temporal gyrus (LpMTG). ME sequences provided significantly better activation magnitude than SE sequences in the left ventral anterior temporal lobe (LvATL).

MB sequences offered significantly better activation precision than SB sequences in the LvATL, right inferior temporal gyrus (RITG) and left inferior frontal gyrus pars triangularis (LIFGpt). MEdn sequences offered significantly better activation precision than ME in the LvATL, RITG, LpMTG, LIFGpt and left frontal pole (LFP). There was no significant difference in either direction between MBodd sequences and SB sequences.

4.3.3.2 Exploratory MVPA

Table 4.2 also shows the *p*-values for *t*-tests comparing our MVPA metric (mean between-task cosine dissimilarity minus mean within-task cosine dissimilarity) between sequences. The ME sequences produced significantly better performance than SE sequences in the LvATL and LpMTG. MEdn sequences produced better performance than ME sequences in every ROI. All other comparisons failed to reach significance.

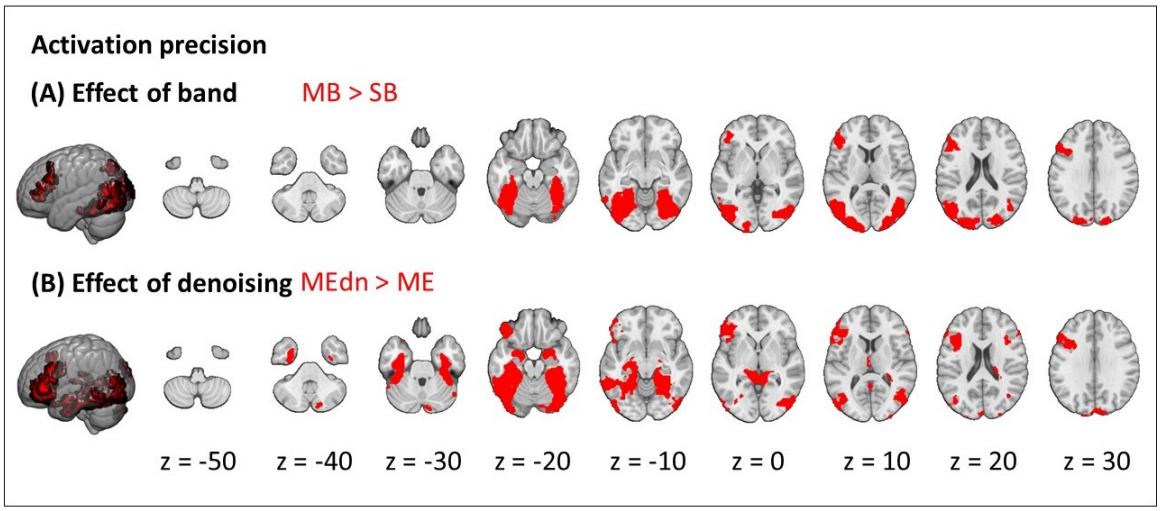

Figure 4.4: Effects on activation precision: (A) effect of multiband (MB>SB, red); (B) effect of ME-ICA denoising (MEdn>ME, red). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

4.3.4 Whole-brain analysis

Figure 4.3 shows the results for activation magnitude. Figure 4.3A shows a single cluster, within the right fusiform gyrus and lateral inferior occipital cortex, that showed greater activation magnitude with pTx than with SESB (pTx>SESB). Figure 4.3B shows the main effect of ME over SE – clusters in the fusiform and inferior temporal gyri bilaterally, plus the left orbitofrontal cortex (cluster and peak information is provided in Appendix D, Table D.1).

Figure 4.4 shows the results for activation precision. Both the main effect of MB over SB (MB>SB) and the main effect of ME-ICA denoising over ME without ICA denoising (MEdn>ME) extended down the temporal lobe and included frontal regions (Appendix D, Table D.1). There was no significant difference between downsampled MB sequences and SB sequences (MBodd>SB or SB>MBodd).

To summarise, these results aligned with results from our ROI analyses – ME (but not pTx) increased activation magnitude and both MB and ME-ICA denoising increased activation precision (main effects of the contrast of interest (S>C) plus all reverse contrasts, are shown in Appendix D, Figures D.1 – D.6).

4.3.5 Slice leakage analysis

Seed and possible artefact locations for an example participant are shown in Figure 4.5A. Figure 4.5B shows activation magnitude for an example seed and its corresponding possible artefact

locations for all participants in each MB sequence and its corresponding sequences. Figure 4.5C shows MVPA performance (violin plots for all other peaks are shown in Appendix D, Figure D.7). *t*-tests were conducted at each location, but no comparison reached statistical significance (*p*-values are shown in Appendix D, Table D.2). We therefore concluded that there was no evidence for slice leakage in either of our MB datasets.

4.4 Discussion

The tSNR of fMRI varies across the brain. This is especially evident in the vATLs, which are located next to the air-filled sinuses and are therefore affected by signal dropout and distortions (Halai et al., 2014, 2015, 2024). This study was the first 7T-fMRI study systematically comparing pTx, ME and MB as methods for improving sensitivity in these regions while maintaining sensitivity across the brain. We found that pTx improved activation magnitude in posterior temporal and occipital regions while ME improved activation magnitude across multiple areas of the semantic network. Both MB and ME-ICA denoising resulted in improved activation precision extending down the temporal lobe and including frontal regions. In an exploratory analysis we found that ME and ME-ICA improved MVPA performance but MB did not. No slice leakage artefacts were associated with our multiband sequences.

Although parallel transmit produced better activation magnitude in posterior temporal and occipital regions than SESB, activation magnitude in the vATL was comparable for both sequences. These results replicated those of Ding et al. (2022) who failed to find improved contrast in anterior temporal regions with pTx (although they did find improved tSNR). Note that, despite visible signal dropout on the EPIs (shown in Appendix D, Figure D.8), both sequences were able to identify semantic activity extending into ventral anterior temporal regions; this finding reflects the increased sensitivity of 7T-fMRI compared to 3T-fMRI (in which semantic activity is rarely observed without multi-echo (Halai et al., 2014, 2015, 2024) or another method of recovering signal, such as spin-echo (Binney et al., 2010; Embleton et al., 2010). That a standard sequence can detect signal in the vATLs might come as a surprise to many neuroimaging researchers as 7T-fMRI is not only associated with increased sensitivity, but also with exacerbated magnetic susceptibility artefacts. However, we acknowledge that our "standard" sequence is not, for example, ultra-high-resolution and so the sequence is not truly "standard" for the field of 7T-fMRI.

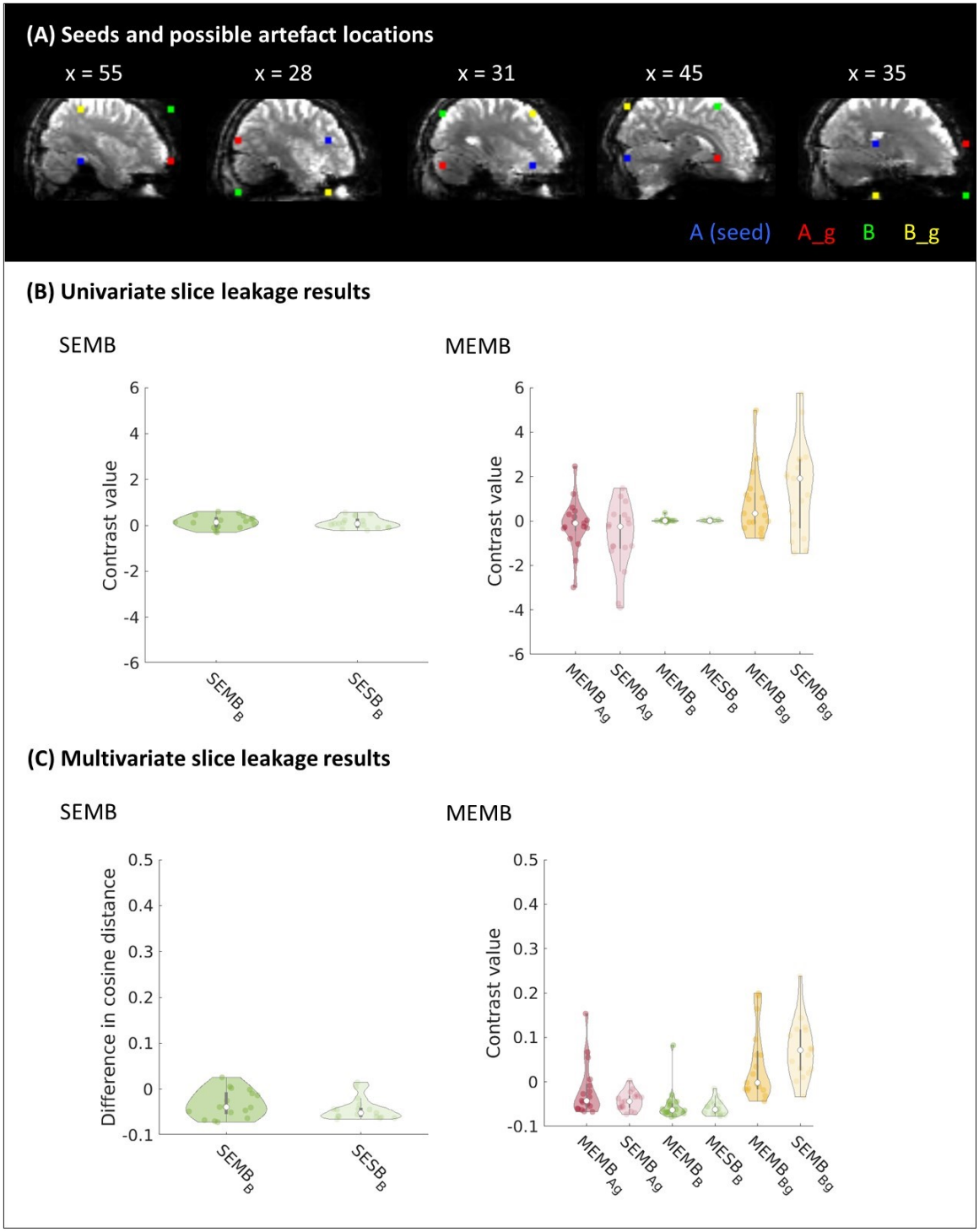

Figure 4.5: Slice leakage analysis. (A) Seed and possible artefact locations for a single participant. A (blue) is the seed location, B (green) is the possible artefact location based on phase shift, Ag (red) is the possible artefact location in the same slice as A based on GRAPPA, and Bg (yellow) is the possible artefact location in the same slice as B based on GRAPPA. All spheres have a radius of 4 voxels and are overlaid on one volume of multi-echo multiband (MEMB) data for a single participant in that participant's native space. Informed consent was obtained from the participant for this image to be published. (B) Mean activation magnitude (contrast betas) within each sphere for each MB sequence and the corresponding control sequence for the first seed location (corresponding plots for other seeds are shown in Appendix D, Figure D.7;

Figure 4.5: statistics for all seeds are shown in Appendix D, Table D.2). (C) Mean MVPA performance within each sphere for each MB sequence and the corresponding control sequence. The MVPA metric is the mean between-task cosine dissimilarity minus mean within-task cosine dissimilarity over all possible pairs of task blocks.

Our factorial design enabled us to disentangle the effects of ME and MB. As hypothesised, ME improved activation magnitude and these effects were localised to inferior temporal and orbitofrontal areas in both the ROI and the whole-brain analyses. These results demonstrated that using multiple echoes, including one very short echo, is a successful method of increasing activation magnitude in areas of magnetic susceptibility where T2* is particularly short (Halai et al., 2014, 2015, 2024; Jung et al., 2017). ME also opens up the opportunity for sophisticated denoising such as tedana (DuPre et al., 2021; Kundu et al., 2011, 2013), which improved activation precision in agreement with previous findings (Amemiya et al., 2019; Gonzalez-Castillo et al., 2016). ME was found to improve MVPA performance compared to SE sequences and, in turn, MEdn sequences performed significantly better than ME sequences without denoising in every region of interest (with no detrimental effects). Future studies employing multivariate methods should strongly consider using both ME during acquisition and ME-ICA during preprocessing.

As hypothesised, multiband improved activation precision in both the ROI and the whole-brain analyses, including in areas of inhomogeneity, and in agreement with other work (Puckett et al., 2018). After downsampling the MB timeseries to match the number of volumes in the SB timeseries, there was no longer any difference between the SB and MB sequences; this suggests that it is the increase in the number of volumes that accounts for the benefits of multiband. Additionally, there was no evidence for leakage of signal into the simultaneously-acquired slice (McNabb et al., 2020; Todd et al., 2016), at least for our modest multiband factor of 2. Therefore, combining MB with ME results in both larger activation magnitude and better activation precision, including in the vATLs, and there are other benefits too – MB can help reduce the longer TR associated with ME and may decrease noise aliasing.

This study provides evidence to use ME and/or MB modified sequences to detect signal in the vATLs. However, it is important to note that individual 7T MRI scanners have their own hardware and software limitations; this means that each site must optimise and test feasible parameters locally. For example, our system did not allow us to take advantage of the higher spatial resolution offered by 7T-fMRI while retaining an adequately short first echo for our

multi-echo sequences. Puckett et al. (2018) used comparable voxel sizes but had a shorter first TE compared to this study (9.9 ms and 11.8 ms, respectively); Miletić et al. (2020) had higher resolution (1.6 mm^3) and a shorter first echo of 9.66 ms. We also acknowledge that pTx can be combined with other methods – for example, MB (e.g. Wu et al., 2016) – but this was not implemented on our scanner at the time of running this study and therefore needs to be empirically tested in the future. Therefore, rather than identifying the limits of what is possible, we offer an accessible framework that will enable researchers to leverage 7T-fMRI for targeting regions affected by susceptibility artefacts despite the hardware and software constraints of individual scanners.

4.5 Conclusion

In this study we compared pTx, ME and MB as methods of improving sensitivity in the vATL, an area plagued by signal dropout. Both pTx and ME improved activation magnitude, but only ME showed improvements in anterior temporal regions. MB and ME-ICA improved activation precision in areas of inhomogeneity. Exploratory results suggested that both ME and ME-ICA may also benefit MVPA. We demonstrated that a multi-echo, multiband sequence can detect signal in the vATLs while maintaining sensitivity across the whole brain and is therefore a versatile choice for future studies using high field strength and investigating the functional roles of the vATLs (Lambon Ralph et al., 2017).

Data and code availability: Data will be made publicly available upon peer review and acceptance. Code is publicly available at: https://github.com/slfrisby/7TOptimisation/.

Acknowledgements: This work was supported by an MRC Unit grant (SUAG/019 G116768) to M. M. C., an MRC programme grant (MR/R023883/1) and intramural funding (MC_UU_00005/18) to M. A. L. R, and an MRC Career Development Award (MR/V031481/1) to A. D. H. We would like to thank the participants and the Wolfson Brain Imaging Centre radiographers.

Competing interests: The authors declare no conflicts of interest.

Supplementary material: Supplementary results can be found in Appendix D.

Chapter 5

Decoding semantics with 7T-fMRI: Convergent evidence and divergent discovery

Foreword

Building on the methodological foundation laid in Chapter 4, in this Chapter I assess the capacity of 7T-fMRI to detect semantic representations and thus address the second aim of this thesis. I evaluate the strengths and limitations of four multivariate decoding methods for revealing properties of semantic representations and thus address the third aim of this thesis.

This Chapter is a manuscript in preparation:

Frisby, S. L., Cox, C. R., Halai, A. D., Lambon Ralph, M. A., & Rogers, T. T. (in prep.). Decoding semantics with 7T-fMRI: convergent evidence and divergent discovery.

I designed this study together with my co-authors. I recruited and scanned all participants and preprocessed the data. With my co-authors I developed the analysis pipeline, including making improvements to the WISC MVPA toolbox and its high-throughput computing implementation. I then conducted all analyses and wrote the first draft of the manuscript, which was then edited collaboratively.

Abstract

The hub-and-spoke model of semantic representation is supported by convergent evidence from neuropsychology, computational modelling, noninvasive brain stimulation and intracranial electrophysiology. However, multivariate analyses of fMRI data produce results that rarely include the ventrolateral anterior temporal lobe (vATL) semantic "hub". We analysed distortion-corrected 7T-fMRI data using regularised logistic regression classifiers to decode

animacy and Representational Similarity Learning to decode fine-grained semantic structure. We found evidence for graded, multidimensional semantic representations in the vATL; our results indicate that signal inhomogeneity and methodological assumptions are the most likely explanations of the vATL's absence from previous studies. Capitalising on the spatial coverage that other methods lack, we also found evidence for graded, multidimensional semantic structure in posterior temporal and occipitotemporal cortex – a discovery that could drive further development of the hub-and-spoke model.

Keywords: semantic representation, 7T-fMRI, decoding, multivariate pattern analysis

5.1 Introduction

Semantic cognition is our ability to store knowledge about the world and to use this knowledge to interpret the meaning of language, objects, and events. Convergent evidence from neuropsychology (Bozeat et al., 2000; Hodges & Patterson, 2007), computational modelling (Jackson et al., 2021; Rogers et al., 2004; Rogers & McClelland, 2004), noninvasive brain stimulation (Binney et al., 2010; Lambon Ralph et al., 2009; Pobric et al., 2007), and intracranial electrophysiology (C. R. Cox et al., 2024; Matoba et al., 2024; Rogers et al., 2021; Sato et al., 2021; Shimotake et al., 2015) identifies a distributed semantic representation system with a bilateral ventrolateral anterior temporal lobe (vATL) "hub" (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004). Evidence from functional magnetic resonance imaging (fMRI), however, is at odds with this account – individual studies produce contradictory results (Frisby et al., Chapter 2) that highlight a variety of regions but rarely the vATL (e.g. Connolly et al., 2012; Devereux et al., 2013, 2018; Huth et al., 2016; Pereira et al., 2018). This study aims to resolve this apparent contradiction in the literature. Employing distortion-corrected ultra-high-field fMRI (7T-fMRI) and four methods of multivariate decoding, we address two important questions – (1) why fMRI often fails to discover semantic structure in the vATL, and (2) whether other regions of the brain encode semantic structure similar to that observed in the vATL.

5.1.1 Background and motivation

The hub-and-spoke model of semantic representation (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004) proposes that modality-specific semantic information is encoded in modality-specific areas of cortex ("spokes"). The vATL functions as a "hub" that both connects all spokes (enabling retrieval of an object's multimodal properties following input from a single sensory modality) and "re-represents" information in a way that expresses overall semantic similarity rather than just similarity of appearance, name or functional role (Lambon Ralph et al., 2010; Lambon Ralph & Patterson, 2008; Rogers et al., 2004). This overall semantic similarity structure is *multidimensional* – a tiger is similar to a cat in overall appearance, but differs in size and behaviour. It is also *graded* – although a robin, a wren, and an ostrich are all similar, the robin and the wren are more similar than either is to the ostrich (C. R. Cox et al., 2024).

The vATL's crucial role in semantic representation is supported by multiple sources of evidence. First, in semantic dementia, atrophy of the vATL causes multimodal semantic impairment but spares other cognitive abilities (Bozeat et al., 2000; Hodges & Patterson, 2007; Snowden et al., 1989; Warrington, 1975). Second, computational modelling demonstrates the importance of a multimodal hub – only models with a hidden layer receiving input from all sensory modalities are capable of representing overall semantic similarity structure (Jackson et al., 2021; Rogers & McClelland, 2004), and, when the hidden layer is damaged, a pattern of impairment resembling semantic dementia is observed (Rogers et al., 2004). Third, semantic impairments can be induced by transcranial magnetic stimulation (TMS) of the vATL (Binney et al., 2010; Lambon Ralph et al., 2009; Pobric et al., 2007, 2010a, 2010b).

In recent years, multivariate decoding, which analyses patterns of activity over multiple neural features (voxels, sensors or electrodes) to characterise what information those patterns represent, has been applied to ECoG data recorded from the vATL cortical surface of patients undergoing surgery for intractable seizures. Consistent with evidence from other sources, semantic information can be decoded both from changes in voltage (C. R. Cox et al., 2024; Rogers et al., 2021) and changes in spectral power (Clarke, 2020; Frisby et al., Chapter 3, Rupp et al., 2017). Notably, Rogers et al. (2021), using regularised logistic regression classifiers to decode animacy, found that semantic information was represented by the vATL by a pattern of activity that changed *dynamically* in time – at a given electrode, a positive deflection in voltage might indicate that a stimulus is animate at one timepoint but indicate that a stimulus is inanimate 100 ms later. These dynamic changes were more predominant in anterior regions of

electrode coverage – more posteriorly, decoding accuracy was equally high but the relationship between voltage deflections and animacy was much more stable. Additionally, C. R. Cox et al. (2024) decoded semantic structure in the vATL using Representational Similarity Learning (RSL; Oswal et al., 2016). In RSL, a similarity structure is modelled (for example, using feature norms; Dilkina & Lambon Ralph, 2012; Frisby et al., Chapter 2), then decomposed into multiple independent dimensions. Linear regression models then learn to predict the coordinates of each stimulus on each dimension; significant correlations between true and predicted coordinates indicate that the model is a good representation of the similarity structure in the brain. (C. R. Cox et al., 2024) discovered similarity structure in the vATL that was both graded and multidimensional, as predicted by the hub-and-spoke model.

5.1.2 Question 1: Why does fMRI often fail to discover semantic structure in the vATL?

In light of the evidence from ECoG, it may seem surprising that studies applying similar multivariate methods to fMRI data usually fail to identify semantic similarity structure in the vATL during semantic tasks (e.g. Connolly et al., 2012; Devereux et al., 2013, 2018; Huth et al., 2016; Pereira et al., 2018). There are at least three possible reasons for this. First, the proximity of the vATL to the air-filled sinuses can cause signal dropout and distortions. Univariate fMRI studies that use distortion correction, such as spin-echo or multi-echo acquisition, do detect activation related to semantic processing in the vATL (Binney et al., 2010; Embleton et al., 2010; Frisby et al., Chapter 4; Halai et al., 2014, 2015, 2024; Visser et al., 2010). Second, the limited temporal resolution of fMRI may be insufficient to detect a code that changes dynamically from millisecond to millisecond, an issue most likely to affect structure discovery in the anteriormost portion of the vATL (Rogers et al., 2021).

The third possibility is methodological. Multivariate methods each encapsulate hypotheses about representational structure that constrain the kind of neural code that they can detect and these assumptions can lead to drastically different results (Frisby et al., Chapter 2). As one example, stimulus category can be decoded using regularised logistic regression classifiers with LASSO (L1) regularisation (Tibshirani, 1996), which assumes that the neural code is sparse (only a few of the features under consideration encode the target structure) and uncorrelated (important features exhibit uncorrelated activity), but makes no assumption about how neural features are distributed anatomically. Alternatively, sparse-overlapping-sets LASSO

(SOSLASSO) regularisation (Rao et al., 2013, 2016) assumes that important features are relatively sparse but also in roughly the same anatomical locations within and across participants. C. R. Cox & Rogers (2021) demonstrated that, when used to decode stimulus category from fMRI data, SOSLASSO revealed a far more anatomically extensive pattern of activity (including the vATL) than did LASSO. As a second example, representational similarity analysis (RSA; Kriegeskorte et al., 2008), which promises to reveal graded, multidimensional similarity structure in neural activity, has failed to reveal anything more complex than a binary animacy distinction in ECoG data from the vATL (Y. Chen et al., 2016; but see also Clarke, 2020). This may be because RSA treats all features as equally important – it can yield null results if only a subset of possible features encodes the information of interest. Consistent with this supposition, C. R. Cox et al. (2024) successfully decoded graded and multidimensional semantic structure in the same data using RSL, in which decoding models assign different coefficients to each feature to reflect their differing importance. Successful decoding depended, however, on the choice of regularisation penalty – models fitted using LASSO regularisation were unsuccessful, but models fitted with grOWL regularisation (which incorporates gentle and plausible assumptions about the dimensionality of the neural representation; see Appendix E) were successful.

The first aim of this study was to test these three hypotheses about why discovery of semantic structure has been so elusive in fMRI studies. We tested for semantic structure in our data using two decoding approaches. To evaluate whether vATL areas differentiate animate and inanimate stimuli, we used regularised logistic regression classifiers to decode binary animacy (Rogers et al., 2021). To evaluate whether vATL areas encode graded, multidimensional semantic structure, we applied RSL to decode stimulus coordinates within a three-dimensional target semantic space. Significant correlations between predicted and true coordinates on more than one dimension provide evidence that the ROI encodes multidimensional semantic structure. Significant correlations within animate and inanimate domains arise only if semantic structure is truly graded – i.e. if the model does more than predict one value for all animate stimuli and another for all inanimate stimuli (C. R. Cox et al., 2024).

First, to evaluate whether the discrepancy arises from signal inhomogeneity in vATL, we collected data using a multi-echo, multiband acquisition sequence optimised for recovering signal in the vATL without sacrificing signal quality elsewhere in the brain (Frisby et al., Chapter 4). A null result thus would suggest that failure to discover semantic structure is not solely attributable to poor signal in this key area.

Second, to evaluate whether the dynamically-changing animacy code observed in more anterior vATL affects signal detection with fMRI, we defined a vATL region of interest (ROI) based on electrode coverage in a previous ECoG study (Frisby et al., Chapter 3). We then applied regularised logistic regression to data either from the entire ROI or from the anterior and posterior half of the ROI (separately). Recall that both anterior and posterior regions showed equally good decoding accuracy from ECoG data, but with much more pronounced dynamic change in the anterior portion. If dynamic representational change affects decoding accuracy in fMRI, we should therefore see *worse* decoding in the anterior half than the posterior half, in contrast to the ECoG result. Equally good decoding would disconfirm the hypothesis that dynamic representational change impacts decodability in fMRI.

Third, to assess whether results depend upon the decoding method employed, we tested for semantic structure using regularisation penalties that make relatively few and quite different assumptions about the nature of the semantic code (Frisby et al., Chapter 2). We compared logistic regression classifiers with LASSO regularisation (which assumes only that the code is sparse) to classifiers with SOSLASSO regularisation (which additionally assumes that informative voxels are located similarly within and across subjects; C. R. Cox et al., 2024). We also compared RSL models with LASSO regularisation (which assumes sparsity only) to the results of RSL with the neurally-inspired regularisation penalty grOWL (Appendix E).

5.1.3 Question 2: Which (if any) other regions of the brain encode semantic structure?

While the preceding questions emphasise potential limits of fMRI, this technique also offers advantages over other approaches by providing a spatially-resolved view of the whole brain. Most neuropsychological studies examine patients with focal damage; computational models include only regions that are hypothesised to be important; transcranial magnetic stimulation is applied to a small fraction of the cortex at once; and ECoG has very restricted cortical coverage, skewed in favour of regions that are more commonly epileptogenic (temporal regions are disproportionately affected by both sclerosis and glioma; Duffau, 2014; Kanemoto et al., 1996).

Whole-brain coverage is important because multiple competing models of semantic representation predict that semantic structure is encoded outside the vATL – for example, in crossmodal "convergence zones" in posterior temporal cortex (A. R. Damasio, 1989; H. Damasio et al., 2004), within sensory systems (A. Martin, 2007, 2016), or in a second hub in the angular

gyrus (Binder & Desai, 2011). The hub-and-spoke model, which was initially developed using methods that have spatial limitations (including neuropsychology and computational modelling), makes no specific predictions about whether and where graded, multidimensional semantic structure might be found outside the vATL. While the spokes are generally hypothesised to represent unimodal similarity structure, it is not clear whether this is restricted to comparatively early parts of sensory, motor and language systems, or to what extent semantic structure from the hub might permeate other regions (Lambon Ralph et al., 2017).

It should be noted that the field of multivariate fMRI does not provide consistent support for any of these theories. Some multivariate fMRI studies yield results that are consistent with a theory (A. J. Anderson, Lalor, et al., 2019; Dehghani et al., 2017; Fernandino, Binder, et al., 2016) while results from other studies highlight regions whose importance is not predicted by any theory (e.g. Huth et al., 2016). Choice of method (as well as explaining why the vATL rarely appears in multivariate fMRI studies) may again explain the heterogeneity of these findings – different methods encapsulate different assumptions and therefore yield different results.

The second aim of the study, therefore, was to assess where else semantic structure similar to that in the vATL (i.e. graded and multidimensional) was evident in the brain. We applied the same techniques used for our ROI analysis – both regularised logistic regression and RSL. As in the ROI RSL analysis, significant correlations between predicted and true coordinates on more than one dimension provide evidence for multidimensional structure, while significant correlations within domains provide evidence for gradedness.

We also visualised model coefficients on the cortical surface, which enabled us to compare our findings both to results from a univariate analysis and to predictions from previous theoretical and multivariate work (Binder & Desai, 2011; A. R. Damasio, 1989; H. Damasio et al., 2004; Huth et al., 2016; Lambon Ralph et al., 2017; A. Martin, 2007, 2016; Patterson et al., 2007; Rogers et al., 2004).

To examine the implications of different regularisation penalties on decoding (and thus to assess whether differences in methodology could account for contradictions in the literature), we applied and compared regularised logistic regression with LASSO and with SOSLASSO regularisation and RSL with LASSO and with grOWL regularisation.

5.2 Materials and methods

5.2.1 Participants

32 participants were recruited to the study. Of these, five were unable to complete the study either for medical reasons (for example, an incidental finding) or because of technical problems during data acquisition which meant that the scanner needed to be rebooted. This left 27 participants (age range 19 – 50, mean age 27.96 years, 17 female, 10 male) who were all native speakers of British English, were right-handed, had normal or corrected-to-normal vision, and had no neurological or sensory disorders. All participants gave written informed consent and the research was approved by a local National Health Service (NHS) ethics committee.

5.2.2 Stimuli and task

Stimuli were the same 100 line drawings used in previous ECoG work (C. R. Cox et al., 2024; Frisby et al., Chapter 3; Rogers et al., 2021) – 50 animals and 50 inanimate items including vehicles, clothes, musical instruments and a range of other objects (Morrison et al., 1997; https://github.com/slfrisby/7TConvergent/tree/main/stimuli/). There were no significant differences between the categories with respect to age of acquisition, visual complexity, familiarity, word frequency, name agreement, and non-semantic visual structure (Barry et al., 1997; Rogers et al., 2021). PsychoPy 2022.2.5 (Peirce et al., 2019) was used to display stimuli and to record speech. Stimuli for both tasks were rear-projected onto a screen in the MRI scanner, and observers viewed stimuli through a mirror mounted to the head coil directly above the eyes.

The study had a fast event-related design with 4 runs per participant, collected in a single session. Each run began 16 s after the start of the MR sequence and consisted of 100 trials (each stimulus appeared once per run). Each trial consisted of a stimulus presented for 4 s followed by a fixation cross presented for 4 s. 4 sequences of animate and inanimate stimuli were created and counterbalanced across participants. The order of trials was optimised for the contrast between animate and inanimate stimuli using optseq2 (https://surfer.nmr.mgh.harvard.edu/fswiki/optseq2/). For each run and each participant, a random order of animate stimuli and a random order of inanimate stimuli were created; the stimuli were then interspersed according to the sequence generated with optseq2.

Participants were instructed to name each picture out loud as quickly and accurately as possible while keeping head motion to a minimum. To ensure that participants understood the

instructions, they practised the task before entering the scanner while the experimenter provided feedback about their head motion. Practice continued until the experimenter was satisfied that head motion was minimised.

Overt speech was recorded for each trial and responses were scored for accuracy with the study's aim of decoding semantic similarity structure in mind. If the participant did not respond within the 4 seconds during which the stimulus was on-screen, the trial was marked as incorrect. If the participant responded incorrectly but named an object that the rater judged to be visually and semantically similar to the stimulus and at a similar level of specificity (e.g. "wasp" for bee), the trial was marked as correct; if the object that they named was semantically dissimilar or was named at a less specific level (e.g. "insect" for bee), the trial was marked as incorrect. It was important that neural responses did not reflect a "blend" of the semantics of multiple stimuli so, if the participant gave two synonyms (e.g. "gun, pistol") the trial was marked as correct; if they named two different objects ("kettle, iron") the trial was marked as incorrect even if one of the names was correct. Trials containing fillers such as "um" and "err" in addition to a correct response were marked as correct because the participants were neurologically healthy and there was no reason to believe that these utterances reflected anything other than normal speech processes. The mean and standard deviation of the accuracy scores were calculated.

5.2.3 Image acquisition

All images were acquired on a whole-body 7T Siemens MAGNETOM Terra MRI scanner (Siemens Healthcare, Erlangen, Germany) with an 8Tx32Rx head coil (Nova Medical, USA).

An MP2RAGE anatomical scan was acquired with the following parameters: 224 sagittal slices (interleaved acquisition), FOV 240 × 225.12 × 240 mm, voxel size 0.75 × 0.75 × 0.75 mm, TR 4300 ms, inversion times 840 and 2370 ms, TE 1.99 ms, flip angles 5° and 6°, GRAPPA acceleration factor 3. Next, manual B_0-shimming was performed over the volume to be used for echo-planar imaging (EPI). The aim was to reduce the water linewidth, defined as full width of the spectrum at half height, to below 40 Hz. However, for some participants the adjustments proved time-consuming and so adjustment time was capped at 30 minutes from the start of scanning after which the best shim parameters were adopted. The actual average water linewidth was 38.80 Hz (standard deviation = 5.93 Hz).

All four functional runs of EPI used the multi-echo multiband sequence optimised by Frisby et al. (Chapter 4) with the following parameters: 48 axial slices (interleaved acquisition),

FOV 210 × 210 × 210 mm to cover the whole brain in most participants (visual inspection ensured that the vATL was included in the FOV for all participants and the FOV was tilted up at the nose to avoid ghosting of the eyes into the vATL), TR = 1510 ms, 3 echoes, TE_1 = 11.80 ms, TE_2 = 27.05 ms, TE_3 = 42.30 ms, multiband factor 2, GRAPPA acceleration factor 3, phase partial Fourier 7/8, voxel size 2.5 mm isotropic (no gap), flip angle 63º, and A-P phase encoding direction. Immediately *before* each run, 5 additional volumes were acquired with the phase encoding direction changed to P-A. These volumes were to be used for distortion correction during preprocessing.

5.2.4 Data analysis

All analysis code is available at: https://github.com/slfrisby/7TConvergent/.

5.2.4.1 Preprocessing

For ease of data sharing we converted all DICOMs to BIDS format (Gorgolewski et al., 2016) using *heudiconv* (v1.1.1; Halchenko et al., 2024).

Standard reproducible preprocessing pipelines designed for 3T-fMRI, such as *fMRIprep* (Esteban et al., 2019; Gorgolewski et al., 2011) perform poorly on 7T-fMRI EPI data; therefore, the analysis pipeline was split into stages using different software packages. The MP2RAGE T1w (combined) image first had its background noise removed using O'Brien regularisation (O'Brien et al., 2014) and was then submitted to CAT12 for segmentation (Gaser et al., 2023; in SPM12; https://www.fil.ion.ucl.ac.uk/spm/). The bias-corrected and global-intensity-corrected T1w (combined) image produced was provided to as input to *fMRIPrep* 21.0.1. The T1w (combined) image was corrected for intensity non-uniformity with *N4BiasFieldCorrection* (Tustison et al., 2010) and skull-stripped with a Nipype implementation of the *antsBrainExtraction.sh* workflow (ANTs). Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM) and grey matter (GM) was performed on the brain-extracted T1w (combined) image using *fast* (FSL 6.0.5.1; Zhang et al., 2001). Volume-based spatial normalisation to standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with *antsRegistration.sh* (ANTs 2.3.3, Avants et al., 2009, 2011; https://github.com/ANTsX/ANTs/).

Functional preprocessing was performed using in-house code, composed of functions from AFNI (v.18.3.03; R. W. Cox, 1996; R. W. Cox & Hyde, 1997; https://afni.nimh.nih.gov), FSL (v.5.0.10; Andersson et al., 2003; Jenkinson et al., 2012; S. M. Smith, 2002; S. M. Smith et al.,

2004; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/), tedana (v.24.0.0; DuPre et al., 2021; Kundu et al., 2011, 2013; https://tedana.readthedocs.io/en/stable/index.html) and ANTs (v. 2.2.0; Avants et al., 2009, 2011; https://github.com/ANTsX/ANTs/). EPIs were slice-time corrected to the middle slice using *3dTshift* (AFNI). Next, EPIs from the first echo were motion corrected with *3dvolreg* and *3dAllineate* using the first volume of the first run acquired as a reference. The same transforms were applied to the EPIs from the second and third echoes. EPIs were then skull-stripped using *BET* (FSL) to create a participant-specific brain mask. *tedana* was used to combine data from all echoes optimally based on T2* weighting (Posse, 2012) and to denoise the data using ICA. *tedana* conducts denoising by decomposing data using PCA and ICA, classifying components according to whether the signal scales linearly with TE (as BOLD signal does), and reconstruct the data using only BOLD-like components. The brain mask created with *BET* was used as the mask for this stage.

Next, unwarping was conducted using *topup* and *applytopup* (FSL). Field displacement maps were calculated using ten volumes (five with A-P phase encoding direction, extracted from the start of each functional run, and five with P-A phase encoding direction, collected separately before each run) and the resulting correction was applied to all images.

Finally, the mean EPI for each run was coregistered to the skull-stripped native structural image using a rigid-body registration in *AntsRegistrationSyN.sh* (ANTs). This produced transforms from native EPI space to native T1 space. For all decoding analyses, images remained in native space and were not smoothed. However, the transforms from native EPI to native T1 (for the first run) and the transforms from native T1 to standard space (MNI152NLin2009cAsym), with *antsApplyTransforms* (ANTs), were used to calculate the MNI-space coordinates that each voxel *would* have *were* images to be transformed into MNI space. (These coordinates were used for visualisation and were provided to classifiers trained with SOSLASSO regularisation, which take anatomical information as input.) The transform from native EPI to native T1 space was also used to transform the participant's grey-matter mask into native EPI space. For the univariate analysis, images were transformed into standard space (MNI152NLin2009cAsym) with *antsApplyTransforms* (ANTs) and smoothed with a 6 mm FWHM Gaussian filter in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/).

5.2.4.2 1st-level (within-participant) GLM

Data were processed using the general linear model (GLM) approach implemented in SPM12 in MATLAB r2023b. For decoding, each stimulus was modelled as an individual boxcar function, 1.5 seconds in length because the stimulus in our stimulus set with the longest mean naming time had a mean naming time of 1323 ms (Morrison et al., 1997; periods between stimuli were modelled implicitly). These boxcar functions were subsequently convolved with SPM's difference of gammas haemodynamic response function. The six motion parameters extracted during preprocessing were used as regressors of no interest. Data from all four runs was concatenated, so an additional constant was included for each run. The micro-time resolution was set as the number of slices (n = 48), the micro-time onset was set as the reference slice for slice-time correction (n = 24), and the high-pass filter cutoff was 128 seconds. Each participant's grey matter mask in native EPI space was used as a mask for the analysis. The parameter estimation method was restricted maximum likelihood estimation (ReML) and serial correlations were accounted for using an autoregressive AR(1) model during estimation. This produced one beta image per stimulus in native space. For univariate analysis, the same model was run on the normalised, smoothed data with identical parameters except that a grey matter mask in standard space (MNI152NLin2009cAsym) was used and that stimuli named incorrectly were modelled as additional regressors of no interest. The univariate contrasts of interest were greater activation for animate stimuli named correctly than inanimate stimuli named correctly (A>I) and greater activation for inanimate stimuli named correctly than animate stimuli named correctly (I>A). This produced contrast images in MNI space.

5.2.4.3 Univariate analysis – 2nd-level (across-participant) GLM

Group-level t-tests of each contrast of interest (A>I, I>A) were conducted in SPM12. The group t-maps were assessed for significance by using a voxel-height threshold of $p<0.001$ to define clusters and then a cluster-defining family-wise-error-corrected threshold of $p<0.05$ for statistical inference. Group-level results were then projected to the pial surface (fsaverage template; Dale et al., 1999; Fischl et al., 1999) using *-volume-to-surface-mapping* with trilinear interpolation in Connectome Workbench 1.5.0 and the thresholded maps were plotted on the pial surface using *plot_surf_stat_map* in *nilearn* implemented in Python 3.9.

5.2.4.4 Multivariate decoding

5.2.4.4.1 Decoding approach

A detailed overview of all four decoding methods is provided in Appendix E.

To prepare data for decoding, beta images from all 4 presentations of each stimulus were averaged across runs. If a participant had named a stimulus incorrectly on one to three occasions, the beta image(s) corresponding to the incorrect trial(s) were not included in the averaging. If the participant named the stimulus incorrectly on all four trials, the values for each voxel were interpolated with the median value for that voxel over all other stimuli.

We tested for semantic structure using two decoding methods. First, to assess whether and in which regions animate and inanimate stimuli could be reliably differentiated, we used regularised logistic regression classifiers to decode binary animacy (Rogers et al., 2021). One set of classifiers was trained using LASSO regularisation (Tibshirani, 1996), which assumes that the neural code is sparse (only a few of the features under consideration encode the target structure) and uncorrelated (important features exhibit uncorrelated activity). Classifiers were fitted using *glmnet* in MATLAB r2023b (Friedman et al., 2010), using 3 – 0.2 as the range of possible λ values. Each classifier took, as input features, beta values from multiple voxels from a single participant. A second set of classifiers was trained using SOSLASSO regularisation (C. R. Cox & Rogers, 2021), which assumes that the neural code is relatively sparse but also in roughly the same anatomical location within and across participants. Classifiers were fitted using the WISC MVPA toolbox in MATLAB r2018b (https://github.com/crcox/WISC_MVPA/) with hyperband budget set to 25 (all other parameters were set as defaults, including a set size of 18 mm with 9 mm overlap and 3 – 0.2 as the range of possible λ values). These classifiers took beta values from multiple voxels from multiple participants as input and, though data were in native space, the MNI-space coordinates that each voxel *would* have *were* images to be transformed into MNI space were also provided as input to the classifier.

Second, we used RSL to assess whether and where graded, multidimensional semantic structure was represented. Target semantic dimensions were modelled following the approach of C. R. Cox et al. (2024). First, a representational similarity matrix (RSM) was constructed that expressed the semantic similarity between each possible pair of stimuli (operationalised using feature verification norms; Dilkina & Lambon Ralph, 2012); then, singular value decomposition (SVD) was applied to to extract three components accounting for 89.5 % of the variance in the whole RSM (81.1 %, 4.4 %, and 4.0 %, respectively). Figure 5.1 shows the coordinates of all 100

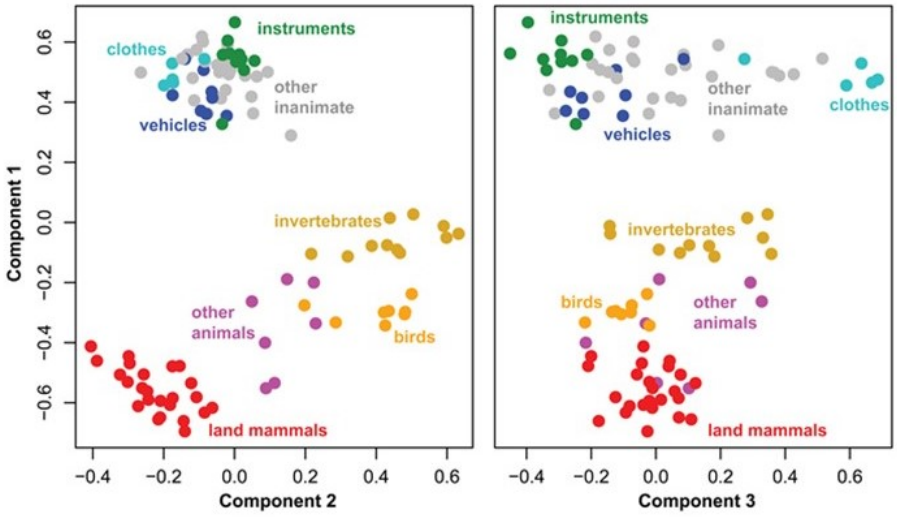

Figure 5.1: Coordinates of all stimuli on each target semantic dimension. Colours indicate category membership – land mammals (red), birds (orange), invertebrates (brown), other animals (magenta), instruments (green), clothes (light blue), vehicles (dark blue), and other inanimate objects (grey). Reprinted with permission from C. R. Cox et al. (2024).

stimuli on each dimension. RSL models were fitted independently for each participant to predict the coordinates of each stimulus on the three target semantic dimensions from beta values from multiple voxels. One set of models used LASSO regularisation and were trained using *glmnet* in MATLAB r2023b, using 6 – 0 as the range of possible λ values. A second set of models used grOWL regularisation, which assumes that the neural code is sparse, correlated (important features exhibit correlated activity), and *not* "axis-aligned" with the target similarity space (for further detail, see Appendix E). grOWL models were trained using the WISC MVPA toolbox in MATLAB r2018a with hyperband budget set to 25 (all other parameters were set as defaults, including 6 – 0 as the range of possible λ values).

For both logistic regression classifiers and RSL models, accuracy was assessed via ten-fold nested cross-validation. In each outer loop, ten out of 100 stimuli (five animate, five inanimate) were defined as the outer-loop hold-out set. A second ten stimuli (five animate, five inanimate) were defined as the inner-loop hold-out set. The remaining 80 stimuli were used as training data to search for the best hyperparameter(s) (defined as the hyperparameter(s) that minimised the mean squared error or, for grOWL, the Frobenius norm of the difference between the target matrix and the decoding matrix; see Appendix E). Models with these hyperparameter values were assessed on the inner-loop hold-out set. The folds were then reassigned so that ten stimuli previously used for training became the inner-loop hold-out set. Once all combinations of

inner-loop hold-out set and training data had been explored, the best-performing hyperparameter value(s) were used to train a final model that was assessed on the outer-loop hold-out set. The procedure was then repeated with the nine other possible hold-out sets. In the first instance, stimuli were assigned randomly to the ten hold-out folds, but this order was then fixed so that models fitted with every kind of regularisation were assessed using the same fold configuration.

The mean classification accuracy (for regularised logistic regression classifiers) or correlation between true and predicted coordinates (for RSL models) was calculated over all hold-out folds. Statistical significance was assessed via permutation testing. This was done for two reasons. First, since we used a fast event-related design rather than a slow event-related design, the haemodynamic response overlapped between trials. This means that, if temporally adjacent stimuli were divided between training and test set, the stimulus in the training set may contain information about the stimulus in the test set and so may produce classification hold-out accuracies above 0.5 or hold-out correlations above 0. Second, cross-validated correlation (as was calculated for RSL models) is known to produce predicted coordinates that sometimes correlate *negatively* with the target coordinates in a hold-out set – the mean correlation under the null hypothesis may be below zero (C. R. Cox et al., 2024; Zhou et al., 2017); therefore, the magnitude of the real correlation can be understood only with reference to the permutation distribution. In permutation testing, the stimulus labels were permuted before model training (as for hold-out folds, the permutation orders were generated at random but then fixed so that the permutations for each model were comparable) and the procedure was repeated 100 times to produce 100 simulated accuracy or correlation values for each participant. A group-level empirical null distribution was simulated by randomly sampling one of the 100 values from each participant and computing a group average 10,000 times (Stelzer et al., 2013). One-tailed p-values were calculated (where m is the number of values in the permutation distribution and b is the number of values in the permutation distribution larger than the true value; C. R. Cox et al., 2024; Phipson & Smyth, 2010):

$$p = \frac{b + 1}{m + 1}$$

To test whether one model or ROI produced a higher accuracy or correlation than another, a group-level permutation distribution of differences was created using the differences

between the values in both permutation distributions and using the group-level simulation procedure described above. A *p*-value was then calculated for the true difference. Statistical significance for all tests was defined after adjusting *p*-values to control the false discovery rate at $\alpha = 0.05$ (Benjamini & Hochberg, 1995).

Ten-fold nested cross-validation does not provide a clear picture of which voxels are informative – each of the ten models selects its own set of coefficients (C. R. Cox & Rogers, 2021). Therefore, to visualise coefficients, one model was trained for each participant using an inner loop of ten stimuli to identify the best hyperparameter(s) but no outer loop (i.e. all 90 stimuli that were not in the current inner loop hold-out set were used for hyperparameter exploration; therefore, all 100 stimuli were used for training). 100 permutation models were also trained for each participant. Coefficients were converted to one NIFTI volume per participant at native resolution (2.5 mm isotropic) in MNI space using SPM12. Coefficients both for models trained on real data and for models trained on permuted data were projected to the pial surface (fsaverage template; Dale et al., 1999; Fischl et al., 1999) using *-volume-to-surface-mapping* with trilinear interpolation and smoothed on the surface at 6 mm FWHM using *–metric-smoothing* in Connectome Workbench 1.5.0. Two metrics were calculated – (1) the proportion of participants in which each vertex was selected and (2) the proportion of negative coefficients that each vertex received (since animals were coded as 0 and inanimate objects as 1, a negative coefficient indicates that a positive beta value is associated with increased probability that the stimulus is animate). The map of the proportion of negative coefficients was thresholded using the map of coefficient selection – vertices were included in the map only if they were selected significantly more frequently in the real distribution than they were in the permutation distribution, as assessed via binomial tests with *p*-values adjusted to control the false discovery rate at $\alpha = 0.05$ (Benjamini & Hochberg, 1995). Finally, the thresholded map was plotted on the pial surface using *plot_surf_stat_map* in *nilearn* implemented in Python 3.9.

5.2.4.4.2 Experimental questions

5.2.4.4.2.1 Question 1: Why does fMRI often fail to discover semantic structure in the vATL?

An ROI was constructed based on the electrode coverage (all electrode coordinates are available at https://github.com/slfrisby/7TConvergent/). Electrodes in the temporal lobe were converted to one NIFTI volume at native resolution (2.5 mm isotropic) in MNI space using SPM12 and smoothed at 10mm FWHM. In order to exclude regions that lay outside the temporal lobe, a

mask was created by taking the Harvard-Oxford cortical and subcortical atlases, removing the temporal lobe from the atlas, and adding the cerebellum from the AAL 2 atlas (Tzourio-Mazoyer et al., 2002; since the Harvard-Oxford atlas does not include a cerebellum). Only voxels that lay outside this atlas mask were included in the ROI. After masking the ROI was smoothed again at 4mm FWHM because, on inspection, this created a smooth shape with space around the temporal lobes – since the ROIs were to be backprojected into native space, it was important to ensure that individual participants' temporal lobes would be fully encompassed. Since electrode coverage was far more extensive in the left than the right hemisphere, a mirror-image of the left ROI was used as the right ROI. The final ROIs in standard space (MNI152NLin2009cAsym) are shown projected to the surface in Figures 2 and 3 and are available at https://github.com/slfrisby/7TConvergent/tree/main/ROI/. The ROIs were backprojected into each participant's native space using the transforms from native EPI to native T1 and the transforms from native T1 to standard space generated with *antsRegistration.sh*. Only beta values from voxels that overlapped with both the ROI and the participant's native-space grey matter mask were provided as input to decoding models.

To test for the presence of semantic structure, we trained both regularised logistic regression classifiers (to decode binary animacy) and RSL models (to decode fine-grained semantic structure) on beta values from one ROI at a time (not from both hemispheres simultaneously). In the RSL analysis, significant correlations between predicted and true coordinates on more than one dimension would be taken as evidence that the discovered semantic structure was multidimensional. To assess whether the structure was graded, we evaluated RSL models on stimuli from within one domain only (animate or inanimate) – correlations within-domain will arise only if the model does more than predict one value for all animate stimuli and another for all inanimate stimuli.

We had three hypotheses about why the vATL may not appear in multivariate fMRI studies of semantic representation. The first hypothesis was that semantic structure would be obscured by signal inhomogeneity. Use of a multi-echo, multiband acquisition sequence made us as confident as possible that signal inhomogeneity would not obstruct signal discovery; a null result would therefore suggest that failure to discover semantic structure is not solely attributable to poor signal in the vATL.

The second hypothesis was that decoding accuracy may be degraded by dynamic representational change in more anterior areas (previously observed with ECoG by Rogers et al.

(2021). To evaluate this hypothesis, we bisected our ROI between its anteriormost and posteriormost coordinates in MNI space to create anterior and posterior half-ROIs. We trained separate regularised logistic regression classifiers within each half. Since Rogers et al. (2021) found more dynamic change in more anterior aspects of the ROI, and because fMRI lacks temporal resolution, comparable decoding accuracy in the anterior and the posterior half would falsify our hypothesis that dynamically-changing semantic codes in very anterior areas are difficult to detect with fMRI. Conversely, successful decoding in the posterior half and a null result in the anterior half would support the hypothesis that dynamic properties prevent signal discovery with fMRI. We also trained RSL models on each half for completeness although, because dynamic representation of multidimensional semantics has not yet been documented with time-resolved imaging methods, we had no clear hypotheses about relative correlations in each half and so could not conduct a formal test for dynamism.

Our third hypothesis was that discovery of semantic structure depends on the assumptions encapsulated by the analysis method. We examined the implications of different regularisation penalties on decoding by comparing results from logistic regression with LASSO regularisation to results from logistic regression with SOSLASSO regularisation and by comparing results from RSL with LASSO regularisation to results from RSL with grOWL regularisation.

5.2.4.4.2.2 Question 2: Which (if any) other regions of the brain encode semantic structure?

To test for the presence of semantic structure at the whole-brain level, we trained both regularised logistic regression classifiers (to decode binary animacy) and RSL models (to decode fine-grained semantic structure) on beta values from the whole brain (within the participant's native-space grey matter mask). RSL also enabled us to test whether this code exhibited some of the same properties as the semantic code in the vATL (multidimensionality and gradedness) – significant correlations between predicted and true coordinates on more than one dimension would be taken as evidence that the discovered semantic structure was multidimensional and significant correlations within-domain would be taken as evidence that the structure was graded.

The primary aim of the whole-brain analysis was to explore the spatial extent of any graded, multidimensional semantic representations that we discovered (an opportunity not afforded by methods such as neuropsychology, computational modelling, noninvasive brain stimulation and intracranial electrophysiology). We therefore visualised coefficients from all models on the cortical surface. This allowed us to compare our findings both to results from our

univariate analysis and to the predictions of different theories of semantic representation, including the hub-and-spoke model.

We hypothesised that methodological differences may account for the heterogeneity of the findings in the literature. Therefore, we examined the implications of different regularisation penalties on decoding by comparing results from logistic regression with LASSO regularisation to results from logistic regression with SOSLASSO regularisation and by comparing results from RSL with LASSO regularisation to results from RSL with grOWL regularisation. We compared both accuracy and spatial extent.

5.3 Results

5.3.1 Excluded participants

No participants were excluded because of excessive head motion, as defined by two metrics – framewise displacement (Power et al., 2012) and DVARS (which indexes the rate of change of BOLD signal across the entire brain between each pair of successive frames of data; Afyouni & Nichols, 2018). All subsequent analyses were conducted on all 27 participants who were able to complete the study successfully.

5.3.2 Behavioural results

The participants named the stimuli with high accuracy (mean = 96.85 %, standard deviation 2.71 %). Three participants had at least one run with accuracy less than two standard deviations below the group mean. However, since none of these participants had accuracy less than 88 % on any run, the decision was made to include them.

5.3.3 Question 1: Why does fMRI often fail to discover semantic structure in the vATL?

In contrast to neuropsychology, computational modelling, noninvasive brain stimulation and intracranial electrophysiology (which support the theory that the vATL encodes graded, multidimensional semantic structure), multivariate analyses of fMRI data rarely reveal semantic structure in the vATL. Accordingly, the first aim of this study was to test three hypotheses about why discovery of semantic structure has been so elusive in fMRI studies – signal inhomogeneity,

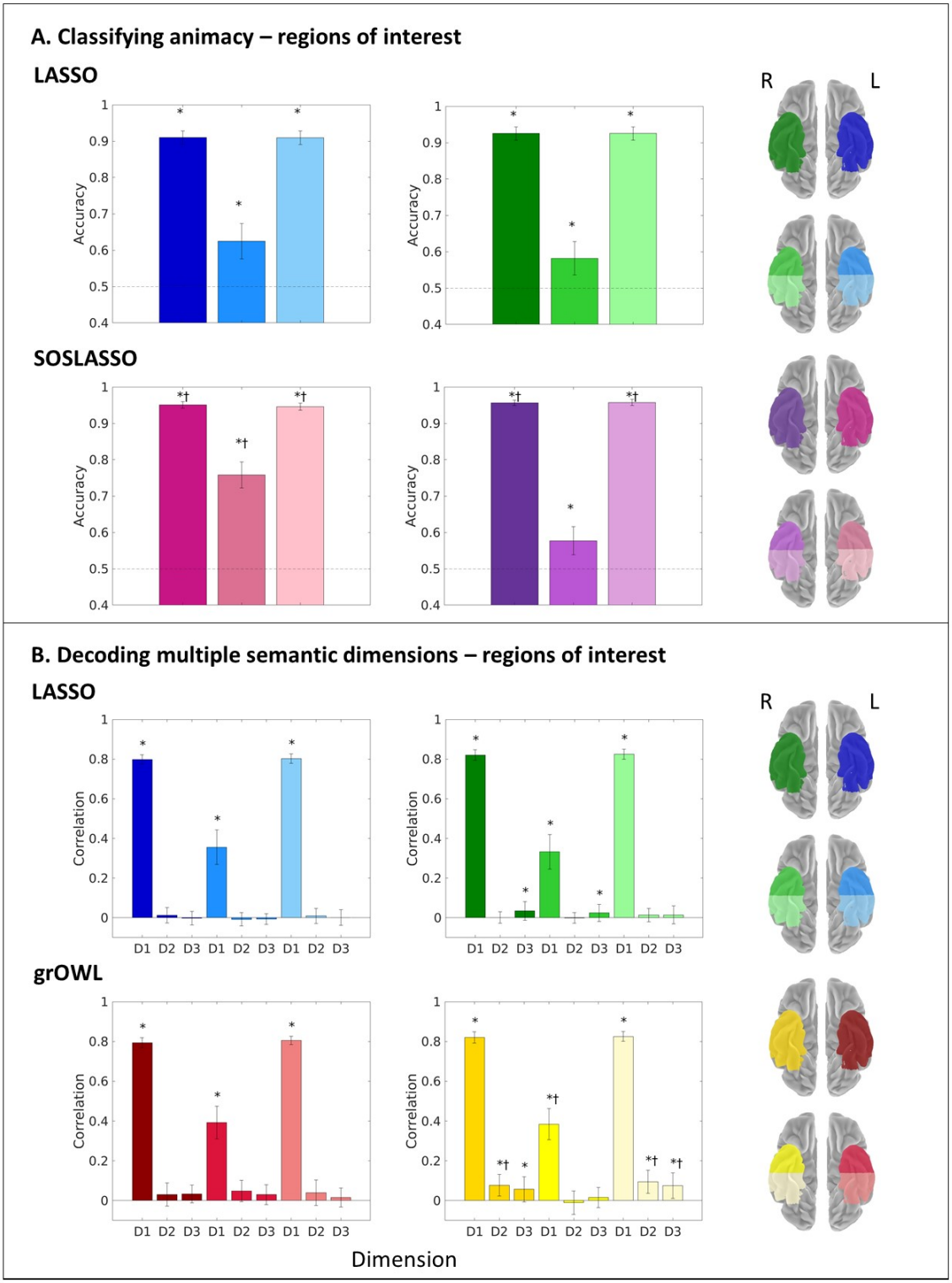
A. Classifying animacy – regions of interest
LASSO
SOSLASSO
Accuracy
R
L
B. Decoding multiple semantic dimensions – regions of interest
LASSO
grOWL
Correlation
D1 D2 D3 D1 D2 D3 D1 D2 D3
Dimension

Figure 5.2

Figure 5.2: Decoding results in regions of interest. Surface plots (right) indicate which bar corresponds to which ROI or half-ROI. (A) Mean and 95 % confidence interval of the hold-out accuracy for regularised logistic regression classifiers trained on contrast beta values to discriminate animate from inanimate stimuli. Classifiers are trained using LASSO regularisation on the left ROI and half-ROIs (upper left, shades of blue), using LASSO regularisation on the right ROI and half-ROIs (upper right, shades of green), using SOSLASSO regularisation on the left ROI and half-ROIs (lower left, shades of pink), or using SOSLASSO regularisation on the right ROI and half-ROIs (lower right, shades of purple). Stars (*) indicate a significant difference between classifier accuracy and chance (as determined by permutation test, probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Daggers (†) indicate a significant difference between classifier accuracy using that regularisation penalty and accuracy in the same ROI using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (B) Mean and 95 % confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Models are trained using LASSO regularisation on the left ROI and half-ROIs (upper left, shades of blue), using LASSO regularisation on the right ROI and half-ROIs (upper right, shades of green), using grOWL regularisation on the left ROI and half-ROIs (lower left, shades of red), or using grOWL regularisation on the right ROI and half-ROIs (lower right, shades of yellow). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension in the same ROI using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$).

dynamic representation of semantic structure, and assumptions implicit in decoding methods.

5.3.3.1 Is semantic structure present in the vATL?

5.3.3.1.1 Binary animacy

Figure 5.2A shows the results from decoding animacy using logistic regression with LASSO or SOSLASSO regularisation, and comparing accuracy in the whole ROI, the anterior half, and the posterior half in each hemisphere. All classifiers trained on all ROIs performed better than chance with both regularisation functions. Classification accuracy was very high (hold-out accuracy $\cong 0.95$) for decoders fitted across the whole ROI or to the posterior half. This result confirmed that binary semantic structure was represented within our ROI.

5.3.3.1.2 Multidimensional semantic structure

Figure 5.2B shows the mean hold-out correlation between predicted and true coordinates on each target semantic dimension, for the full ROI and each half-ROI in each hemisphere, using

RSL models with LASSO or grOWL regularisation trained on all 100 stimuli. Models fitted with both regularisation penalties showed very good decoding of the first dimension of the target semantic space in both hemispheres and all ROIs. In the right hemisphere, models fitted with both regularisation penalties produced significant correlations on at least one other dimension that, though small, were reliably better than the permutation mean (dimension 3 for LASSO, and both dimension 2 and dimension 3 for grOWL). In the left hemisphere, the second and third dimensions were not reliably decoded with either regularisation penalty for any ROI.

These results confirmed the presence of multidimensional semantic structure in the right ROI.

5.3.3.1.3 Graded semantic structure

Significant correlations within animate and inanimate domains arise only if semantic structure is truly graded – i.e. if the model does more than predict one value for all animate stimuli and another for all inanimate stimuli (C. R. Cox et al., 2024).

Figure 5.3 shows correlations calculated for animate and inanimate stimuli separately. For animate stimuli, results depended on the regularisation penalty – whereas models fitted with LASSO exhibited reliable within-domain decoding of the first target dimension only, models fitted with grOWL discovered within-domain structure for both the first and second dimensions in both anterior and posterior halves of the ROI. Within the inanimate domain, although the correlation coefficients were generally smaller overall, reliable decoding was observed with both forms of regularisation for multiple dimensions across both hemispheres. Whereas these patterns were somewhat heterogeneous for models regularised with LASSO, grOWL systematically showed significant decoding of dimensions 1 and 3 in both hemispheres.

When the two domains were considered separately, multidimensional semantic structure was observed in the left temporal lobe for both, despite the fact that correlations on only one dimension were observed when all stimuli were considered together (Figure 5.2). Note that dimension 2 separates out the animate stimuli, but not the inanimate stimuli, and the reverse is true for dimension 3 (Figure 5.1); significant correlations within-domain usually emerged on dimension 2 for animate stimuli and on dimension 3 for inanimate stimuli.

These results confirmed that graded semantic structure was represented within our ROI.

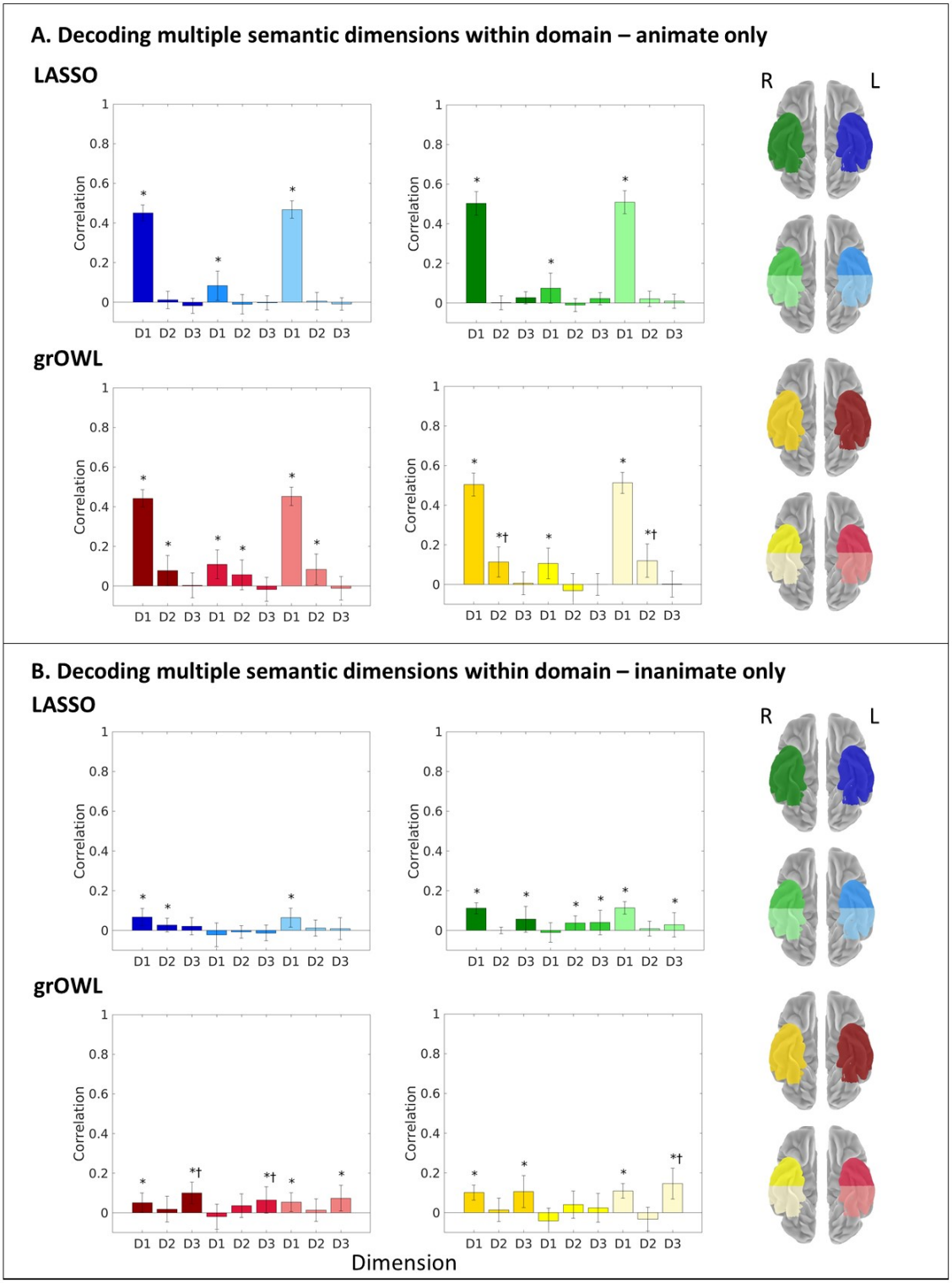
A. Decoding multiple semantic dimensions within domain – animate only
LASSO
grOWL
B. Decoding multiple semantic dimensions within domain – inanimate only
LASSO
grOWL
Correlation
Dimension
D1 D2 D3 D1 D2 D3 D1 D2 D3
R
L

Figure 5.3

Figure 5.3: Hold-out correlations within-domain in regions of interest. Mean and 95 % confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Surface plots (right) indicate which bar corresponds to which ROI or half-ROI. Models are trained using LASSO regularisation on the left ROI and half-ROIs (upper left, shades of blue), using LASSO regularisation on the right ROI and half-ROIs (upper right, shades of green), using grOWL regularisation on the left ROI and half-ROIs (upper left, shades of red), or using grOWL regularisation on the right ROI and half-ROIs (upper right, shades of yellow). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension in the same ROI using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (A) Hold-out correlations for animate stimuli only; (B) Hold-out correlations for inanimate stimuli only.

5.3.3.2 Does dynamic representational change challenge the discovery of semantic structure with fMRI?

Returning to Figure 5.2, we examined the difference in classification accuracy between the anterior and the posterior halves of the ROI. Since Rogers et al. (2021) found more dynamic change in more anterior aspects of the ROI, and because fMRI lacks temporal resolution, comparable decoding accuracy in the anterior and the posterior half would falsify our hypothesis that dynamically-changing semantic codes in very anterior areas are difficult to detect with fMRI. Conversely, successful decoding in the posterior half and a null result in the anterior half would support the hypothesis that dynamic properties prevent signal discovery with fMRI.

Figure 5.2A shows that accuracy in the anterior half-ROI was significantly worse than accuracy in the posterior half-ROI (permutation $p<0.05$) in both hemispheres and with both regularisation penalties. These results were thus consistent with the proposal that the dynamic change arising in the anterior half weakened the ability to detect this structure with fMRI. However, accuracy in the anterior half remained significant, so this dynamism was not a complete barrier to signal discovery.

Figure 5.2B shows the results of RSL models trained on each half-ROI for completeness. Consistent with the classification analysis, worse decoding of the first target dimension was observed in the anterior half-ROI than the posterior half ($r \cong 0.4$ vs. $r \cong 0.8$). However, dynamic representation of multidimensional semantics has not yet been characterised with time-resolved imaging methods. We therefore chose not to conduct a formal test of the

differences between half-ROIs on each dimension or to interpret differences as evidence for dynamic coding of multidimensional semantic structure.

5.3.3.3 Does choice of decoding method challenge the discovery of semantic structure with fMRI?

Multivariate methods each encapsulate hypotheses about representational structure that constrain the kind of neural code that they can detect and these assumptions can lead to different results (Frisby et al., Chapter 2). We explored the impact of methodological decisions on decoding in the vATL by comparing the effects of different regularisation penalties.

Figure 5.2A compares logistic regression with LASSO regularisation to logistic regression with SOSLASSO regularisation. In all cases ROIs except for the right anterior half-ROI, SOSLASSO produced significantly higher classification accuracy than LASSO (permutation $p<0.05$).

Figure 5.2B shows that, for the RSL analyses, models with grOWL regularisation produced significant correlations on all three dimensions, whereas models with LASSO regularisation produced significant correlations on only two.

The impact of regularisation penalty was even more striking for models evaluated within-domain. Figure 5.3 shows that, for animate stimuli, models fitted with LASSO exhibited reliable within-domain decoding of the first target dimension only; by contrast, models fitted with grOWL discovered within-domain structure for both the first and second dimensions, in both anterior and posterior halves of the ROI. Within the inanimate domain, grOWL systematically showed significant decoding of dimensions 1 and 3 in all ROIs in both hemispheres, whereas LASSO exhibited less consistent results.

These results therefore suggested that choice of regularisation penalty impacts accuracy in classification analyses, and both size of correlation and number of dimensions decoded in RSL.

5.3.4 Question 2: Which (if any) other regions of the brain encode semantic structure?

The first set of analyses assessed the potential limits of fMRI, but this technique also offers advantages over other approaches by providing a spatially-resolved view of the whole brain. Various theories of semantic representation hypothesise that semantic similarity structure might

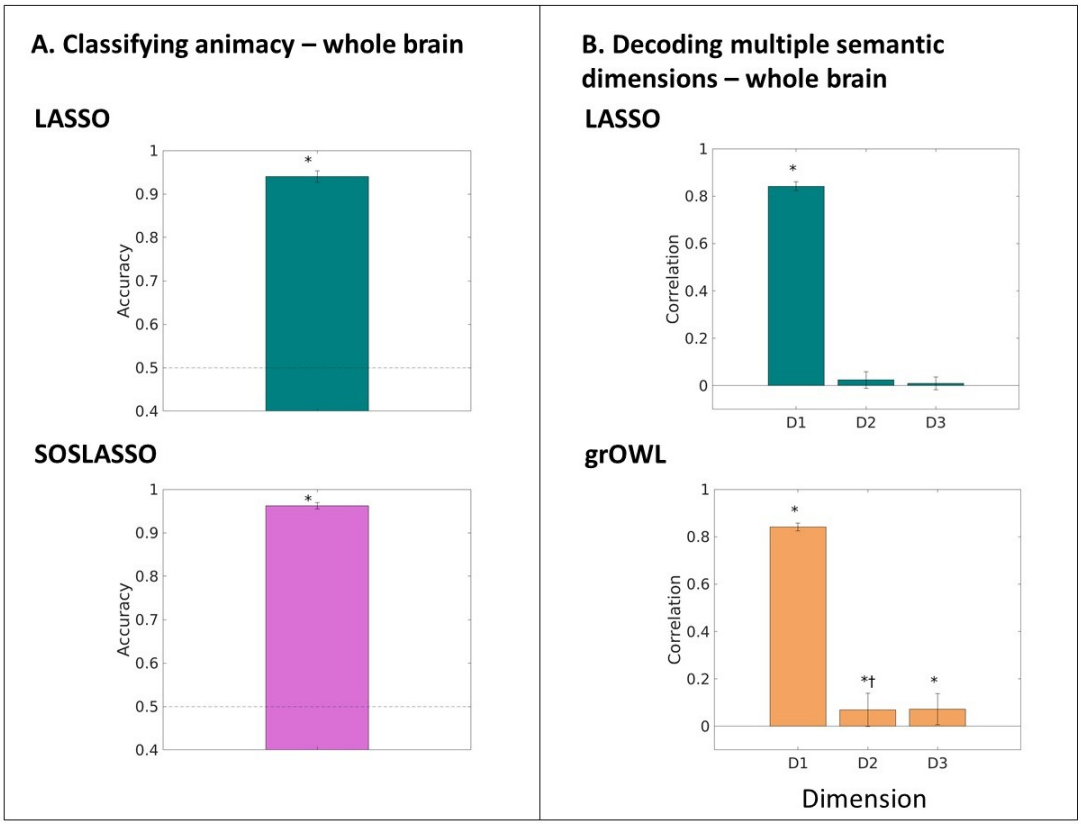

Figure 5.4: Decoding results at the whole-brain level. (A) Mean and 95 % confidence interval of the hold-out accuracy for regularised logistic regression classifiers trained on contrast beta values to discriminate animate from inanimate stimuli. Classifiers are trained using LASSO regularisation (top, teal) or using SOSLASSO regularisation (bottom, magenta). Stars (*) indicate a significant difference between classifier accuracy and chance (as determined by permutation test, probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (B) Mean and 95 % confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Models are trained using LASSO regularisation (top, teal) or using grOWL regularisation (bottom, orange). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$).

be encoded in the angular gyrus, in posterior temporal cortex, or within modality-specific systems. Each theory has been supported by at least some multivariate fMRI studies and choice of method may account for the heterogeneity of these findings. Accordingly, the second aim of this study was to assess where else graded, multidimensional semantic similarity structure was evident in the brain, and whether choice of regularisation penalty impacted the results obtained.

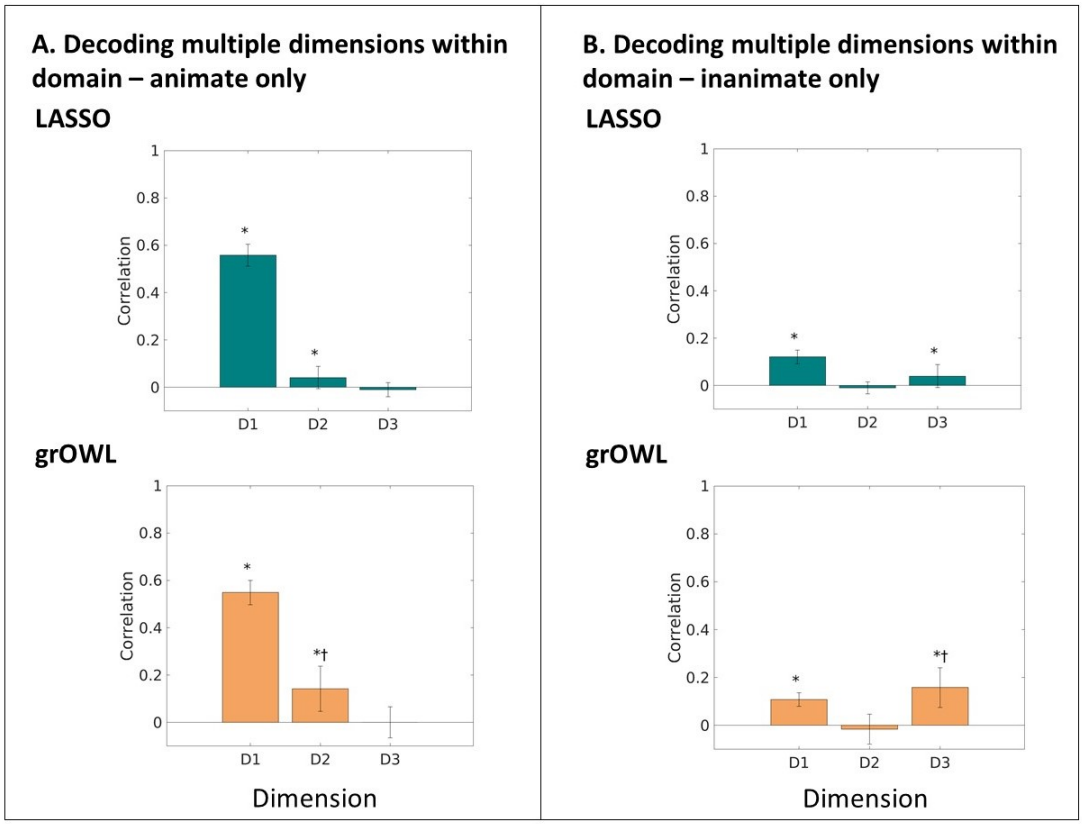

Figure 5.5: Hold-out correlations within-domain at the whole-brain level. Mean and 95 % confidence interval of the hold-out correlation for RSL models trained on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions (dimension 1, D1; dimension 2, D2; and dimension 3, D3). Models are trained using LASSO regularisation (top, teal) or using grOWL regularisation (bottom, orange). Stars (*) indicate a significant correlation between predicted and target coordinates (as determined by permutation test, probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Daggers (†) indicate a significant difference between correlation on that dimension using that regularisation penalty and correlation on the same dimension using the other regularisation penalty (probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (A) Hold-out correlations for animate stimuli only; (B) Hold-out correlations for inanimate stimuli only.

5.3.4.1 Is semantic structure present at the whole-brain level?

5.3.4.1.1 Binary animacy

Figure 5.4A shows the results from decoding animacy with logistic regression with LASSO or SOSLASSO regularisation. As in the ROI analysis, classification accuracy was very high (hold-out accuracy $\cong 0.95$). This result confirmed the presence of information about binary animacy.

5.3.4.1.2 Multidimensional semantic structure

Figure 5.4B shows results from decoding fine-grained semantic structure with RSL using LASSO

or grOWL regularisation trained on all 100 stimuli. Models fitted with both regularisation penalties showed very good decoding of the first target dimension ($r \cong 0.8$). However, evidence for *multidimensional* structure depended on the regularisation penalty – models fitted with grOWL produced significant correlations on dimension 2 and dimension 3 that, though small, were reliably better than the permutation mean, whereas models fitted with LASSO did not.

5.3.4.1.3 Graded semantic structure

Figure 5.5 shows the same correlations calculated for animate and inanimate stimuli separately. Despite the fact that only RSL with grOWL revealed multidimensional structure when all stimuli were considered together, both RSL with LASSO and RSL with grOWL produced significant correlations on two dimensions within-domain. For animate stimuli, these were dimension 1 and dimension 2 (which separates out the animate stimuli) and, for inanimate stimuli, the dimensions were dimension 1 and dimension 3 (which separates out the inanimate stimuli.

These findings confirmed the presence of graded structure at the whole-brain level.

5.3.4.2 Where is semantic structure present at the whole-brain level?

5.3.4.2.1 Binary animacy

To explore the spatial extent of the graded, multidimensional semantic representations that we discovered, we visualised coefficients from all models on the cortical surface. Figure 5.6 compares model coefficients from the regularised logistic regression analyses to the results from the whole-brain univariate analysis. As shown in Figure 5.6A, the univariate results highlighted posterior temporal, occipital and parietal regions. Regions showing significantly greater activation for animate than inanimate stimuli were interspersed with regions showing significantly greater activation for inanimate than animate stimuli, although, from the ventral perspective, there appeared to be a preference for animate stimuli laterally and for inanimate stimuli medially.

Figures 5.6B and 5.6C show the coefficients from logistic regression classifiers with LASSO (Figure 5.6B) or with SOSLASSO regularisation (Figure 5.6C). Both methods highlighted posterior temporal and occipitotemporal regions. On the whole, lateral occipitotemporal regions exhibited more consistently negative coefficients (meaning that, in most participants, a positive beta value is associated with an increased probability that the stimulus is animate). Medial regions exhibited more consistently positive coefficients. Lateral

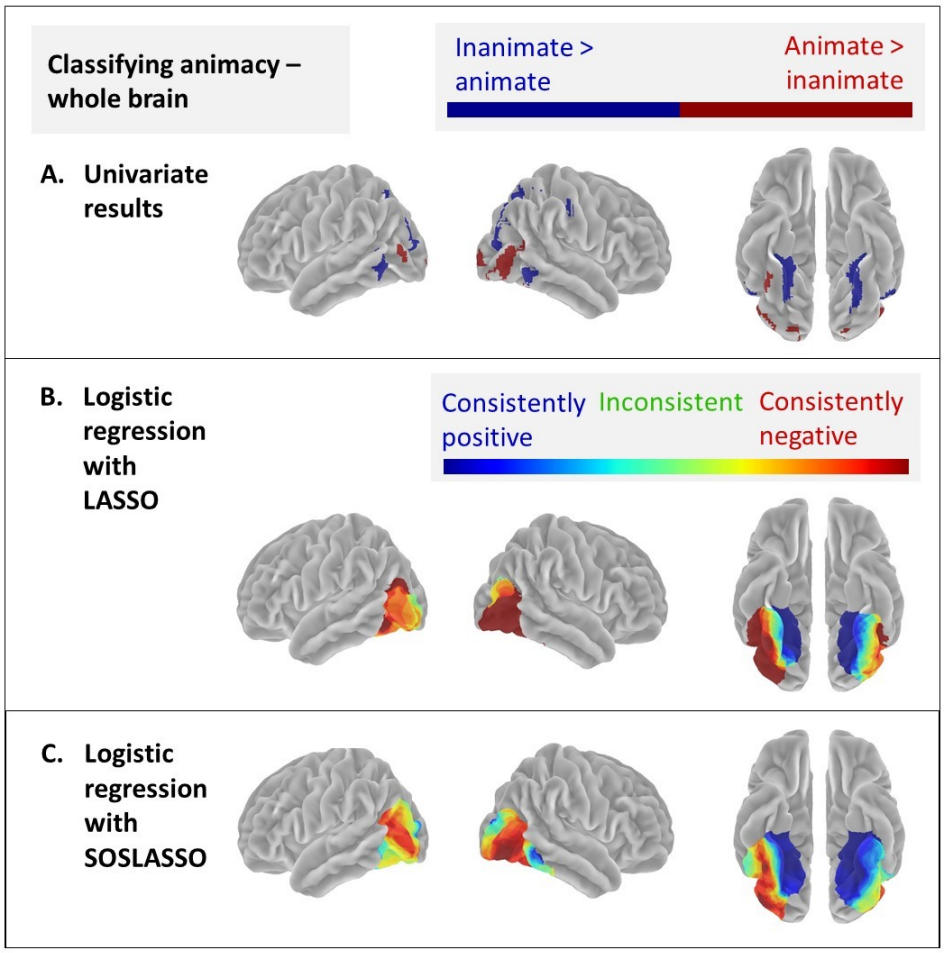

Figure 5.6: Coefficients for classification analyses. (A) Univariate effects (A>I, red; I>A, blue) are shown projected to the cortical surface and are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$. (B) Coefficients for logistic regression classifiers trained with LASSO regularisation on contrast beta values to discriminate animate from inanimate stimuli are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at $\alpha = 0.05$ (see Methods). Warm colours indicate that coefficients are mostly negative across participants (since animals were coded as 0 and inanimate objects as 1, a negative coefficient indicates that a positive beta value is associated with increased probability that the stimulus is animate); cool colours indicate that coefficients are mostly positive across participants (a positive beta value is associated with increased probability that the stimulus is inanimate); green shades indicate that the region was selected in multiple participants but that the sign of the coefficient was not consistent across participants. (C) Coefficients for logistic regression classifiers trained with SOSLASSO regularisation on contrast beta values to discriminate animate from inanimate stimuli are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at $\alpha = 0.05$ (see Methods). The colour scaling is the same as in (B).

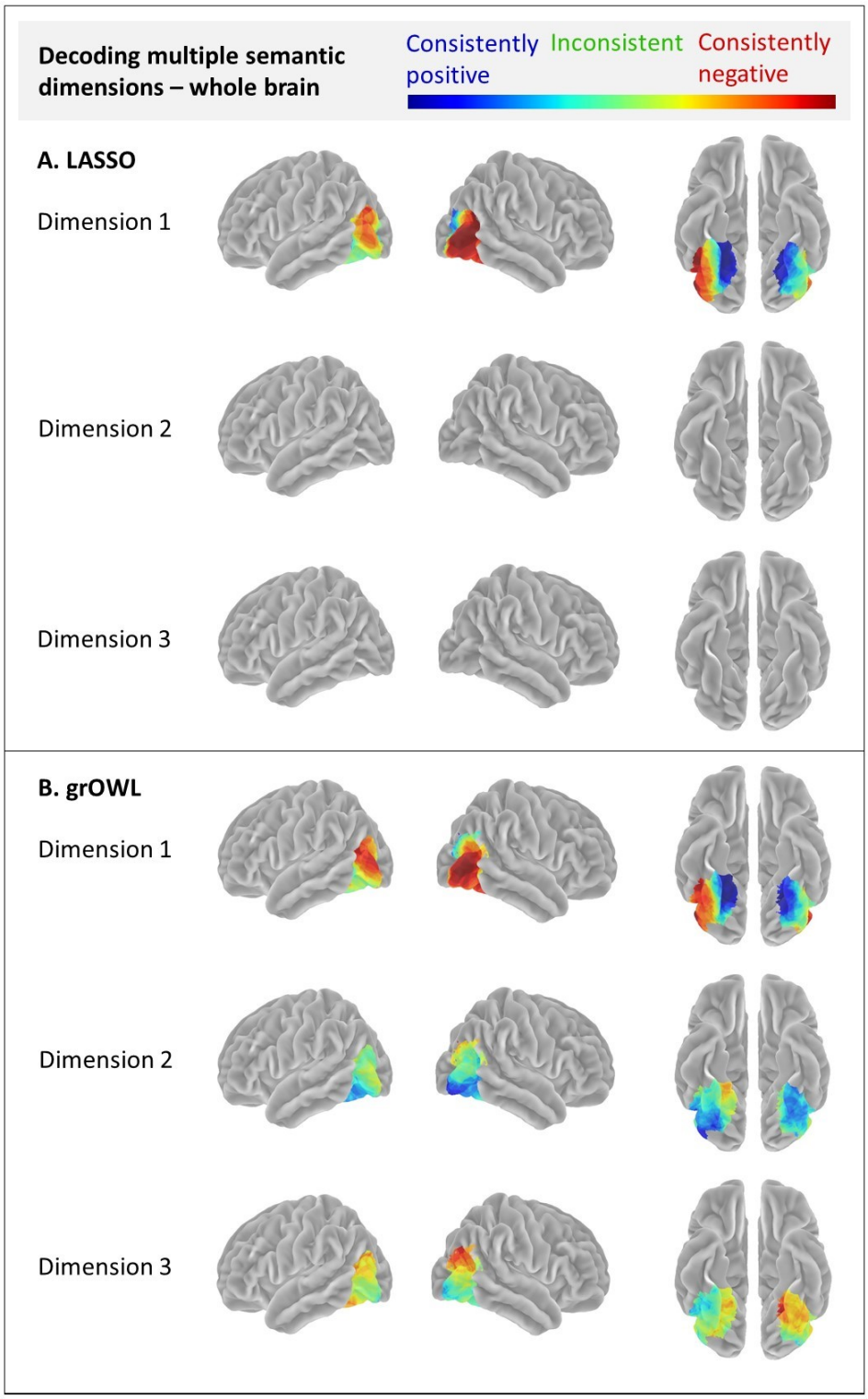
Decoding multiple semantic dimensions – whole brain
Consistently positive
Inconsistent
Consistently negative
A. LASSO
Dimension 1
Dimension 2
Dimension 3
B. grOWL
Dimension 1
Dimension 2
Dimension 3

Figure 5.7

Figure 5.7: Coefficients for RSL analyses. (A) Coefficients for RSL models trained with LASSO regularisation on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at $\alpha = 0.05$ (see Methods). Warm colours indicate that coefficients are mostly negative across participants; cool colours indicate that coefficients are mostly positive across participants; green shades indicate that the region was selected in multiple participants but that the sign of the coefficient was not consistent across participants. (B) Coefficients for RSL models trained with grOWL regularisation on contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions are shown projected to the cortical surface and thresholded via permutation testing to control the false discovery rate at $\alpha = 0.05$ (see Methods). The colour scaling is the same as in (A).

and medial regions were bridged by regions that were consistently selected across participants, but in which the sign of the coefficient was not consistent across participants (these regions were absent from the univariate results).

The vATL was selected in neither case.

5.3.4.2.2 Graded, multidimensional semantic structure

Figure 5.7 shows the coefficients from RSL models fitted with LASSO (Figure 5.7A) or with grOWL (Figure 5.7B). Again, posterior temporal and occipitotemporal regions were selected in both cases. For the first dimension, coefficients showed the same directional trends that were observed in the regularised logistic regression analysis – across participants, negative coefficients were consistently found in lateral regions and positive coefficients were consistently found in medial regions. Note that dimension 1 distinguishes between animate and inanimate stimuli (Figure 5.1), which is a likely explanation for this correspondence between decoding methods.

For the second and third dimensions, only models fitted with grOWL yielded coefficients that were in a sufficiently consistent location across participants to survive permutation thresholding. Again, posterior temporal and occipitotemporal regions were selected; although there were some regions in which coefficient directions were consistent in direction across participants (e.g. posterior ventrolateral occipitotemporal regions for dimension 2), there were also far greater swathes (compared to results for dimension 1) in which coefficients differed in direction across participants.

Again, the vATL was not selected by any analyses.

5.3.4.3 Does choice of decoding method challenge the discovery of semantic strucutre with fMRI?

One possible explanation for the heterogeneity of findings in the literature is the use of different decoding methods, each of which encapsulates different assumptions. We explored the impact of methodological decisions by comparing the effects of different regularisation penalties.

Figure 5.4A shows that there was no difference in classification accuracy when SOSLASSO instead of LASSO was used as the regularisation penalty. Figures 5.6B and 5.6C show that semantic similarity structure was highlighted in the same regions with both methods.

By contrast, the RSL results differed depending on the regularisation penalty. Figure 5.4B shows that only RSL with grOWL was successful at decoding more than one dimension when all stimuli were considered together; Figure 5.5 shows that RSL with LASSO could successfully decode two dimensions within-domain, but was outperformed by RSL with grOWL when decoding dimension 2 for animate stimuli and dimension 3 for inanimate stimuli (permutation $p<0.05$). Figure 5.7 shows that, when visualising coefficients on the cortical surface, models trained with grOWL yielded significant results for all three dimensions. However, for models trained with LASSO, coefficients for dimensions 2 and 3 did not survive permutation thresholding. Therefore, for RSL, there was a clear advantage of using grOWL rather than LASSO for regularisation in whole-brain analyses.

These RSL results supported the hypothesis that different methods (that make different assumptions) produce different patterns of results; this could therefore explain the heterogeneity of multivariate results in the literature.

5.4 Discussion

Convergent evidence from neuropsychology (Bozeat et al., 2000; Hodges & Patterson, 2007), computational modelling (Jackson et al., 2021; Rogers et al., 2004; Rogers & McClelland, 2004), noninvasive brain stimulation (Binney et al., 2010; Lambon Ralph et al., 2009; Pobric et al., 2007), and intracranial electrophysiology (C. R. Cox et al., 2024; Matoba et al., 2024; Rogers et al., 2021; Sato et al., 2021; Shimotake et al., 2015) supports the hub-and-spoke model of semantic cognition (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004). Functional magnetic resonance imaging (fMRI), however, yields mixed and contradictory results (Frisby et al., Chapter 2) that rarely highlight the vATL "hub" (e.g. Connolly et al., 2012; Devereux et al.,

2013, 2018; Huth et al., 2016; Pereira et al., 2018). This study aims to account for this discrepancy by addressing two important questions – (1) why fMRI often fails to discover semantic structure in the vATL, and (2) whether other regions of the brain encode semantic structure similar to that observed in the vATL. In line with convergent evidence from other methods, we discovered dynamic, graded, multidimensional semantic structure in the vATL, despite the threats to discovery posed by signal inhomogeneity, dynamic coding, and methodological assumptions. We also found evidence for graded, multidimensional semantic structure in posterior temporal and occipitotemporal cortex.

5.4.1 Question 1: Why does fMRI often fail to discover semantic structure in the vATL?

In this study, we discovered evidence for semantic representation in the vATL, both when using regularised logistic regression to decode binary animacy and when using RSL to decode fine-grained semantic structure. Furthermore, our RSL results support the claim that the semantic code is both graded and multidimensional. These results align with the predictions of the hub-and-spoke model (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004) and converge with evidence from neuropsychology (Bozeat et al., 2000; Hodges &, 2004), noninvasive brain stimulation (Binney et al., 2010; Lambon Ralph et al., 2009; Patterson, 2007), computational modelling (Jackson et al., 2021; Rogers et al., 2004; Rogers & McClelland, 2004; Pobric et al., 2007), and intracranial electrophysiology (Cox et al., 2024; Matoba et al., 2024; Rogers et al., 2021; Sato et al., 2021; Shimotake et al., 2015).

Our results diverge from other multivariate studies using fMRI, which rarely highlight the vATL (e.g. Connolly et al., 2012; Devereux et al., 2013, 2018; Huth et al., 2016; Pereira et al., 2018). The first aim of our study was to explore three possible reasons for this difference. The first possible reason was signal inhomogeneity, caused by the proximity of the vATL to the air-filled sinuses. To remediate this challenge, we employed a multi-echo, multiband acquisition sequence. We subsequently found reliable decoding of both binary animacy information and graded, multidimensional semantic similarity structure even in the most anterior parts of vATL. Thus, while signal inhomogeneity may have obscured such structure in previous work, the current study suggests that this barrier can be overcome with multi-echo, multiband acquisition. Collecting data without distortion correction (to enable a direct comparison) was beyond the scope of this study; others have directly tested the link between acquisition sequence and

correlation-based decoding accuracy (Frisby et al., Chapter 4; Halai et al., 2024) and future studies should extend this test to decoding methods that use machine learning (including regularised logistic regression and RSL).

The second possible reason we explored was that the limited temporal resolution of fMRI may be insufficient to detect a code that changes dynamically from millisecond to millisecond (Rogers et al., 2021). Although animacy information was reliably decoded in both halves of the ROI, hold-out accuracy was significantly lower in the anterior half-ROI than the posterior half-ROI. This result contrasts with previous results using ECoG, which found equally good decoding in posterior and anterior areas despite the rapidly changing code observed in the more anterior portion of the ROI. The discrepancy in accuracy between halves in fMRI is thus consistent with the view that dynamic coding in anterior vATL may somewhat hinder the discovery of semantic structure with fMRI. However, accuracy in the anterior half-ROI did not drop below the threshold for significance, suggesting that dynamism is not a complete barrier to signal discovery and so is unlikely to account for the consistent absence of the vATL in previous studies.

The third possible reason we explored is choice of method – multivariate methods each encapsulate hypotheses about representational structure that constrain the kind of neural code that they can detect and these assumptions can lead to drastically different results (Frisby et al., Chapter 2). Within the ROI, logistic regression with SOSLASSO regularisation yielded higher classification accuracy than logistic regression with LASSO. In our RSL analysis, models fitted with grOWL systematically decoded dimensions 1 and 2 within the animate domain and dimensions 1 and 3 within the inanimate domain, whereas models fitted with LASSO exhibited a less reliable pattern of results. Additionally, any significant differences in correlations between the methods were always in favour of grOWL. Together these findings support the conclusion that different analytic approaches can lead to differing conclusions about the role of vATL in semantic processing.

Moreover, SOSLASSO and grOWL make assumptions that are explicitly based on proposals about how neural systems represent information (C. R. Cox et al., 2024; C. R. Cox & Rogers, 2021; Appendix E). Therefore, the current results suggest that, when multivariate decoding models adopt gentle and neurally-inspired assumptions, results of fMRI analysis cohere better with results from other methods. However, as argued in Frisby et al. (Chapter 2), the most important consideration when choosing a decoding strategy should be the kind of neural code

one expects to detect, and other researchers' expectations about the nature of the neural code may differ from our own expectations. Therefore, the best regularisation penalty for future studies is the regularisation penalty that aligns with those individual researchers' explicit hypotheses.

5.4.2 Question 2: Which (if any) other regions of the brain encode semantic structure?

The whole-brain analysis identified graded, multidimensional semantic similarity structure (similar to that discovered in the vATL in the ROI analysis) in posterior temporal and occipitotemporal cortex. Note that multivariate methods provide an insight into a further property of this representation – inconsistency (meaning that representation of the same information is associated with different directions of change in activation across individuals; Frisby et al., Chapter 2). Both regularised logistic regression and RSL highlighted posterior temporal regions that were consistently selected across participants, but in which the direction of the model coefficient varied between participants (as shown by green shades in Figure 5.7B). These regions were not evident in the univariate analysis, which is capable of finding signal only when it is consistent across participants; this result therefore reinforces the utility of multivariate approaches for revealing, not only where semantic structure is represented, but how.

Visualising the spatial extent of this graded, multidimensional semantic structure enabled us to compare our findings to the predictions of different theories of semantic representation (Binder & Desai, 2011; A. R. Damasio, 1989; H. Damasio et al., 2004; Huth et al., 2016; Lambon Ralph et al., 2017; A. Martin, 2007, 2016; Patterson et al., 2007; Rogers et al., 2004). Some of these theories do predict the selection of the posterior temporal and occipitotemporal cortex – some propose that it contains multimodal "convergence zones" (e.g. Binder & Desai, 2011; A. R. Damasio, 1989; H. Damasio et al., 2004), while others propose that it represents visual semantic features (A. Martin, 2016). Other theories implicate the posterior temporal cortex, but as part of a broader network including the angular gyrus (Binder & Desai, 2011) or indeed most of the cortex (e.g. Huth et al., 2016); the absence of those additional regions means that our results support those theories less well. At first glance, the absence of the vATL suggests that our results also align poorly with the predictions of the hub-and-spoke model – however, our ROI analysis establishes that semantic structure is present within the vATL. Note that most tests of the hub-and-spoke model have been focused on the vATL hub – relatively little

work has attempted to characterise representational structure within the spokes, a set of regions which may include the posterior temporal and occipitotemporal cortex. Preliminary evidence from computational implementations of the hub-and-spoke model suggests that spoke regions do come to represent overall semantic similarity structure (rather than similarity structure within a particular modality) during semantic tasks (L. Chen et al., 2017; Jackson et al., 2021), but it may be only the spoke that is processing the sensory input that does so (C. R. Cox, 2016). The results from our study are consistent with these exploratory findings – we presented visual stimuli and found semantic similarity structure in visual association cortex. Importantly, our results do not adjudicate between the competing theories of semantic representation; future studies should be designed in such a way that they can do so (Frisby et al., Chapter 2).

The absence of the vATL from the coefficient maps in Figures 5.6 and 5.7 could be explained by multiple methodological factors. First, the regularisation penalties we used encapsulated the assumption of sparsity – non-zero coefficients were placed on only a small subset of the possible array of features – and may therefore have selected only a subset of the important signal. The degree of sparsity is dictated by the range of possible λ values; since this study represents the first application of SOSLASSO and grOWL to 7T-fMRI, we used toolbox defaults, but this may have produced an "over-sparse" solution. Future work should seek to optimise these hyperparameter ranges. Second, our multi-echo, multiband acquisition sequence did not mitigate the effects of signal dropout and distortions perfectly (temporal signal-to-noise ratio is shown in Appendix F, Figure F.1); sparse models given the whole brain from which to choose features may prefer to select voxels in which BOLD tracks the changes in stimuli more faithfully. Third, we found evidence for dynamic representation of semantic structure (animacy) in the anterior half-ROI; sparse models that are given the whole brain may prefer voxels that represent semantic structure in a temporally stable fashion.

There is also a fourth reason for the absence of the vATL. Recall that we employ permutation testing to evaluate when a voxel is selected by a decoding model more often than expected from a null distribution. In our decoding analyses, particularly in RSL, voxels in the vATL are selected far more frequently by models trained on permuted data than are voxels elsewhere in the brain (C. R. Cox, 2016; permutation distributions are shown in Appendix F, Figures F.2 and F.3). This means that coefficients in the vATL are unlikely to survive permutation thresholding even when they are selected by most real models. The reason for this phenomenon is unknown and may be a consequence of signal inhomogeneity (cf. Appendix F, Figure F.1).

To summarise, the current results suggest that, when hyperparameters are fine-tuned, acquisition sequences are further improved, and the reasons for the spatial properties of the permutation distributions are clarified, coefficient maps from decoding analyses may highlight the vATL and thereby confirm the predictions of the hub-and-spoke model (Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004). By contrast, there are no obvious methodological reasons why regions that others have predicted to be important, such as the angular gyrus (Binder & Desai, 2011), are not evident in this analysis.

We also assessed whether differences in methodology could account for heterogeneous results in the literature. In the whole-brain analysis, there was no clear benefit of using SOSLASSO for regularised logistic regression rather than LASSO. Accuracy was comparable with both methods, possibly because it was already at ceiling. The spatial distribution of coefficients was the same with both methods, in contrast to a previous study that discovered a much more anatomically extensive network distinguishing face from nonface stimuli using SOSLASSO than using LASSO (C. R. Cox & Rogers, 2021). When the barriers described above (including "over-sparsifying" and nonuniform permutation distributions) are addressed, regularised logistic regression with SOSLASSO may reveal a more widespread network coding animacy. Turning to RSL, models fitted with grOWL revealed representation of all three dimensions in posterior temporal and occipitotemporal cortex. By contrast, models for LASSO revealed a maximum of two dimensions and no coefficients for dimensions 2 and 3 survived permutation thresholding. As in the ROI analysis, these results suggest that the gentle and neurally-inspired assumptions made by grOWL can benefit decoding. In summary, differences between analysis methods were observed even with an "assumption-light" decoding approach based on regularised regression; methodological assumptions are therefore a plausible explanation for the heterogeneity of findings in the literature (Frisby et al., Chapter 2).

5.5 Conclusion

This work addressed address two important questions – (1) why fMRI often fails to discover semantic structure in the vATL, and (2) whether other regions of the brain encode semantic structure similar to that observed in the vATL. It provided two important answers – (1) signal inhomogeneity and methodological assumptions are the most likely explanations of the vATL's absence from previous decoding studies using fMRI; and (2) 7T-fMRI can also detect

representation of graded, multidimensional semantic structure in posterior temporal and occipitotemporal cortex; a discovery that future iterations of the hub-and-spoke model should seek to characterise in detail.

Data and code availability: Data will be made publicly available upon peer review and acceptance. Code is publicly available at: https://github.com/slfrisby/7TConvergent/.

Acknowledgements: This work was supported by an EPS study visit grant (G118497) to S. L. F., by an MRC Career Development Award (MR/V031481/1) to A. D. H., by an MRC programme grant (MR/R023883/1) and intramural funding (MC_UU_00005/18) to M. A. L. R., and by an NSF grant (NCS-FO 219903 DRL) to T. T. R. We would like to thank the participants and the Wolfson Brain Imaging Centre radiographers.

Competing interests: The authors declare no conflicts of interest.

Supplementary material: Supplementary methods can be found in Appendix E and supplementary results can be found in Appendix F.

Chapter 6

General Discussion

Representation of semantic information enables us to engage with the world in a meaningful way – to comprehend and produce language, identify and use objects, and understand and participate in events that involve us. This thesis aimed (1) to develop a theoretical framework with which to describe the nature of neural representations, including semantic representations; (2) to assess and compare the capacities of electrocorticography (ECoG) and 7 tesla functional magnetic resonance imaging (7T-fMRI) to detect semantic representations; and (3) to evaluate the strengths and limitations of multivariate analysis methods, in particular methods based on regularised regression, for revealing properties of semantic representations.

This Chapter serves three purposes – to summarise the results from Chapters 2, 3, 4 and 5 (since each Chapter functions as a standalone paper or manuscript, detailed discussions can be found within each Chapter), to highlight themes that span the four Chapters, and to propose directions for future research.

6.1 Chapter summary

6.1.1 Chapter 2: Decoding semantic representations in mind and brain: a theoretical framework and review

In Chapter 2, I reviewed theories of semantic representation and proposed a new theoretical framework for describing representations. Specifically, I posed six questions about the computational and neural characteristics of representations. By answering these questions, representations can be classified as:

1. *Category* representations (composed of discrete, independent units that each correspond to a concept), *feature-based* representations (composed of multiple independently-interpretable features), or *vector-space* representations (composed of a pattern across representational units, the meanings of which cannot be independently interpreted)

2. *Self-contained* (encapsulating semantic information within itself such that mere activation of the representation brings about retrieval/inference) or *grounded* (requiring the generation of modality-specific surface representations to produce retrieval/inference)

3. *Homogeneous* (consisting of units that all adopt the same activation state when representing a concept) or *heterogeneous* (consisting of units that adopt different activation states when representing a concept)

4. *Consistent* (associated with the same direction of change in activation across individuals) or *inconsistent* (associated with different directions of change in activation across individuals)

5. *Independent* (consisting of units that express the presence or absence of the same semantic information irrespective of the states of other units) or *conjoint* (consisting of units that express different semantic information depending on the states of other units)

6. *Contiguous* (composed of units residing in the same brain region) or *dispersed* (composed of units residing in different brain regions; note that representations can be contiguous or dispersed within or across individuals as shown in Figure 2.2C)

This framework crystallises the predictions or assumptions that any theory makes, thereby situating each theory relative to others in the field and making the vast array of different approaches tractable.

Having laid out the space of possible cognitive and neural representations, I assessed the capabilities of popular multivariate methods for revealing each type. Crucially, I demonstrated that each method encapsulates assumptions (not always made explicit) about how the brain represents information, which means that there are types of code to which individual methods are insensitive. I explained how multivariate pattern classification approaches are, by nature, assumption-free, but must be used in conjunction with a feature selection method (such as an ROI or searchlight) that restricts the types of code they can detect. I explained how encoder/decoder approaches (also known as generative approaches) fail to detect some conjoint codes and how they depend upon feature selection for interpretability. I explained how representational similarity analysis can detect category, feature-based and vector-space representations, but only when an appropriate target similarity matrix is used, and how it can detect dispersed representations only when it is *not* (as is almost always the case) used in conjunction with ROIs or searchlights.

To conclude, in this Chapter I suggested why findings in the literature are often contradictory; identified crucial links between cognitive theory, data collection, and analysis that can help to better connect neuroimaging to mechanistic theories of semantic cognition; and provided a field guide to contemporary multivariate methods for brain imaging.

6.1.2 Chapter 3: All spectral frequencies of neural activity reveal sematic representation in the human anterior ventral temporal cortex

In Chapter 3 I probed the temporal dynamics of semantic representations. Most previous work decoding semantic information from ECoG data has focused on voltage (Y. Chen et al., 2016; C. R. Cox et al., 2024; H. Liu et al., 2009; Nagata et al., 2022; Rogers et al., 2021) or spikes (Kraskov et al., 2007; Reber et al., 2019). The role of time-frequency power and phase in semantic representation has been investigated far less thoroughly (cf. Clarke, 2020; Rupp et al., 2017; Wang et al., 2011). In this study I asked (1) whether semantic information could be decoded from time-frequency power and/or phase, and, if so, from which frequency bands; and (2) whether this code exhibited the same "deep, distributed, dynamic" properties previously observed both in ECoG voltage data and in a neural network model of semantic representation (Rogers et al., 2004, 2021). Specifically, I explored the following four properties:

1. *Constant decodability*. Neural activity predicts stimulus category at every time point once activation reaches the vATL "hub".

2. *Local temporal generalisation*. Classifiers generalise best to time windows close to the window on which they were trained and more poorly to time windows further away from the training time.

3. *Widening generalisation window*. The temporal window over which classifiers generalise grows wider over time.

4. *Change in code direction*. Increased or decreased neural activity can signify different semantic information at different points in time.

I studied ECoG data recorded from grid electrodes on the cortical surface of the vATL of patients (n = 19) undergoing surgery for intractable epilepsy. I devised a bespoke, clinically-informed preprocessing pipeline to ready the data for decoding (explored further in

Appendix C) and decoded animacy using logistic regression with LASSO regularisation (Tibshirani, 1996) – a method that encapsulates very few assumptions about the nature of the neural code and, importantly for this study, is ideally suited to assessing whether the time-frequency code (like the neural network model) exhibits local temporal generalisation and/or a widening generalisation window.

I found (1) that it is possible to decode semantic information from power in every frequency range between theta and high gamma and, although individual frequency ranges contained enough information to drive decoding above chance, decoding from the whole frequency spectrum at once produced the highest decoding accuracy; (2) that only when the decoder was trained on power from all frequencies at once did the full range of deep, distributed, dynamic properties observed by Rogers et al. (2021) emerge. From these findings I concluded that semantic information is not simply present redundantly in multiple frequencies (Rupp et al., 2017; Wang et al., 2011), but is represented in such a way that multiple frequencies must be considered together for the full set of properties to be revealed (termed a *transfrequency* representation). If this conclusion is correct, then to speak of a transfrequency code and a voltage code is simply to describe the same phenomenon in two different ways (cf. Miller, 2010).

6.1.3 Chapter 4: Optimising 7T-fMRI for imaging the anterior temporal lobe

In Chapter 4 I laid the foundations for studies investigating semantic representations with 7T-fMRI (and thus capitalising on its improved tSNR relative to 3T-fMRI; Morris et al., 2019). The vATLs are located next to the air-filled sinuses and so are affected by magnetic field inhomogeneity and the resulting signal drop-out and distortions; methods for tackling these problems (Halai et al., 2014, 2024) had not yet been fully evaluated at 7T (cf. Ding et al., 2022). I compared three methods of improving sensitivity in the vATLs: (1) parallel transmit, which uses multiple transmit elements, controlled independently, to homogenise the B_1 pulse applied to the tissue; (2) multi-echo, which entails collection of multiple volumes at different echo times following a single radiofrequency pulse and opens the door to denoising via multi-echo independent component analysis (ME-ICA); and (3) multiband, in which multiple slices are acquired simultaneously, which can be translated into increased statistical power.

I found that parallel transmit improved activation magnitude in posterior temporal/occipital regions while multi-echo improved activation magnitude across multiple areas of the semantic network. Both multiband and ME-ICA denoising resulted in improved

activation precision extending rostrally along the temporal lobe and including frontal regions. In an exploratory analysis I found that multi-echo and ME-ICA improved decoding of task condition. I demonstrated that a multi-echo, multiband sequence can detect signal in the vATL while maintaining sensitivity across the whole brain and is therefore a versatile choice for future studies using high field strength and investigating the functional roles of the vATL, including the study in Chapter 5.

6.1.4 Chapter 5: Decoding semantics with 7T-fMRI: Convergent evidence and divergent discovery

In Chapter 5 I assessed the capabilities of 7T-fMRI for revealing semantic representations. Convergent evidence from neuropsychology (Bozeat et al., 2000; Hodges & Patterson, 2007; Snowden et al., 1989; Warrington, 1975), computational modelling (Jackson et al., 2021; Rogers et al., 2004; Rogers & McClelland, 2004), noninvasive brain stimulation (Binney et al., 2010; Lambon Ralph et al., 2009; Pobric et al., 2007), and ECoG (C. R. Cox et al., 2024; Matoba et al., 2024; Rogers et al., 2021; Sato et al., 2021; Shimotake et al., 2015) supports the theory that the vATL functions as a crucial semantic "hub". Multivariate studies of fMRI data, however, produce results that contradict each other and that rarely highlight the vATL (e.g. Connolly et al., 2012; Devereux et al., 2013, 2018; Huth et al., 2016; Pereira et al., 2018). In this study I asked (1) why fMRI often fails to discover semantic structure in the vATL, and (2) whether other regions of the brain encode semantic structure similar to that observed in the vATL.

I acquired 7T-fMRI data, using the novel whole-brain acquisition sequence developed in Chapter 4, while healthy participants named the same pictures that the ECoG patients named. I applied four decoding methods, designed to ensure that assumptions about the nature of the neural code were minimised and well-justified. To decode animacy I used regularised logistic regression with LASSO regularisation and with sparse-overlapping-sets LASSO (SOSLASSO) regularisation (C. R. Cox & Rogers, 2021), and to decode fine-grained semantic similarity structure I used representational similarity learning (RSL) with LASSO regularisation and with group-ordered-weighted LASSO (grOWL) regularisation (C. R. Cox et al., 2024).

Applying each of these methods within an ROI defined based on the coverage of the ECoG electrodes in Chapter 3, I established that dynamic (Rogers et al., 2021), graded, multidimensional (Clarke, 2020; C. R. Cox et al., 2024) semantic representations in the vATL are visible with 7T-fMRI. I concluded that my use of a distortion-corrected acquisition sequence

(Chapter 4) and assumption-light analysis methods are the most likely reasons for the difference between this study and previous work. In particular, I found that both logistic regression with SOSLASSO regularisation and RSL with grOWL regularisation improved decoding performance relative to either method with standard LASSO, and that RSL with grOWL enabled the detection of more dimensions of representation.

Applying the four decoding methods to the whole brain, I discovered graded, multidimensional semantic structure in the posterior temporal and occipitotemporal cortex. Most tests of the hub-and-spoke model have focused on the vATL "hub" (cf. L. Chen et al., 2017; C. R. Cox, 2016; Jackson et al., 2021), so the role of the posterior temporal and occipitotemporal cortex warrants more detailed characterisation. This whole-brain analysis did not highlight semantic representations in the ATL, which may be because the vATL is selected with disproportionate frequency in the permutation distribution (Appendix F, Figures F.2 and F.3), or because my decoding models assume that semantic representations are sparser than they actually are, prefer to select voxels in regions free of magnetic field inhomogeneities, and/or prefer to select voxels in regions with temporally stable codes. As in the ROI analysis, I found that RSL with grOWL revealed more dimensions of representation than RSL with LASSO did (there was no benefit of using SOSLASSO for logistic regression, possibly because performance was at ceiling with both SOSLASSO and standard LASSO).

To conclude, this study's findings converge with the hub-and-spoke model by detecting semantic representations in the vATL and diverge by detecting multidimensional semantic structure in posterior temporal and occipitotemporal cortex – a discovery that future iterations of the hub-and-spoke model should seek to characterise in detail.

6.2 Common themes

6.2.1 How is semantic information represented in the brain?

The capability to discern how the brain *does* represent semantic information depends critically on our ability to articulate hypotheses about how the brain *could* represent semantic information (and then to adjudicate between those hypotheses). The theoretical framework proposed in Chapter 2 breaks this daunting question down into six smaller and more precise questions about the computational characteristics of representations and their instantiation by one or more spatially distinct neural populations. Although this characterisation of representations is not

exhaustive – i.e., there are more than six questions that can and should be asked about representations – it represents a systematic, unifying theoretical framework in which to situate theories about, and methods for discovering, semantic representations in the brain.

The analytic approach in Chapter 5 was designed so as not to pre-empt the answers to any of the six questions laid out in Chapter 2. The study in Chapter 5 also addressed one of those questions directly – I found evidence that semantic representations are graded and multidimensional, which suggests that semantic dimensions are not category representations (e.g. Collins & Loftus, 1975; Collins & Quillian, 1969; Rosch, 1975). However, because neither grOWL nor LASSO dictates that the brain's similarity space must be axis-aligned with the target similarity space (Appendix E), it is unclear (from these results alone) whether semantic representations are feature-based representations (with interpretable dimensions; A. J. Anderson, Binder, et al., 2019; Cree et al., 1999; Tyler et al., 2000) or vector-space representations (with uninterpretable dimensions; Lambon Ralph et al., 2017; Patterson et al., 2007; Rogers et al., 2004). The whole-brain analyses in Chapter 5 also suggest preliminary answers to four of the other questions in Chapter 2. First, the whole-brain analysis highlighted posterior temporal and occipitotemporal regions, which are known to be specialised for processing visual information; this suggests that semantic representations are grounded (Barsalou, 2003; Glenberg, 2010; Glenberg & Robertson, 2000). Second, the whole-brain maps feature regions where coefficients are positive and regions where coefficients are negative (across participants; Figures 5.6 and 5.6), suggesting that semantic representations are heterogeneous. Third, the whole-brain maps feature some regions which were consistently selected across participants, but in which coefficients were not consistent in direction across participants, suggesting that semantic representations are inconsistent across individuals. Fourth, semantic information could be detected from the vATL (in the ROI analysis) all the way to posterior occipitotemporal cortex (in the whole-brain analysis). However, coefficients were placed in similar neural locations in each participant (even when analysis methods did not encapsulate the assumption that they would be), which suggests that the semantic code is dispersed within individuals, but relatively contiguous across individuals (Binder & Desai, 2011; H. Damasio et al., 2004; Lambon Ralph et al., 2017). However, these conclusions are tentative for reasons explored below.

Chapter 2 focused primarily on the computational and spatial properties of semantic representations. Chapter 5 suggests that semantic representations are also dynamic (Clarke, 2020; Jones et al., 2015; Yuste, 2015). Chapter 3 probes their temporal characteristics in detail.

The results of Chapter 3 suggest that the semantic code is not only deep, distributed, and dynamic, it is also *transfrequency* – when (and only when) all frequencies are considered together, it exhibits local temporal generalisation, a widening generalisation window, and changes in code direction, just like a parallel distributed processing instantiation of the hub-and-spoke model (Rogers et al., 2004, 2021; Rogers & McClelland, 2004). This finding is not fully accounted for by theories proposing that semantic information is present redundantly in multiple frequencies (Lam et al., 2016; Rupp et al., 2017), or that different frequencies are used to transmit the same information over different cortical distances (Canolty & Knight, 2010; Lachaux et al., 2012). This result raises the question of whether it is wise to think of the semantic code as a frequency code at all. Since voltage reflects the sum of activity at all frequencies, perhaps we do not need to posit a code more complex than the voltage code characterised by (Rogers et al., 2021) – the presence of semantic information in time-frequency power does not entail that it is being used for computation (Watrous, Fell, et al., 2015).

6.2.2 Where are semantic representations?

In Chapter 3, I used ECoG to search for semantic information in the vATL and, unsurprisingly given the wealth of evidence of its importance from patient studies (e.g. Hodges & Patterson, 2007; Noppeney et al., 2006; Snowden et al., 1989; Tyler et al., 2000; Warrington, 1975; Warrington & Shallice, 1984), brain stimulation (Binney et al., 2010; Lambon Ralph et al., 2009; Matoba et al., 2024; Pobric et al., 2007, 2010a, 2010b), computational modelling (L. Chen et al., 2017; Rogers et al., 2004) and other imaging (e.g. Mollo et al., 2017), I was successful in decoding animacy. The 7T-fMRI results in Chapter 5 converge with these results converge with Chapter 3 – semantic information was decodable even from the anterior half of the temporal lobe ROI. This finding was a positive surprise given the vATL's absence in many multivariate analyses of fMRI data (Connolly et al., 2012; Devereux et al., 2013, 2018; Huth et al., 2016; Pereira et al., 2018).

However, Chapter 5 also highlights the representation of semantic information in posterior occipitotemporal regions. This result is compatible with several theories of semantic representation (L. Chen et al., 2017; C. R. Cox, 2016; A. R. Damasio, 1989; H. Damasio et al., 2004; Jackson et al., 2021; Lambon Ralph et al., 2017; A. Martin, 2016; Patterson et al., 2007; Rogers et al., 2004) and highlights an intrinsic limitation of the methods employed in this thesis. Semantic representations both express conceptual similarity structure *and* enable retrieval and

inference. Methods such as RSL are ideally suited to identifying regions expressing conceptual similarity structure, but the *necessity* and *sufficiency* of those areas for retrieval and inference cannot be established. Convergence with methods that test the capabilities of the brain when damaged – either naturally (Hodges & Patterson, 2007; Noppeney et al., 2006; Snowden et al., 1989; Tyler et al., 2000; Warrington, 1975; Warrington & Shallice, 1984) or artificially and reversibly (Binney et al., 2010; Lambon Ralph et al., 2009; Pobric et al., 2007, 2010a, 2010b) – is essential.

6.2.3 What sorts of evidence can be brought to bear on the nature of semantic representations?

In this thesis I use two imaging modalities to investigate semantic representations – ECoG and 7T-fMRI. Chapter 3 delved deeper into this issue, asking in which derivatives of ECoG data semantic information could be detected – time-frequency power, time-frequency phase, and/or voltage. Semantic information was found both in time-frequency power (Lam et al., 2016; Merker, 2013; Rupp et al., 2017; Solomon et al., 2019) and voltage (Y. Chen et al., 2016; C. R. Cox et al., 2024; H. Liu et al., 2009; Nagata et al., 2022; Rogers et al., 2021). (It was not evident in time-frequency phase, despite perspectives that phase might represent semantic information either directly (Clarke, 2020; Clarke et al., 2018) or jointly with power via phase-amplitude coupling (Aru et al., 2015; Canolty et al., 2006; Canolty & Knight, 2010; Hermes et al., 2014; Lisman & Jensen, 2013; Sederberg et al., 2006).) However, it was only when all frequencies were considered together that time-frequency data exhibited the same range of deep, distributed, dynamic properties identified in the voltage code by Rogers et al. (2021). This raises the question of whether time-frequency decomposition of ECoG data is a useful method for studying semantic representation in the vATL – since voltage reflects the sum of activity at all frequencies, studying voltage may be a more parsimonious method that leads to the same conclusions. If voltage is chosen as the subject of study, preprocessing pipelines should be adapted to preserve information contained in voltage (Appendix C).

Chapter 5 (underpinned by Chapter 4) established that both binary category information and graded, multidimensional semantic structure could be decoded from 7T-fMRI data, just as they could from ECoG data (C. R. Cox et al., 2024; Rogers et al., 2021). Although previous research questioned whether the dynamism of the semantic code would render it invisible to fMRI (Rogers et al., 2021), this result demonstrates that some aspects of semantic

representations (such as their dimensionality) can be studied noninvasively in healthy participants. Chapter 5 also demonstrates that whole-brain imaging can reveal aspects of semantic representations to which ECoG (with its limited spatial coverage) is insensitive. Of course, fMRI is not suited to addressing every research question – for example, the dynamic properties of the semantic code, as revealed by ECoG, warrants further characterisation and, for this, time-resolved methods are required. In addition to ECoG, whole-brain, time-resolved methods that can be applied in large samples of healthy participants, such as magnetoencephalography (MEG), will be invaluable.

This thesis also explored the impact of analytical decisions on the discovery of semantic representations. Chapter 2 highlighted the link between theory and analysis, illustrating the impact of analytic choices on which types of representation can be discovered. Chapter 5 explored the same issue from a practical perspective, comparing the impact of different regularisation penalties on the same 7T-fMRI dataset. Regularisation penalties that incorporate explicit, gentle, thoughtful assumptions about the nature of the semantic code (for example, the assumption that semantic representations are located in roughly, though not exactly, the same place within and across individuals; C. R. Cox & Rogers, 2021) can produce more accurate classification and/or larger cross-validated correlations, discover more dimensions of representation, and reveal the spatial properties of representation more faithfully than penalties that do not incorporate prior knowledge. However, the results of Chapter 5 should not be taken to mean that incorporating more assumptions is always beneficial to decoding. As illustrated in Chapter 2, many assumptions implicit in analysis approaches will blind a study to representations of certain types – for example, encoder models (e.g. A. J. Anderson et al., 2021; Fernandino, Humphries, et al., 2016; Huth et al., 2012, 2012; Mitchell et al., 2008; Tang et al., 2021) cannot detect some kinds of conjoint codes and the use of ROIs for feature selection (e.g. Asyraff et al., 2021; Devereux et al., 2018; Dijkstra et al., 2021; Liuzzi et al., 2021; Wurm & Caramazza, 2019) prohibits the detection of extremely dispersed codes.

Additionally, the results of Chapter 5 should not be taken to mean that SOSLASSO and grOWL should be applied indiscriminately. As argued in Chapter 2, different analysis methods are suited to answering different experimental questions. Therefore, the best course of action is for researchers to select a method that is capable of detecting the kind of semantic code that they hypothesise to exist (and, where possible, methods that can adjudicate between competing theories). This is exemplified in Chapter 3, in which I selected LASSO as the regularisation

penalty because it produces very sparse solutions in which many features receive coefficients of 0. If the information used by a classifier trained with LASSO regularisation at time t is present at time $t \pm n$, the classifier will perform well at time $t \pm n$; other units, which may be in different states at the two timepoints, will receive coefficients of zero and therefore will not affect classifier performance (Rogers et al., 2021). Therefore, although LASSO did not prove to be the best regularisation penalty for revealing multidimensional semantic structure in 7T-fMRI data, it was an ideal method to use for addressing questions about local temporal generalisation (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; King & Dehaene, 2014). and widening generalisation windows (Rogers et al., 2021).

6.3 Future directions

6.3.1 Improvements to methodology

The findings in this thesis reflect the knowledge and technologies available at the time of writing rather than the limits of what is possible. While conducting these experiments I encountered diverse methodological challenges, to wit:

1. *Incomplete cleaning of saccades and microsaccades from ECoG data.* Saccadic and microsaccadic activity has the potential to contaminate ECoG (Clarke, 2020; Jerbi et al., 2009; Kovach et al., 2011; Yuval-Greenberg et al., 2008) and animate and inanimate stimuli may elicit different patterns of eye movement (G. W. Humphreys & Forde, 2001; Appendix C); therefore, there is a notional possibility that saccadic and microsaccadic activity could contribute to decoding performance. Although I assessed whether animacy could be decoded from EOG data alone (Appendix C, Figure C.2), the sample size was not large enough to produce conclusive results. In addition, our novel, clinically-informed preprocessing pipeline featured an ICA-based method of microsaccade detection and removal (Clarke, 2020) that failed to remove any microsaccades. There is scope for improvement to this procedure.

2. *Suboptimal tSNR of 7T-fMRI data.* The multi-echo, multiband 7T-fMRI acquisition sequence optimised in Chapter 4 detects univariate contrast extending into the vATL (Appendix D, Figure D.8), however, it fails to restore temporal signal-to-noise to the levels achieved across the rest of the brain (Appendix F, Figure F.1), which could have a deleterious effect

on decoding. It is hoped that future improvements to fMRI hardware and software will enable further improvement to acquisition sequences.

3. *Idiosyncratic hardware and software limitations of 7T scanners.* Our system did not allow me to take advantage of the higher spatial resolution offered by 7T-fMRI while retaining an adequately short first echo for our multi-echo sequences (cf. Miletić et al., 2020; Puckett et al., 2018). Additionally, parallel transmit can be combined with other methods – for example, multiband (e.g. Wu et al., 2016) – but this was not implemented on our scanner at the time of running this study and therefore needs to be empirically tested in the future.

4. *Computational limitations of novel decoding pipelines.* Both SOSLASSO and RSL are used for only the second time in this thesis (cf. C. R. Cox et al., 2024; C. R. Cox & Rogers, 2021). Both of these techniques are implemented using the WISC MVPA toolbox, which runs on the largest high-throughput computing cluster in North America at the University of Wisconsin-Madison. Interactions between parameters in this computationally demanding process are not fully understood – for example, my decision to use a hyperband budget of 25 (which may result in suboptimal models) was made because only then would the analysis run to completion. Ongoing research aims to further our understanding of the best ways to implement this pipeline.

5. *Nonuniform permutation distributions.* When constructing a permutation distribution for models based on regularised regression, voxels in the vATL are selected far more frequently than voxels elsewhere (Appendix F, Figures F.2 and F.3). The reason for this is unknown (and may be a consequence of reduced signal quality), but must be established before this issue can be addressed and unbiased conclusions about the dispersedness (see Chapter 2) of semantic representations can be drawn.

6.3.2 Adjudication between competing theories

In Chapter 2, I argued that researchers studying semantic representations should not merely choose an analysis method that can detect a neural code of the kind that they hypothesise to exist – rather, they should choose a method that is capable of adjudicating between hypotheses about the types of code that are possible.

My methods of choice in Chapters 3 and 5 were based on regularised regression. Regression (in and of itself) enables the detection of representations irrespective of the answers

to *any* of the six questions in Chapter 2. It must be used in conjunction with a regularisation penalty and the particular regularisation penalty used incorporates assumptions, but the penalties that I used did not blind me to code of any of the types listed in Table 2.1. This openness was advantageous to me since my aim was to establish whether there was evidence for semantic representation of *any* kind in the vATL or in the whole brain with 7T-fMRI or ECoG. However, the tentative conclusions discussed above – that the semantic code is either feature-based or vector-space, grounded, heterogeneous, inconsistent, and dispersed within individuals but contiguous across individuals – are tentative because this study was not designed to *falsify* any of those conclusions. Future work should seek to characterise the nature of semantic representations directly by designing studies that can be used to distinguish between representations of different types.

6.3.3 A taxonomy of time

Chapter 2 gives clear and exhaustive (if very broad) descriptions of how informative neural units might be situated relative to one another spatially – they may be contiguous or dispersed, within or across individuals. This framework may be sufficient for studies that lack temporal resolution, such as the study in Chapter 5. However, Chapter 3 describes representations that are dynamic and transfrequency. At present, "dynamic" is defined with reference to a neural network model, and "transfrequency" is defined only in the context of these results. Multiple questions therefore remain. First, it is unclear what the alternative possibilities are – a "static code" has not been defined. Second, the distinctions between codes with different temporal properties are poorly articulated – for example, the difference between a static and a dynamic code could be categorical or graded, and there could be subtypes of each. Third, it is unclear what other temporal dimensions may exist and may be useful for describing the way that representations vary over time. Fourth, it is unclear what benefit representing information via temporally-varying codes may convey to the brain.

Semantic theory still lags behind current practice and future theoretical work should seek to address this gap.

6.4 Conclusion

In this thesis I have developed a theoretical framework with which to describe the nature of neural representations, including semantic representations. I have shown that ECoG can be used to reveal a transfrequency code in the vATL; I have shown that 7T-fMRI can be used both to reveal the graded, multidimensional properties of semantic representations in the vATL and to identify the representation of semantic structure elsewhere. I have established that neurally-inspired regularisation penalties (SOSLASSO and grOWL), can be beneficial for decoding but that the best decoding methods for future studies are those that are carefully selected to complement the research question.

References

Abel, T. J., Rhone, A. E., Nourski, K. V., Kawasaki, H., Oya, H., Griffiths, T. D., Howard, M. A., & Tranel, D. (2015). Direct physiologic evidence of a heteromodal convergence region for proper naming in human left anterior temporal lobe. *Journal of Neuroscience*, *35*(4), 1513–1520. https://doi.org/10.1523/JNEUROSCI.3387-14.2015

Acosta-Cabronero, J., Patterson, K., Fryer, T. D., Hodges, J. R., Pengas, G., Williams, G. B., & Nestor, P. J. (2011). Atrophy, hypometabolism and white matter abnormalities in semantic dementia tell a coherent story. *Brain*, *134*(7), 2025–2035. https://doi.org/10.1093/brain/awr119

Adlam, A.-L. R., Patterson, K., Rogers, T. T., Nestor, P. J., Salmond, C. H., Acosta-Cabronero, J., & Hodges, J. R. (2006). Semantic dementia and fluent primary progressive aphasia: Two sides of the same coin? *Brain*, *129*(11), 3066–3080. https://doi.org/10.1093/brain/awl285

Afyouni, S., & Nichols, T. E. (2018). Insight and inference for DVARS. *NeuroImage*, *172*, 291–312. https://doi.org/10.1016/j.neuroimage.2017.12.098

Amemiya, S., Yamashita, H., Takao, H., & Abe, O. (2019). Integrated multi-echo denoising strategy improves identification of inherent language laterality. *Magnetic Resonance in Medicine*, *81*(5), 3262–3271. https://doi.org/10.1002/mrm.27620

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Raizada, R. D. S., Lin, F., & Lalor, E. C. (2019). An integrated neural decoder of linguistic and experiential meaning. *Journal of Neuroscience*, *39*(45), 8969–8987. https://doi.org/10.1523/JNEUROSCI.2575-18.2019

Anderson, A. J., Kiela, D., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Raizada, R. D. S., Grimm, S., & Lalor, E. C. (2021). Deep artificial neural networks reveal a distributed cortical network encoding propositional sentence-level meaning. *Journal of Neuroscience*, *41*(18), 4100–4119. https://doi.org/10.1523/JNEUROSCI.1152-20.2021

Anderson, A. J., Lalor, E. C., Lin, F., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Raizada, R. D. S., Grimm, S., & Wang, X. (2019). Multiple regions of a cortical network commonly encode the meaning of words in multiple grammatical positions of read sentences. *Cerebral Cortex*, *29*(6), 2396–2411. https://doi.org/10.1093/cercor/bhy110

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429. https://doi.org/10.1037/0033-295X.98.3.409

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., Singer, W., & Vicente, R. (2015). Untangling cross-frequency coupling in neuroscience. *Current Opinion in Neurobiology*, *31*, 51–61. https://doi.org/10.1016/j.conb.2014.08.002

Arya, R. (2019). Similarity of spatiotemporal dynamics of language-related ECoG high-gamma modulation in Japanese and English speakers. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *130*(8), 1403–1404. https://doi.org/10.1016/j.clinph.2019.05.006

Asyraff, A., Lemarchand, R., Tamm, A., & Hoffman, P. (2021). Stimulus-independent neural coding of event semantics: Evidence from cross-sentence fMRI decoding. *NeuroImage*, *236*, 118073. https://doi.org/10.1016/j.neuroimage.2021.118073

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Avants, B. B., Tustison, N., & Johnson, H. (2009). *Advanced Normalization Tools (ANTS)*.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 560–585. https://doi.org/10.1080/783663595

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*(5–6), 513–562. https://doi.org/10.1080/01690960344000026

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barth, M., Breuer, F., Koopmans, P. J., Norris, D. G., & Poser, B. A. (2016). Simultaneous multislice (SMS) imaging techniques. *Magnetic Resonance in Medicine*, *75*(1), 63–81. https://doi.org/10.1002/mrm.25897

Başar-Eroglu, C., Strüber, D., Schürmann, M., Stadler, M., & Başar, E. (1996). Gamma-band responses in the brain: A short review of psychophysiological correlates and functional significance. *International Journal of Psychophysiology*, *24*(1–2), 101–112. https://doi.org/10.1016/S0167-8760(96)00051-7

Behrmann, M., & Plaut, D. C. (2014). Bilateral hemispheric processing of words and faces: Evidence from word impairments in prosopagnosia and face impairments in pure alexia. *Cerebral Cortex*, *24*(4), 1102–1118. https://doi.org/10.1093/cercor/bhs390

Behrmann, M., Scherf, K. S., & Avidan, G. (2016). Neural mechanisms of face perception, their emergence over development, and their breakdown. *WIREs Cognitive Science*, *7*(4), 247–263. https://doi.org/10.1002/wcs.1388

Benítez-Burraco, A., & Murphy, E. (2019). Why brain oscillations are improving our understanding of language. *Frontiers in Behavioral Neuroscience*, *13*. https://doi.org/10.3389/fnbeh.2019.00190

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.

Bertrand, O., Bohorquez, J., & Pernier, J. (1994). Time-frequency digital filtering based on an invertible wavelet transform: An application to evoked potentials. *IEEE Transactions on Biomedical Engineering*, *41*(1), 77–88. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/10.277274

Bhavsar, S., Zvyagintsev, M., & Mathiak, K. (2014). BOLD sensitivity and SNR characteristics of parallel imaging-accelerated single-shot multi-echo EPI for fMRI. *NeuroImage*, *84*, 65–75. https://doi.org/10.1016/j.neuroimage.2013.08.007

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

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience*, *11*(1), 80–93. https://doi.org/10.1162/089892999563265

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M., & Lambon Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: Evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex*, *20*(11), 2728–2738. https://doi.org/10.1093/cercor/bhq019

Booth, A. E., & Waxman, S. R. (2008). Taking stock as theories of word learning take shape. *Developmental Science*, *11*(2), 185–194. https://doi.org/10.1111/j.1467-7687.2007.00664.x

Boyacioğlu, R., Schulz, J., Koopmans, P. J., Barth, M., & Norris, D. G. (2015). Improved sensitivity and specificity for resting state and task fMRI with multiband multi-echo EPI compared to multi-echo EPI at 7 T. *NeuroImage*, *119*, 352–361. https://doi.org/10.1016/j.neuroimage.2015.06.089

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. https://doi.org/10.1016/S0028-3932(00)00034-8

Bressler, S. L. (1990). The gamma wave: A cortical information carrier? *Trends in Neurosciences*, *13*(5), 161–162. https://doi.org/10.1016/0166-2236(90)90039-D

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, *12*(2–3), 177–210. https://doi.org/10.1080/016909697386844

Caballero-Gaudes, C., Moia, S., Panwar, P., Bandettini, P. A., & Gonzalez-Castillo, J. (2019). A deconvolution algorithm for multi-echo functional MRI: Multi-echo Sparse Paradigm Free Mapping. *NeuroImage*, *202*, 116081. https://doi.org/10.1016/j.neuroimage.2019.116081

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Computational Biology*, *10*(12), e1003963. https://doi.org/10.1371/journal.pcbi.1003963

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, *313*(5793), 1626–1628. https://doi.org/10.1126/science.1128115

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, *14*(11), 506–515. https://doi.org/10.1016/j.tics.2010.09.001

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. https://doi.org/10.1016/S1364-6613(03)00159-1

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. https://doi.org/10.1162/089892998563752

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–1. https://doi.org/10.1167/13.10.1

Carota, F., Kriegeskorte, N., Nili, H., & Pulvermüller, F. (2017). Representational similarity mapping of distributional semantics in left inferior frontal, middle temporal, and motor cortex. *Cerebral Cortex*, *27*(1), 294–309. https://doi.org/10.1093/cercor/bhw379

Cervenka, M. C., Corines, J., Boatman-Reich, D. F., Eloyan, A., Sheng, X., Franaszczuk, P. J., & Crone, N. E. (2013). Electrocorticographic functional mapping identifies human cortex critical for auditory and visual naming. *NeuroImage*, *69*, 267–276. https://doi.org/10.1016/j.neuroimage.2012.12.037

Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., Cash, S. S., & Halgren, E. (2011). First-pass selectivity for semantic categories in human anteroventral temporal lobe. *Journal of Neuroscience*, *31*(49), 18119–18129. https://doi.org/10.1523/JNEUROSCI.3122-11.2011

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. https://doi.org/10.1038/13217
Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025
Chen, L., Lambon Ralph, M. A., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, *1*(3), 0039. https://doi.org/10.1038/s41562-016-0039
Chen, L., & Rogers, T. T. (2014). Revisiting domain-general accounts of category specificity in mind and brain. *WIREs Cognitive Science*, *5*(3), 327–344. https://doi.org/10.1002/wcs.1283
Chen, L., & Rogers, T. T. (2015). A model of emergent category-specific activation in the posterior fusiform gyrus of sighted and congenitally blind populations. *Journal of Cognitive Neuroscience*, *27*(10), 1981–1999. https://doi.org/10.1162/jocn_a_00834
Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., Fukuyama, H., Takahashi, R., Ikeda, A., & Lambon Ralph, M. A. (2016). The 'when' and 'where' of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex*, *79*, 1–13. https://doi.org/10.1016/j.cortex.2016.02.015
Chiou, R., Humphreys, G. F., Jung, J., & Lambon Ralph, M. A. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation' systems. *Cortex*, *103*, 100–116. https://doi.org/10.1016/j.cortex.2018.02.018
Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*(1), 27755. https://doi.org/10.1038/srep27755
Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635
Clarke, A. (2020). Dynamic activity patterns in the anterior temporal lobe represents object semantics. *Cognitive Neuroscience*, *11*(3), 111–121. https://doi.org/10.1080/17588928.2020.1742678
Clarke, A., Devereux, B. J., & Tyler, L. K. (2018). Oscillatory dynamics of perceptual to conceptual transformations in the ventral visual pathway. *Journal of Cognitive Neuroscience*, *30*(11), 1590–1605. https://doi.org/10.1162/jocn_a_01325
Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *Journal of Neuroscience*, *34*(14), 4766–4775. https://doi.org/10.1523/JNEUROSCI.2828-13.2014
Cohen, A. D., Nencka, A. S., Lebel, R. M., & Wang, Y. (2017). Multiband multi-echo imaging of simultaneous oxygenation and flow timeseries for resting state connectivity. *PLOS ONE*, *12*(3), e0169253. https://doi.org/10.1371/journal.pone.0169253
Cohen, A. D., Nencka, A. S., & Wang, Y. (2018). Multiband multi-echo simultaneous ASL/BOLD for task-induced functional MRI. *PLOS ONE*, *13*(2), e0190427. https://doi.org/10.1371/journal.pone.0190427
Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.
Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.

https://doi.org/10.1016/S0022-5371(69)80069-1
Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Abdi, H., & Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012
Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013
Cope, T. E., Sohoglu, E., Peterson, K. A., Jones, P. S., Rua, C., Passamonti, L., Sedley, W., Post, B., Coebergh, J., Butler, C. R., Garrard, P., Abdel-Aziz, K., Husain, M., Griffiths, T. D., Patterson, K., Davis, M. H., & Rowe, J. B. (2023). Temporal lobe perceptual predictions for speech are instantiated in motor cortex and reconciled by inferior frontal cortex. *Cell Reports*, *42*(5). https://doi.org/10.1016/j.celrep.2023.112422
Coutanche, M. N. (2013). Distinguishing multi-voxel patterns and mean activation: Why, how, and what does it tell us? *Cognitive, Affective, & Behavioral Neuroscience*, *13*(3), 667–673. https://doi.org/10.3758/s13415-013-0186-2
Cox, C. R. (2016). *Testing neurocognitive predictions of the hub-and-spoke model of semantic memory with network representational similarity analysis*. University of Wisconsin-Madison.
Cox, C. R., & Rogers, T. T. (2021). Finding distributed needles in neural haystacks. *Journal of Neuroscience*, *41*(5), 1019–1032. https://doi.org/10.1523/JNEUROSCI.0904-20.2020
Cox, C. R., Rogers, T. T., Shimotake, A., Kikuchi, T., Kunieda, T., Miyamoto, S., Takahashi, R., Matsumoto, R., Ikeda, A., & Lambon Ralph, M. A. (2024). Representational similarity learning reveals a graded multidimensional semantic space in the human anterior temporal cortex. *Imaging Neuroscience*, *2*, 1–22. https://doi.org/10.1162/imag_a_00093
Cox, C. R., Seidenberg, M. S., & Rogers, T. T. (2015). Connecting functional brain imaging and parallel distributed processing. *Language, Cognition and Neuroscience*, *30*(4), 380–394. https://doi.org/10.1080/23273798.2014.994010
Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270. https://doi.org/10.1016/S1053-8119(03)00049-1
Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014
Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L
Craddock, M., Martinovic, J., & Müller, M. M. (2016). Accounting for microsaccadic artifacts in the EEG using independent component analysis and beamforming. *Psychophysiology*, *53*(4), 553–565. https://doi.org/10.1111/psyp.12593
Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*(3), 371–414. https://doi.org/10.1207/s15516709cog2303_4
Crone, N. E., Boatman, D., Gordon, B., & Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. *Clinical Neurophysiology*, *112*(4), 565–582. https://doi.org/10.1016/S1388-2457(00)00545-9
Crone, N. E., & Hao, L. (2002). Functional dynamics of spoken and signed word production: A case study using electrocorticographic spectral analysis. *Aphasiology*, *16*(9), 903–927. https://doi.org/10.1080/02687030244000383
Crone, N. E., Korzeniewska, A., & Franaszczuk, P. J. (2011). Cortical gamma responses: Searching

high and low. *International Journal of Psychophysiology*, *79*(1), 9–15. https://doi.org/10.1016/j.ijpsycho.2010.10.013
Crone, N. E., Miglioretti, D. L., Gordon, B., Sieracki, J. M., Wilson, M. T., Uematsu, S., & Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. *Brain*, *121*(12), 2271–2299. https://doi.org/10.1093/brain/121.12.2271
Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395
Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, *1*(1), 123–132. https://doi.org/10.1162/neco.1989.1.1.123
Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1), 179–229. https://doi.org/10.1016/j.cognition.2002.07.001
Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., & Poldrack, R. A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage*, *97*, 271–283. https://doi.org/10.1016/j.neuroimage.2014.04.037
Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: Practices and pitfalls. *Annals of the New York Academy of Sciences*, *1296*(1), 108–134. https://doi.org/10.1111/nyas.12156
de Cheveigné, A. (2023). *Is EEG is better left alone?* (p. 2023.06.19.545602). bioRxiv. https://doi.org/10.1101/2023.06.19.545602
Dehghani, M., Boghrati, R., Man, K., Hoover, J., Gimbel, S. I., Vaswani, A., Zevin, J. D., Immordino-Yang, M. H., Gordon, A. S., Damasio, A., & Kaplan, J. T. (2017). Decoding the neural representation of story meanings across languages. *Human Brain Mapping*, *38*(12), 6096–6106. https://doi.org/10.1002/hbm.23814
Delorme, A. (2023). EEG is better left alone. *Scientific Reports*, *13*(1), 2372. https://doi.org/10.1038/s41598-023-27528-0
Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009
Demetriou, L., Kowalczyk, O. S., Tyson, G., Bello, T., Newbould, R. D., & Wall, M. B. (2018). A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI. *NeuroImage*, *176*, 404–416. https://doi.org/10.1016/j.neuroimage.2018.05.011
Derby, S., Miller, P., Murphy, B., & Devereux, B. (2018). *Using sparse semantic embeddings learned from multimodal text and image data to model human conceptual knowledge* (arXiv:1809.02534). arXiv. http://arxiv.org/abs/1809.02534
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. https://doi.org/10.1523/JNEUROSCI.3809-13.2013
Devereux, B. J., Clarke, A., & Tyler, L. K. (2018). Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Scientific Reports*, *8*(1), 10636. https://doi.org/10.1038/s41598-018-28865-1
Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional*

transformers for language understanding (arXiv:1810.04805). arXiv. http://arxiv.org/abs/1810.04805

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), 589–600. https://doi.org/10.1006/nimg.2000.0595

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Computational Biology*, *13*(4), e1005508. https://doi.org/10.1371/journal.pcbi.1005508

Dijkstra, N., van Gaal, S., Geerligs, L., Bosch, S. E., & van Gerven, M. A. J. (2021). No evidence for neural overlap between unconsciously processed and imagined stimuli. *eNeuro*, *8*(5), ENEURO.0228-21.2021. https://doi.org/10.1523/ENEURO.0228-21.2021

Dilkina, K., & Lambon Ralph, M. A. (2012). Conceptual structure within and between modalities. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00333

Ding, B., Dragonu, I., Rua, C., Carlin, J. D., Halai, A. D., Liebig, P., Heidemann, R., Correia, M. M., & Rodgers, C. T. (2022). Parallel transmit (pTx) with online pulse design for task-based fMRI at 7T. *Magnetic Resonance Imaging*, *93*, 163–174. https://doi.org/10.1016/j.mri.2022.07.003

Dipasquale, O., Sethi, A., Laganà, M. M., Baglio, F., Baselli, G., Kundu, P., Harrison, N. A., & Cercignani, M. (2017). Comparing resting state fMRI de-noising approaches using multi- and single-echo acquisitions. *PLOS ONE*, *12*(3), e0173289. https://doi.org/10.1371/journal.pone.0173289

Drane, D. L., Ojemann, G. A., Aylward, E., Ojemann, J. G., Johnson, L. C., Silbergeld, D. L., Miller, J. W., & Tranel, D. (2008). Category-specific naming and recognition deficits in temporal lobe epilepsy surgical patients. *Neuropsychologia*, *46*(5), 1242–1255. https://doi.org/10.1016/j.neuropsychologia.2007.11.034

Duffau, H. (2014). The huge plastic potential of adult brain and the role of connectomics: New insights provided by serial mappings in glioma surgery. *Cortex*, *58*, 325–337. https://doi.org/10.1016/j.cortex.2013.08.005

Dundas, E. M., Plaut, D. C., & Behrmann, M. (2013). The joint development of hemispheric lateralization for words and faces. *Journal of Experimental Psychology: General*, *142*(2), 348–358. https://doi.org/10.1037/a0029503

DuPre, E., Salo, T., Ahmed, Z., Bandettini, P. A., Bottenhorn, K. L., Caballero-Gaudes, C., Dowdle, L. T., Gonzalez-Castillo, J., Heunis, S., Kundu, P., Laird, A. R., Markello, R., Markiewicz, C. J., Moia, S., Staden, I., Teves, J. B., Uruñuela, E., Vaziri-Pashkam, M., Whitaker, K., & Handwerker, D. A. (2021). TE-dependent analysis of multi-echo fMRI with tedana. *Journal of Open Source Software*, *6*(66), 3669. https://doi.org/10.21105/joss.03669

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, *21*(4), 449–467. https://doi.org/10.1017/S0140525X98001253

Edwards, E., Nagarajan, S. S., Dalal, S. S., Canolty, R. T., Kirsch, H. E., Barbaro, N. M., & Knight, R. T. (2010). Spatiotemporal imaging of cortical activation during verb generation and picture naming. *NeuroImage*, *50*(1), 291–301. https://doi.org/10.1016/j.neuroimage.2009.12.035

Eggert, G. H. (1977). *Wernicke's works on aphasia: A sourcebook and review*. Mouton.

Embleton, K. V., Haroon, H. A., Morris, D. M., Lambon Ralph, M. A. L., & Parker, G. J. M. (2010). Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Human Brain Mapping*, *31*(10), 1570–1587. https://doi.org/10.1002/hbm.20959

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Evans, J. W., Kundu, P., Horovitz, S. G., & Bandettini, P. A. (2015). Separating slow BOLD from non-BOLD baseline drifts using multi-echo fMRI. *NeuroImage*, *105*, 189–197. https://doi.org/10.1016/j.neuroimage.2014.10.051

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339–357. https://doi.org/10.1037/0096-3445.120.4.339

Fernandez, B., Leuchs, L., Sämann, P. G., Czisch, M., & Spoormaker, V. I. (2017). Multi-echo EPI of human fear conditioning reveals improved BOLD detection in ventromedial prefrontal cortex. *NeuroImage*, *156*, 65–77. https://doi.org/10.1016/j.neuroimage.2017.05.005

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*, *26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., & Binder, J. R. (2016). Heteromodal cortical areas encode sensory-motor features of word meaning. *Journal of Neuroscience*, *36*(38), 9763–9769. https://doi.org/10.1523/JNEUROSCI.4095-15.2016

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

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207. https://doi.org/10.1006/nimg.1998.0396

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Forseth, K. J., Kadipasaoglu, C. M., Conner, C. R., Hickok, G., Knight, R. T., & Tandon, N. (2018). A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain*, *141*(7), 2112–2126. https://doi.org/10.1093/brain/awy120

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1). https://doi.org/10.18637/jss.v033.i01

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*(3), 218–229. https://doi.org/10.1006/nimg.1997.0291

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7

Gaser, C., Dahnke, R., Thompson, P. M., Kurth, F., Luders, E., & Alzheimer's Disease Neuroimaging Initiative. (2023). *CAT – A computational anatomy toolbox for the analysis of structural MRI data* (p. 2022.06.11.495736). bioRxiv. https://doi.org/10.1101/2022.06.11.495736

Geschwind, N. (1972). Language and the brain. *Scientific American*, *226*(4), 76–83.

Gilmore, A. W., Agron, A. M., González-Araya, E. I., Gotts, S. J., & Martin, A. (2022). A comparison of single- and multi-echo processing of functional MRI data during overt autobiographical recall. *Frontiers in Neuroscience*, *16*, 854387. https://doi.org/10.3389/fnins.2022.854387

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *WIREs Cognitive Science*, *1*(4), 586–596. https://doi.org/10.1002/wcs.55

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*(3), 379–401. https://doi.org/10.1006/jmla.2000.2714

Gonzalez-Castillo, J., Panwar, P., Buchanan, L. C., Caballero-Gaudes, C., Handwerker, D. A., Jangraw, D. C., Zachariou, V., Inati, S., Roopchansingh, V., Derbyshire, J. A., & Bandettini, P. A. (2016). Evaluation of multi-echo ICA denoising for task based fMRI studies: Block designs, rapid event-related designs, and cardiac-gated fMRI. *NeuroImage*, *141*, 452–468. https://doi.org/10.1016/j.neuroimage.2016.07.049

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Gras, V., Poser, B. A., Wu, X., Tomi-Tricot, R., & Boulant, N. (2019). Optimizing BOLD sensitivity in the 7T Human Connectome Project resting-state fMRI protocol using plug-and-play parallel transmission. *NeuroImage*, *195*, 1–10. https://doi.org/10.1016/j.neuroimage.2019.03.040

Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., & Binder, J. R. (2010). Neural systems for reading aloud: A multiparametric approach. *Cerebral Cortex*, *20*(8), 1799–1815. https://doi.org/10.1093/cercor/bhp245

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. https://doi.org/10.1037/0033-295X.114.2.211

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A model of representational spaces in human cortex. *Cerebral Cortex*, *26*(6), 2919–2934. https://doi.org/10.1093/cercor/bhw068

Hagberg, G. E., Indovina, I., Sanes, J. N., & Posse, S. (2002). Real-time quantification of T2* changes using multiecho planar imaging and numerical methods. *Magnetic Resonance in Medicine*, *48*(5), 877–882. https://doi.org/10.1002/mrm.10283

Halai, A. D., Henson, R. N., Finoia, P., & Correia, M. M. (2024). *Comparing the effect of multi gradient echo and multi band fMRI during a semantic task* (p. 2024.03.20.585909). bioRxiv. https://doi.org/10.1101/2024.03.20.585909

Halai, A. D., Parkes, L. M., & Welbourne, S. R. (2015). Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension. *NeuroImage*, *122*, 214–221. https://doi.org/10.1016/j.neuroimage.2015.05.067

Halai, A. D., Welbourne, S. R., Embleton, K., & Parkes, L. M. (2014). A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Human Brain Mapping*, *35*(8), 4118–4128. https://doi.org/10.1002/hbm.22463

Halchenko, Y. O., Goncalves, M., Ghosh, S., Velasco, P., Castello, M. V. di O., Salo, T., Wodder, J. T., Hanke, M., Sadil, P., Gorgolewski, K. J., Ioanas, H.-I., Rorden, C., Hendrickson, T. J., Dayan, M., Houlihan, S. D., Kent, J., Strauss, T., Lee, J., To, I., … Kennedy, D. N. (2024). HeuDiConv—Flexible DICOM conversion into structured directory layouts. *Journal of Open Source Software*, *9*(99), 5839. https://doi.org/10.21105/joss.05839

Halgren, E., Kaestner, E., Marinkovic, K., Cash, S. S., Wang, C., Schomer, D. L., Madsen, J. R., & Ulbert, I. (2015). Laminar profile of spontaneous and evoked theta: Rhythmic modulation of cortical processing during word integration. *Neuropsychologia*, *76*, 108–124. https://doi.org/10.1016/j.neuropsychologia.2015.03.021

Halgren, E., Wang, C., Schomer, D. L., Knake, S., Marinkovic, K., Wu, J., & Ulbert, I. (2006). Processing stages underlying word recognition in the anteroventral temporal lobe. *NeuroImage*, *30*(4), 1401–1413. https://doi.org/10.1016/j.neuroimage.2005.10.053

Hamberger, M. J. (2007). Cortical language mapping in epilepsy: A critical review. *Neuropsychology Review*, *17*(4), 477–489. https://doi.org/10.1007/s11065-007-9046-6

Hampton, J. A. (2015). Categories, prototypes and exemplars. In N. Riemer (Ed.), *The Routledge Handbook of Semantics*. Routledge.

Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G., & Pietrini, P. (2016). How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage*, *135*, 232–242. https://doi.org/10.1016/j.neuroimage.2016.04.063

Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *NeuroImage*, *23*(1), 156–166. https://doi.org/10.1016/j.neuroimage.2004.05.020

Hargreaves, B. A., Cunningham, C. H., Nishimura, D. G., & Conolly, S. M. (2004). Variable-rate selective excitation for rapid MRI sequences. *Magnetic Resonance in Medicine*, *52*(3), 590–597. https://doi.org/10.1002/mrm.20168

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62*(2), 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*(1), 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Henry, T. R., Buchtel, H. A., Koeppe, R. A., Pennell, P. B., Kluin, K. J., & Minoshima, S. (1998). Absence of normal activation of the left anterior fusiform gyrus during naming in left temporal lobe epilepsy. *Neurology*, *50*(3), 787–790. https://doi.org/10.1212/WNL.50.3.787

Hermes, D., Miller, K. J., Vansteensel, M. J., Edwards, E., Ferrier, C. H., Bleichner, M. G., van Rijen, P. C., Aarnoutse, E. J., & Ramsey, N. F. (2014). Cortical theta wanes for language. *NeuroImage*, *85*, 738–748. https://doi.org/10.1016/j.neuroimage.2013.07.029

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory*, *3*(3–4), 463–495. https://doi.org/10.1080/09658219508253161

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: A unique clinicopathological syndrome. *The Lancet Neurology*, *6*(11), 1004–1014. https://doi.org/10.1016/S1474-4422(07)70266-1

Hodges, J. R., Patterson, K. E., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, *115*(6), 1783–1806. https://doi.org/10.1093/brain/115.6.1783

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Howard, D., & Patterson, K. E. (1992). *The pyramids and palm trees test.*. Thames Valley Test Company.

Humphreys, G. F., Hoffman, P., Visser, M., Binney, R. J., & Lambon Ralph, M. A. (2015). Establishing task- and modality-dependent dissociations between the semantic and default mode networks. *Proceedings of the National Academy of Sciences*, *112*(25), 7857–7862. https://doi.org/10.1073/pnas.1422760112

Humphreys, G. F., Lambon Ralph, M. A., & Simons, J. S. (2021). A unifying account of angular gyrus contributions to episodic and semantic cognition. *Trends in Neurosciences*, *44*(6), 452–463. https://doi.org/10.1016/j.tins.2021.01.006

Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behavioral and Brain Sciences*, *24*(3), 453–476. https://doi.org/10.1017/S0140525X01004150

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. https://doi.org/10.1038/nature17637

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. https://doi.org/10.1016/j.neuron.2012.10.014

Jackson, R. L. (2021). The neural correlates of semantic control revisited. *NeuroImage*, *224*, 117444. https://doi.org/10.1016/j.neuroimage.2020.117444

Jackson, R. L., Rogers, T. T., & Lambon Ralph, M. A. (2021). Reverse-engineering the cortical architecture for controlled semantic cognition. *Nature Human Behaviour*, *5*(6), 774–786. https://doi.org/10.1038/s41562-020-01034-z

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*(8), 2132–2147. https://doi.org/10.1093/brain/awl153

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

Jerbi, K., Freyermuth, S., Dalal, S., Kahane, P., Bertrand, O., Berthoz, A., & Lachaux, J.-P. (2009). Saccade related gamma-band activity in intracerebral EEG: Dissociating neural from ocular muscle activity. *Brain Topography*, *22*(1), 18–23. https://doi.org/10.1007/s10548-009-0078-5

Jia, J., & Yu, B. (2008). On model selection consistency of the elastic net when $p>>n$. *Statistica Sinica*, *20*, 595–611. https://doi.org/10.21236/ada485557

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

Jones, M. N., Willits, J., & Dennis, S. (2015). In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Models of Semantic Memory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199957996.013.11

Jung, J., Williams, S. R., Sanaei Nezhad, F., & Lambon Ralph, M. A. (2017). GABA concentrations in the anterior temporal lobe predict human semantic processing. *Scientific Reports*, *7*(1), 15748. https://doi.org/10.1038/s41598-017-15981-7

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of

concrete noun representation based on the underlying brain codes. *PLoS ONE*, *5*(1), e8622. https://doi.org/10.1371/journal.pone.0008622
Kanemoto, K., Takeuchi, J., Kawasaki, J., & Kawai, I. (1996). Characteristics of temporal lobe epilepsy with mesial temporal sclerosis, with special reference to psychotic episodes. *Neurology*, *47*(5), 1199–1203. https://doi.org/10.1212/WNL.47.5.1199
Kang, J. Y., & Krauss, G. L. (2019). Normal variants are commonly overread as interictal epileptiform abnormalities. *Journal of Clinical Neurophysiology*, *36*(4), 257. https://doi.org/10.1097/WNP.0000000000000613
Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. https://doi.org/10.1038/77664
Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, *107*(25), 11163–11170. https://doi.org/10.1073/pnas.1005062107
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. https://doi.org/10.1523/jneurosci.17-11-04302.1997
Katz, J. J. (1972). *Semantic theory*. Harper & Row.
Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. https://doi.org/10.1038/nature06713
King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002
Kirilina, E., Lutti, A., Poser, B. A., Blankenburg, F., & Weiskopf, N. (2016). The quest for the best: The impact of different EPI sequences on the sensitivity of random effect fMRI group analyses. *NeuroImage*, *126*, 49–59. https://doi.org/10.1016/j.neuroimage.2015.10.071
Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007
Kojima, K., Brown, E. C., Matsuzaki, N., Rothermel, R., Fuerst, D., Shah, A., Mittal, S., Sood, S., & Asano, E. (2013). Gamma activity modulated by picture and auditory naming tasks: Intracranial recording in patients with focal epilepsy. *Clinical Neurophysiology*, *124*(9), 1737–1744. https://doi.org/10.1016/j.clinph.2013.01.030
Koopmans, P. J., Barth, M., Orzada, S., & Norris, D. G. (2011). Multi-echo fMRI of the cortical laminae in humans at 7T. *NeuroImage*, *56*(3), 1276–1285. https://doi.org/10.1016/j.neuroimage.2011.02.042
Kovach, C. K., Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M. A., & Adolphs, R. (2011). Manifestation of ocular-muscle EMG contamination in human intracranial recordings. *NeuroImage*, *54*(1), 213–233. https://doi.org/10.1016/j.neuroimage.2010.08.002
Kovářová, A., Gajdoš, M., Rektor, I., & Mikl, M. (2022). Contribution of the multi-echo approach in accelerated functional magnetic resonance imaging multiband acquisition. *Human Brain Mapping*, *43*(3), 955–973. https://doi.org/10.1002/hbm.25698
Kraskov, A., Quiroga, R. Q., Reddy, L., Fried, I., & Koch, C. (2007). Local field potentials and spikes in the human medial temporal lobe are selective to image category. *Journal of Cognitive Neuroscience*, *19*(3), 479–492. https://doi.org/10.1162/jocn.2007.19.3.479
Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*(Volume 1, 2015), 417–446. https://doi.org/10.1146/annurev-vision-082114-035447
Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain

mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103
Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(4). https://doi.org/10.3389/neuro.06.004.2008
Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386
Kumar, A. A., Steyvers, M., & Balota, D. A. (2022). A critical review of network-based and distributional approaches to semantic memory structure and processes. *Cognitive Science*, *14*(1), 54–77. https://doi.org/10.1111/tops.12548
Kundu, P., Benson, B. E., Baldwin, K. L., Rosen, D., Luh, W.-M., Bandettini, P. A., Pine, D. S., & Ernst, M. (2015). Robust resting state fMRI processing for studies on typical brain development based on multi-echo EPI acquisition. *Brain Imaging and Behavior*, *9*(1), 56–73. https://doi.org/10.1007/s11682-014-9346-4
Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vértes, P. E., Inati, S. J., Saad, Z. S., Bandettini, P. A., & Bullmore, E. T. (2013). Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences*, *110*(40), 16187–16192. https://doi.org/10.1073/pnas.1301725110
Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M., & Bandettini, P. A. (2011). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, *60*(3), 1759–1770. https://doi.org/10.1016/j.neuroimage.2011.12.028
Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., & Bandettini, P. A. (2017). Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *NeuroImage*, *154*, 59–80. https://doi.org/10.1016/j.neuroimage.2017.03.033
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657
Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E., & Crone, N. E. (2012). High-frequency neural activity and human cognition: Past, present and possible future of intracranial EEG research. *Progress in Neurobiology*, *98*(3), 279–301. https://doi.org/10.1016/j.pneurobio.2012.06.008
Lam, N. H. L., Schoffelen, J.-M., Uddén, J., Hultén, A., & Hagoort, P. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *NeuroImage*, *142*, 43–54. https://doi.org/10.1016/j.neuroimage.2016.03.007
Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150
Lambon Ralph, M. A., & Patterson, K. (2008). Generalization and differentiation in semantic memory. *Annals of the New York Academy of Sciences*, *1124*(1), 61–76. https://doi.org/10.1196/annals.1440.006
Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned

by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex*, *19*(4), 832–838. https://doi.org/10.1093/cercor/bhn131

Lambon Ralph, M. A., 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. https://doi.org/10.1073/pnas.0907307107

Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science*, *7*(5), 161–164. https://doi.org/10.1111/1467-8721.ep10836862

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Le Ster, C., Moreno, A., Mauconduit, F., Gras, V., Stirnberg, R., Poser, B. A., Vignaud, A., Eger, E., Dehaene, S., Meyniel, F., & Boulant, N. (2019). Comparison of SMS-EPI and 3D-EPI at 7T in an fMRI localizer study with matched spatiotemporal resolution and homogenized excitation profiles. *PLOS ONE*, *14*(11), e0225286. https://doi.org/10.1371/journal.pone.0225286

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *Journal of Neuroscience*, *28*(35), 8765–8771. https://doi.org/10.1523/JNEUROSCI.1953-08.2008

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, *130*(1), 3–28. https://doi.org/10.1037/0096-3445.130.1.3

Lisman, J., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, *77*(6), 1002–1016. https://doi.org/10.1016/j.neuron.2013.03.007

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*(2), 281–290. https://doi.org/10.1016/j.neuron.2009.02.025

Liu, T. T. (2016). Noise contributions to the fMRI signal: An overview. *NeuroImage*, *143*, 141–151. https://doi.org/10.1016/j.neuroimage.2016.09.008

Liuzzi, A. G., Bruffaerts, R., Dupont, P., Adamczuk, K., Peeters, R., De Deyne, S., Storms, G., & Vandenberghe, R. (2015). Left perirhinal cortex codes for similarity in meaning between written words: Comparison with auditory word input. *Neuropsychologia*, *76*, 4–16. https://doi.org/10.1016/j.neuropsychologia.2015.03.016

Liuzzi, A. G., Ubaldi, S., & Fairhall, S. L. (2021). Representations of conceptual information during automatic and active semantic access. *Neuropsychologia*, *160*, 107953. https://doi.org/10.1016/j.neuropsychologia.2021.107953

Lombardo, M. V., Auyeung, B., Holt, R. J., Waldman, J., Ruigrok, A. N. V., Mooney, N., Bullmore, E. T., Baron-Cohen, S., & Kundu, P. (2016). Improving effect size estimation and statistical power with multi-echo fMRI and its impact on understanding the neural systems supporting mentalizing. *NeuroImage*, *142*, 55–66. https://doi.org/10.1016/j.neuroimage.2016.07.022

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, *32*(3), 251–295. https://doi.org/10.1006/cogp.1997.0651

Lüders, H., Lesser, R. P., Hahn, J., Dinner, D. S., Morris, H. H., Wylie, E., & Godoy, J. (1991). Basal temporal language area. *Brain*, *114*(2), 743–754. https://doi.org/10.1093/brain/114.2.743

Lynch, C. J., Power, J. D., Scult, M. A., Dubin, M., Gunning, F. M., & Liston, C. (2020). Rapid precision functional mapping of individuals using multi-echo fMRI. *Cell Reports*, *33*(12), 108540. https://doi.org/10.1016/j.celrep.2020.108540

Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00165

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*(3), 397–405. https://doi.org/10.1016/j.neuron.2009.07.012

Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, *55*(3), 507–520. https://doi.org/10.1016/j.neuron.2007.07.011

Mahon, B. Z., Schwarzbach, J., & Caramazza, A. (2010). The representation of tools in left parietal cortex is independent of visual experience. *Psychological Science*, *21*(6), 764–771. https://doi.org/10.1177/0956797610370754

Mandler, J. M. (2006). *The foundations of mind: Origins of conceptual thought*. Oxford University Press.

Marko, M., Cimrová, B., & Riečanský, I. (2019). Neural theta oscillations support semantic memory retrieval. *Scientific Reports*, *9*(1), 17667. https://doi.org/10.1038/s41598-019-53813-y

Marques, J. P., & Norris, D. G. (2018). How to choose the right MR sequence for your research question at 7 T and above? *NeuroImage*, *168*, 119–140. https://doi.org/10.1016/j.neuroimage.2017.04.044

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*(1), 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review*, *23*(4), 979–990. https://doi.org/10.3758/s13423-015-0842-3

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

Matoba, K., Matsumoto, R., Shimotake, A., Nakae, T., Imamura, H., Togo, M., Yamao, Y., Usami, K., Kikuchi, T., Yoshida, K., Matsuhashi, M., Kunieda, T., Miyamoto, S., Takahashi, R., & Ikeda, A. (2024). Basal temporal language area revisited in Japanese language with a language function density map. *Cerebral Cortex*, *34*(6), bhae218. https://doi.org/10.1093/cercor/bhae218

McNabb, C. B., Lindner, M., Shen, S., Burgess, L. G., Murayama, K., & Johnstone, T. (2020). Inter-slice leakage and intra-slice aliasing in simultaneous multi-slice echo-planar images. *Brain Structure and Function*, *225*(3), 1153–1158. https://doi.org/10.1007/s00429-020-02053-2

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

McRae, K., Seidenberg, M. S., & de Sa, V. R. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99–130.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Mehrer, J., Spoerer, C. J., Kriegeskorte, N., & Kietzmann, T. C. (2020). Individual differences

among deep neural network models. *Nature Communications*, *11*(1), 5725. https://doi.org/10.1038/s41467-020-19632-w

Merker, B. (2013). Cortical gamma oscillations: The functional key is activation, not cognition. *Neuroscience & Biobehavioral Reviews*, *37*(3), 401–417. https://doi.org/10.1016/j.neubiorev.2013.01.013

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*(1), 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Mesulam, M.-M., Wieneke, C., Hurley, R., Rademaker, A., Thompson, C. K., Weintraub, S., & Rogalski, E. J. (2013). Words and objects at the tip of the left temporal lobe in primary progressive aphasia. *Brain*, *136*(2), 601–618. https://doi.org/10.1093/brain/aws336

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*. https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

Miletić, S., Bazin, P.-L., Weiskopf, N., van der Zwaag, W., Forstmann, B. U., & Trampel, R. (2020). fMRI protocol optimization for simultaneously studying small subcortical and cortical areas at 7 T. *NeuroImage*, *219*, 116992. https://doi.org/10.1016/j.neuroimage.2020.116992

Miller, K. J. (2010). Broadband spectral change: Evidence for a macroscale correlate of population firing rate? *Journal of Neuroscience*, *30*(19), 6477–6479. https://doi.org/10.1523/JNEUROSCI.6401-09.2010

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195. https://doi.org/10.1126/science.1152876

Mitra, P., & Bokil, H. (2007). *Observed Brain Dynamics*. Oxford University Press.

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine*, *63*(5), 1144–1153. https://doi.org/10.1002/mrm.22361

Mollo, G., Cornelissen, P. L., Millman, R. E., Ellis, A. W., & Jefferies, E. (2017). Oscillatory dynamics supporting semantic cognition: MEG evidence for the contribution of the anterior temporal lobe hub and modality-specific spokes. *PLOS ONE*, *12*(1), e0169269. https://doi.org/10.1371/journal.pone.0169269

Morris, L. S., Kundu, P., Costi, S., Collins, A., Schneider, M., Verma, G., Balchandani, P., & Murrough, J. W. (2019). Ultra-high field MRI reveals mood-related circuit disturbances in depression: A comparison between 3-tesla and 7-tesla. *Translational Psychiatry*, *9*(1), 94. https://doi.org/10.1038/s41398-019-0425-6

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 528–559. https://doi.org/10.1080/027249897392017

Murphy, G. L. (2004). *The big book of concepts*. MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037//0033-295x.92.3.289

Nagata, K., Kunii, N., Shimada, S., Fujitani, S., Takasago, M., & Saito, N. (2022). Spatiotemporal target selection for intracranial neural decoding of abstract and concrete semantics. *Cerebral Cortex*, *32*(24), 5544–5554. https://doi.org/10.1093/cercor/bhac034

Nakai, Y., Jeong, J., Brown, E. C., Rothermel, R., Kojima, K., Kambara, T., Shah, A., Mittal, S., Sood, S., & Asano, E. (2017). Three- and four-dimensional mapping of speech and language in patients with epilepsy. *Brain*, *140*(5), 1351–1370. https://doi.org/10.1093/brain/awx051

Nakai, Y., Sugiura, A., Brown, E. C., Sonoda, M., Jeong, J., Rothermel, R., Luat, A. F., Sood, S., & Asano, E. (2019). Four-dimensional functional cortical maps of visual and auditory language: Intracranial recording. *Epilepsia*, *60*(2), 255–267. https://doi.org/10.1111/epi.14648

Nobre, A., & McCarthy, G. (1995). Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *Journal of Neuroscience*, *15*(2), 1090–1098. https://doi.org/10.1523/JNEUROSCI.15-02-01090.1995

Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: Evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of Cognitive Neuroscience*, *25*(11), 1824–1850. https://doi.org/10.1162/jocn_a_00442

Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P., Mummery, C., & Price, C. J. (2006). Temporal lobe lesions and semantic impairment: A comparison of herpes simplex virus encephalitis and semantic dementia. *Brain*, *130*(4), 1138–1147. https://doi.org/10.1093/brain/awl344

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

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54–65. https://doi.org/10.1037/0278-7393.14.1.54

Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). Voxelwise encoding models with non-spherical multivariate normal priors. *NeuroImage*, *197*, 482–492. https://doi.org/10.1016/j.neuroimage.2019.04.012

O'Brien, K. R., Kober, T., Hagmann, P., Maeder, P., Marques, J., Lazeyras, F., Krueger, G., & Roche, A. (2014). Robust T1-weighted structural brain imaging and morphometry at 7T using MP2RAGE. *PLOS ONE*, *9*(6), e99676. https://doi.org/10.1371/journal.pone.0099676

Oswal, U., Cox, C. R., Lambon Ralph, M. A., Rogers, T. T., & Nowak, R. D. (2016). Representational similarity learning with application to brain networks. *Proceedings of The 33rd International Conference on Machine Learning*, 1041–1049. https://proceedings.mlr.press/v48/oswal16.html

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*(4), 580–590.

Panigrahi, A., Simhadri, H. V., & Bhattacharyya, C. (2019). Word2sense: Sparse interpretable word embeddings. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5692–5705). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1570

Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nature Neuroscience*, *21*(4), 474–483. https://doi.org/10.1038/s41593-018-0108-2

Patterson, K., & Hodges, J. R. (2000). Semantic dementia: One window on the structure and organisation of semantic memory. In L. S. Cermak (Ed.), *Handbook of neuropsychology: Memory and its disorders* (2nd ed., Vol. 2, pp. 313–333). Elsevier.

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. https://doi.org/10.1038/nrn2277

Pauen, S. (2002a). Evidence for knowledge–based category discrimination in infancy. *Child Development*, *73*(4), 1016–1033. https://doi.org/10.1111/1467-8624.00454

Pauen, S. (2002b). The global-to-basic level shift in infants' categorical thinking: First evidence from a longitudinal study. *International Journal of Behavioral Development*, *26*(6), 492–499. https://doi.org/10.1080/01650250143000445

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., & Fazio1, F. (1999). The neural correlates of verb and noun processing: A PET study. *Brain*, *122*(12), 2337–2344. https://doi.org/10.1093/brain/122.12.2337

Pereira, F., & Botvinick, M. (2011). Information mapping with pattern classifiers: A comparative study. *NeuroImage*, *56*(2), 476–496. https://doi.org/10.1016/j.neuroimage.2010.05.026

Pereira, F., Detre, G., & Botvinick, M. (2011). Generating text from functional brain images. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00072

Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology*, *33*(3–4), 175–190. https://doi.org/10.1080/02643294.2016.1176907

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., Botvinick, M., & Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-03068-4

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*, *45*(1), S199–S209. https://doi.org/10.1016/j.neuroimage.2008.11.007

Phipson, B., & Smyth, G. K. (2010). Permutation *p*-values should never be zero: Calculating exact *p*-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, *9*(1). https://doi.org/10.2202/1544-6115.1585

Pick, A. (1898). *Beiträge zur Pathologie und pathologischen Anatomie des Centralnervensystems*. Karger. http://archive.org/details/b21294367

Plaut, D. C., & Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. *Cognitive Neuropsychology*, *28*(3–4), 251–275. https://doi.org/10.1080/02643294.2011.609812

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2007). Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences*, *104*(50), 20137–20141. https://doi.org/10.1073/pnas.0707383104

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010a). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia*, *48*(5), 1336–1342. https://doi.org/10.1016/j.neuropsychologia.2009.12.036

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010b). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current Biology*, *20*(10), 964–968. https://doi.org/10.1016/j.cub.2010.03.070

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 border of human visual cortex. *Nature Neuroscience*, *24*(11), 1628–1636. https://doi.org/10.1038/s41593-021-00921-6

Poser, B. A., & Norris, D. G. (2009). Investigating the benefits of multi-echo EPI for fMRI at 7 T. *NeuroImage*, *45*(4), 1162–1172. https://doi.org/10.1016/j.neuroimage.2009.01.007

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227–1235. https://doi.org/10.1002/mrm.20900

Posse, S. (2012). Multi-echo acquisition. *NeuroImage*, *62*(2), 665–671. https://doi.org/10.1016/j.neuroimage.2011.10.057

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *Journal of Neuroscience*, *35*(7), 3276–3284. https://doi.org/10.1523/JNEUROSCI.3446-14.2015

Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage*, *5*(4), 261–270. https://doi.org/10.1006/nimg.1997.0269

Puckett, A. M., Bollmann, S., Poser, B. A., Palmer, J., Barth, M., & Cunnington, R. (2018). Using multi-echo simultaneous multi-slice (SMS) EPI to improve functional MRI of the subcortical nuclei of the basal ganglia at ultra-high field (7T). *NeuroImage*, *172*, 886–895. https://doi.org/10.1016/j.neuroimage.2017.12.005

Rao, N., Cox, C., Nowak, R. D., & Rogers, T. T. (2013). Sparse overlapping sets lasso for multitask learning and its application to fMRI analysis. *Advances in Neural Information Processing Systems*, *26*. https://proceedings.neurips.cc/paper/2013/file/a1519de5b5d44b31a01de013b9b51a80-Paper.pdf

Rao, N., Nowak, R. D., Cox, C., & Rogers, T. (2016). Classification with the sparse group lasso. *IEEE Transactions on Signal Processing*, *64*(2), 448–463. https://doi.org/10.1109/TSP.2015.2488586

Reber, T. P., Bausch, M., Mackay, S., Boström, J., Elger, C. E., & Mormann, F. (2019). Representation of abstract semantic knowledge in populations of human single neurons in the medial temporal lobe. *PLOS Biology*, *17*(6), e3000290. https://doi.org/10.1371/journal.pbio.3000290

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., … & Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770. https://doi.org/10.1038/s41593-019-0520-2

Riddoch, M. J., & Humphreys, G. W. (2003). Visual agnosia. *Neurologic Clinics*, *21*(2), 501–520. https://doi.org/10.1016/S0733-8619(02)00095-6

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. https://doi.org/10.1038/14819

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. https://doi.org/10.1016/S0022-5371(73)80056-8

Risk, B. B., Murden, R. J., Wu, J., Nebel, M. B., Venkataraman, A., Zhang, Z., & Qiu, D. (2021).

Which multiband factor should you choose for your resting-state fMRI study? *NeuroImage*, *234*, 117965. https://doi.org/10.1016/j.neuroimage.2021.117965

Rogers, T. T. (2020). Neural networks as a critical level of description for cognitive neuroscience. *Current Opinion in Behavioral Sciences*, *32*, 167–173. https://doi.org/10.1016/j.cobeha.2020.02.009

Rogers, T. T. (2024). Generalization and Abstraction: Human Memory as a Magic Library. In M. J. Kahana & A. D. Wagner (Eds.), *The oxford handbook of human memory, Two volume pack: Foundations and applications* (pp. 172–214). Oxford University Press.

Rogers, T. T., Cox, C. R., Lu, Q., Shimotake, A., Kikuchi, T., Kunieda, T., Miyamoto, S., Takahashi, R., Ikeda, A., Matsumoto, R., & Lambon Ralph, M. A. (2021). Evidence for a deep, distributed and dynamic code for animacy in human ventral anterior temporal cortex. *eLife*, *10*, e66276. https://doi.org/10.7554/eLife.66276

Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K., & Price, C. J. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(3), 201–213. https://doi.org/10.3758/CABN.6.3.201

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*(1), 205–235. https://doi.org/10.1037/0033-295X.111.1.205

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press.

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rogers, T. T., Patterson, K., Jefferies, E., & Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: Effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220–239. https://doi.org/10.1016/j.neuropsychologia.2015.04.015

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Lawrence Erlbaum Associates.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rotaru, A. S., Vigliocco, G., & Frank, S. L. (2018). Modeling the structure and dynamics of semantic processing. *Cognitive Science*, *42*(8), 2890–2917. https://doi.org/10.1111/cogs.12690

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing, Volume 1: Explorations in the microstructure of cognition: Foundations*. The MIT Press. https://doi.org/10.7551/mitpress/5236.001.0001

Rupp, K., Roos, M., Milsap, G., Caceres, C., Ratto, C., Chevillet, M., Crone, N. E., & Wolmetz,

M. (2017). Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. *NeuroImage*, *148*, 318–329. https://doi.org/10.1016/j.neuroimage.2016.12.074

Ruts, W., De Deyne, S., Ameel, E., Vanpaemel, W., Verbeemen, T., & Storms, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 506–515. https://doi.org/10.3758/BF03195597

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage*, *27*(1), 188–200. https://doi.org/10.1016/j.neuroimage.2005.04.012

Sato, N., Matsumoto, R., Shimotake, A., Matsuhashi, M., Otani, M., Kikuchi, T., Kunieda, T., Mizuhara, H., Miyamoto, S., Takahashi, R., & Ikeda, A. (2021). Frequency-dependent cortical interactions during semantic processing: An electrocorticogram cross-spectrum analysis using a semantic space model. *Cerebral Cortex*, *31*(9), 4329–4339. https://doi.org/10.1093/cercor/bhab089

Sauseng, P., & Klimesch, W. (2008). What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience & Biobehavioral Reviews*, *32*(5), 1001–1013. https://doi.org/10.1016/j.neubiorev.2008.03.014

Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience*, *23*(34), 10809–10814. https://doi.org/10.1523/JNEUROSCI.23-34-10809.2003

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., Tully, M. S., & Kahana, M. J. (2006). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex*, *17*(5), 1190–1196. https://doi.org/10.1093/cercor/bhl030

Seghier, M. L., Fagan, E., & Price, C. J. (2010). Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *Journal of Neuroscience*, *30*(50), 16809–16817. https://doi.org/10.1523/JNEUROSCI.3377-10.2010

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429. https://doi.org/10.1073/pnas.0700622104

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., & Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced *g*-factor penalty. *Magnetic Resonance in Medicine*, *67*(5), 1210–1224. https://doi.org/10.1002/mrm.23097

Sexton, N. J., & Love, B. C. (2022). Reassessing hierarchical correspondences between brain and deep networks through direct interface. *Science Advances*, *8*(28), eabm2219. https://doi.org/10.1126/sciadv.abm2219

Shimotake, A., Matsumoto, R., Ueno, T., Kunieda, T., Saito, S., Hoffman, P., Kikuchi, T., Fukuyama, H., Miyamoto, S., Takahashi, R., Ikeda, A., & Lambon Ralph, M. A. (2015). Direct exploration of the role of the ventral anterior temporal lobe in semantic memory: Cortical stimulation and local field potential evidence from subdural grid electrodes. *Cerebral Cortex*, *25*(10), 3802–3817. https://doi.org/10.1093/cercor/bhu262

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

Simanova, I., Hagoort, P., Oostenveld, R., & van Gerven, M. A. J. (2014). Modality-independent decoding of semantic information from the human brain. *Cerebral Cortex*, *24*(2), 426–434. https://doi.org/10.1093/cercor/bhs324

Simons, J. S., & Lambon Ralph, M. A. (1999). The auditory agnosias. *Neurocase*, *5*(5), 379–406. https://doi.org/10.1080/13554799908402734

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition* (arXiv:1409.1556). arXiv. http://arxiv.org/abs/1409.1556

Smith, E. E., & Medin, D. L. (2013). *Categories and Concepts*. Harvard University Press. https://doi.org/10.4159/harvard.9780674866270

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214–241. https://doi.org/10.1037/h0036351

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

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

Snowden, J., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology*, *2*(3), 124043. https://doi.org/10.1155/1989/124043

Snyder, K. M., Forseth, K. J., Donos, C., Rollo, P. S., Fischer-Baum, S., Breier, J., & Tandon, N. (2023). Critical role of the ventral temporal lobe in naming. *Epilepsia*, *64*(5), 1200–1213. https://doi.org/10.1111/epi.17555

Solomon, E. A., Lega, B. C., Sperling, M. R., & Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the National Academy of Sciences*, *116*(48), 24343–24352. https://doi.org/10.1073/pnas.1906729116

Spiers, H. J. (2020). The hippocampal cognitive map: One space or many? *Trends in Cognitive Sciences*, *24*(3), 168–170. https://doi.org/10.1016/j.tics.2019.12.013

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063

Stryker, M. P. (1989). Is grandmother an oscillation? *Nature*, *338*(6213), 297–298. https://doi.org/10.1038/338297a0

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162. https://doi.org/10.1016/S1364-6613(99)01299-1

Tang, J., LeBel, A., & Huth, A. G. (2021). *Cortical representations of concrete and abstract concepts in language combine visual and linguistic representations* (p. 2021.05.19.444701). bioRxiv. https://doi.org/10.1101/2021.05.19.444701

Tanji, K. (2005). High-frequency-band activity in the basal temporal cortex during picture-naming and lexical-decision tasks. *Journal of Neuroscience*, *25*(13), 3287–3293. https://doi.org/10.1523/JNEUROSCI.4948-04.2005

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Todd, N., Moeller, S., Auerbach, E. J., Yacoub, E., Flandin, G., & Weiskopf, N. (2016). Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: Sensitivity and slice leakage artifacts. *NeuroImage*, *124*, 32–42. https://doi.org/10.1016/j.neuroimage.2015.08.056

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language*, *75*(2), 195–231. https://doi.org/10.1006/brln.2000.2353

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Van Rullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*(6), 655–668. https://doi.org/10.1068/p3029

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, *383*(6597), 254–256. https://doi.org/10.1038/383254a0

Vargas, R., & Just, M. A. (2019). Neural representations of abstract concepts: Identifying underlying neurosemantic dimensions. *Cerebral Cortex*, *30*(4), 2157–2166. https://doi.org/10.1093/cercor/bhz229

Visconti di Oleggio Castello, M., Haxby, J. V., & Gobbini, M. I. (2021). Shared neural codes for visual and semantic information about familiar faces in a common representational space. *Proceedings of the National Academy of Sciences*, *118*(45), e2110474118. https://doi.org/10.1073/pnas.2110474118

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. https://doi.org/10.1162/jocn.2009.21309

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, *5*(4), 520–526. https://doi.org/10.1016/0959-4388(95)80014-X

Wang, W., Degenhart, A. D., Sudre, G. P., Pomerleau, D. A., & Tyler-Kabara, E. C. (2011). Decoding semantic information from human electrocorticographic (ECoG) signals. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 6294-6298). IEEE.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*(4), 635–657. https://doi.org/10.1080/14640747508400525

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–854.

Watrous, A. J., Deuker, L., Fell, J., & Axmacher, N. (2015). Phase-amplitude coupling supports phase coding in human ECoG. *eLife*, *4*, e07886. https://doi.org/10.7554/eLife.07886

Watrous, A. J., Fell, J., Ekstrom, A. D., & Axmacher, N. (2015). More than spikes: Common oscillatory mechanisms for content specific neural representations during perception and memory. *Current Opinion in Neurobiology*, *31*, 33–39. https://doi.org/10.1016/j.conb.2014.07.024

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Wu, X., Schmitter, S., Auerbach, E. J., Uğurbil, K., & Van de Moortele, P.-F. (2016). A generalized slab-wise framework for parallel transmit multiband RF pulse design. *Magnetic Resonance in Medicine*, *75*(4), 1444–1456. https://doi.org/10.1002/mrm.25689

Wurm, M. F., & Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature Communications*, *10*(1), 289. https://doi.org/10.1038/s41467-018-08084-y

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yamadori, A. (2019). Gogi (word weaning) aphasia and its relation with semantic dementia. In J. Bougousslavsky, F. Boller, & M. Iwata (Eds.), *A History of Neuropsychology* (Vol. 44, pp. 30–38). Karger.

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, *22*(2), 297–306. https://doi.org/10.1038/s41593-018-0310-2

Yun, S. D., & Shah, N. J. (2017). Whole-brain high in-plane resolution fMRI using accelerated EPIK for enhanced characterisation of functional areas at 3T. *PLOS ONE*, *12*(9), e0184759. https://doi.org/10.1371/journal.pone.0184759

Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature Reviews Neuroscience*, *16*(8), 487–497. https://doi.org/10.1038/nrn3962

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, *58*(3), 429–441. https://doi.org/10.1016/j.neuron.2008.03.027

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zhou, Y., Vales, M. I., Wang, A., & Zhang, Z. (2017). Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction. *Briefings in Bioinformatics*, *18*(5), 744–753. https://doi.org/10.1093/bib/bbw064

Appendix A

Supplementary information for Chapter 2

Details of the literature review summarised in Table 2.1 can be downloaded at https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(22)00323-0#f0010/.

Appendix B

Supplementary results for Chapter 3

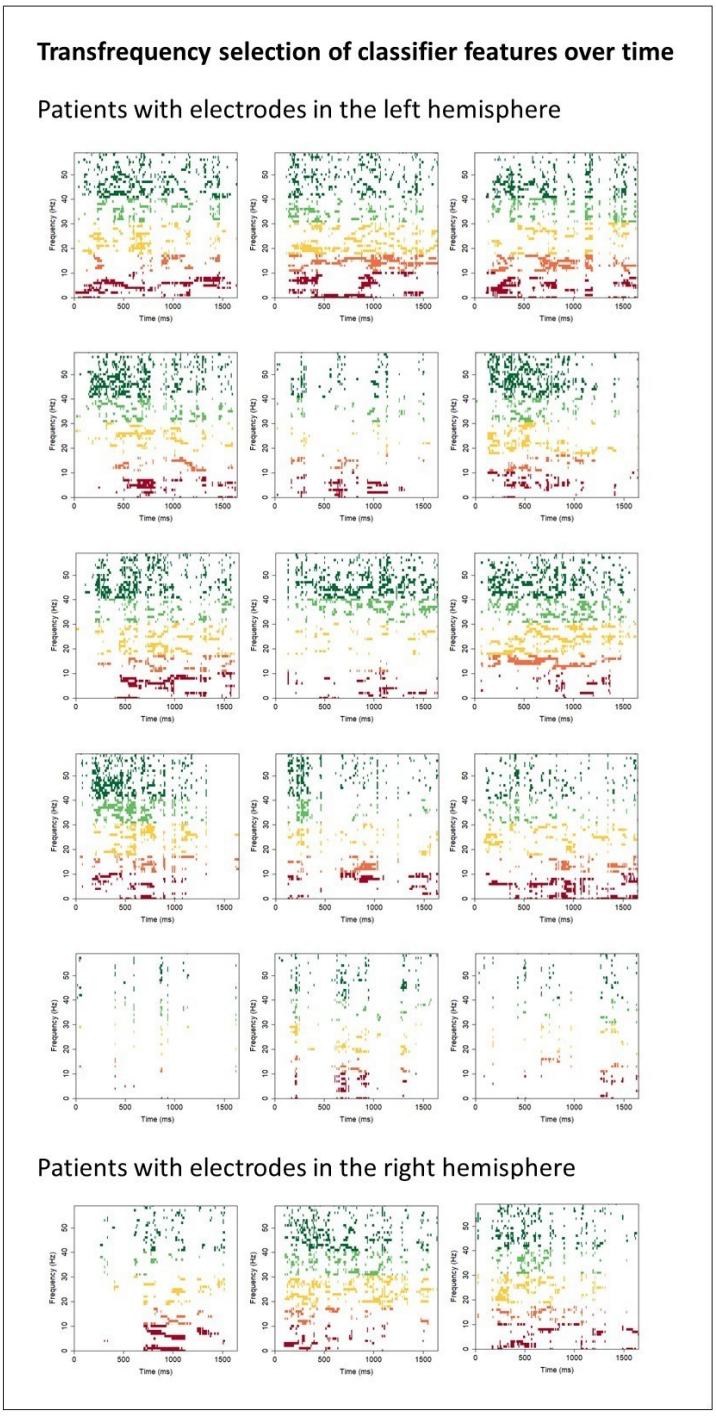

Figure B.1

Figure B.1: Feature selection of each frequency at each timepoint (averaged across electrodes). Classifiers were trained on power frequency features for all frequencies between 4 and 200 Hz. Features with a nonzero coefficient are shown in colour - theta (4 – 7 Hz, red), alpha (12 – 18 Hz, orange), beta (13 – 30 Hz, yellow), gamma (30 – 60 Hz, light green) and high gamma (42 – 200 Hz, dark green). Each plot shows selection for a single patient.

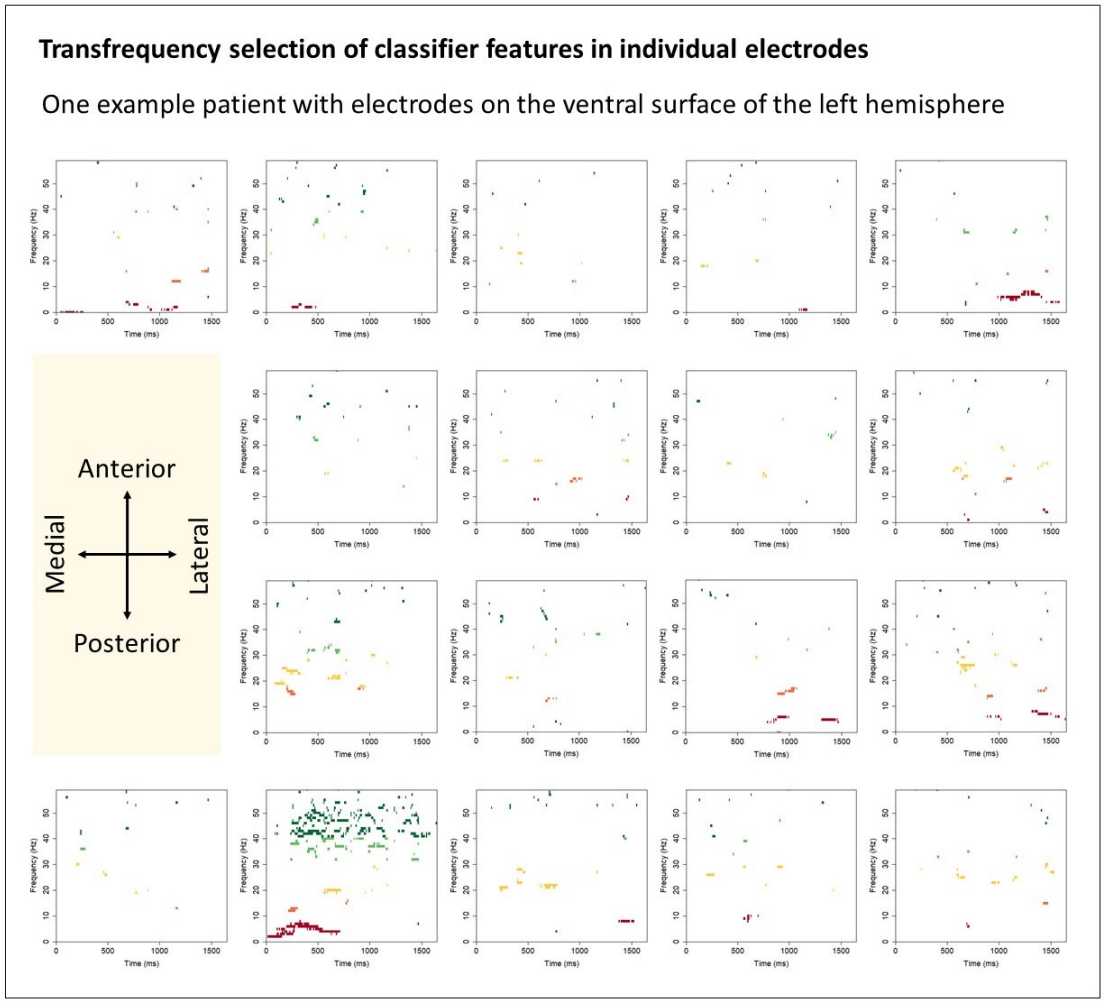

Figure B.2: Feature selection in each electrode for a single patient. Classifiers were trained on power frequency features for all frequencies between 4 and 200 Hz. Features with a nonzero coefficient are shown in colour - theta (4 – 7 Hz, red), alpha (12 – 18 Hz, orange), beta (13 – 30 Hz, yellow), gamma (30 – 60 Hz, light green) and high gamma (42 – 200 Hz, dark green). Each plot shows coefficients for a single electrode. The arrangement of the plots illustrates the layout of the electrodes on the ventral surface of the left hemisphere of this patient's brain.

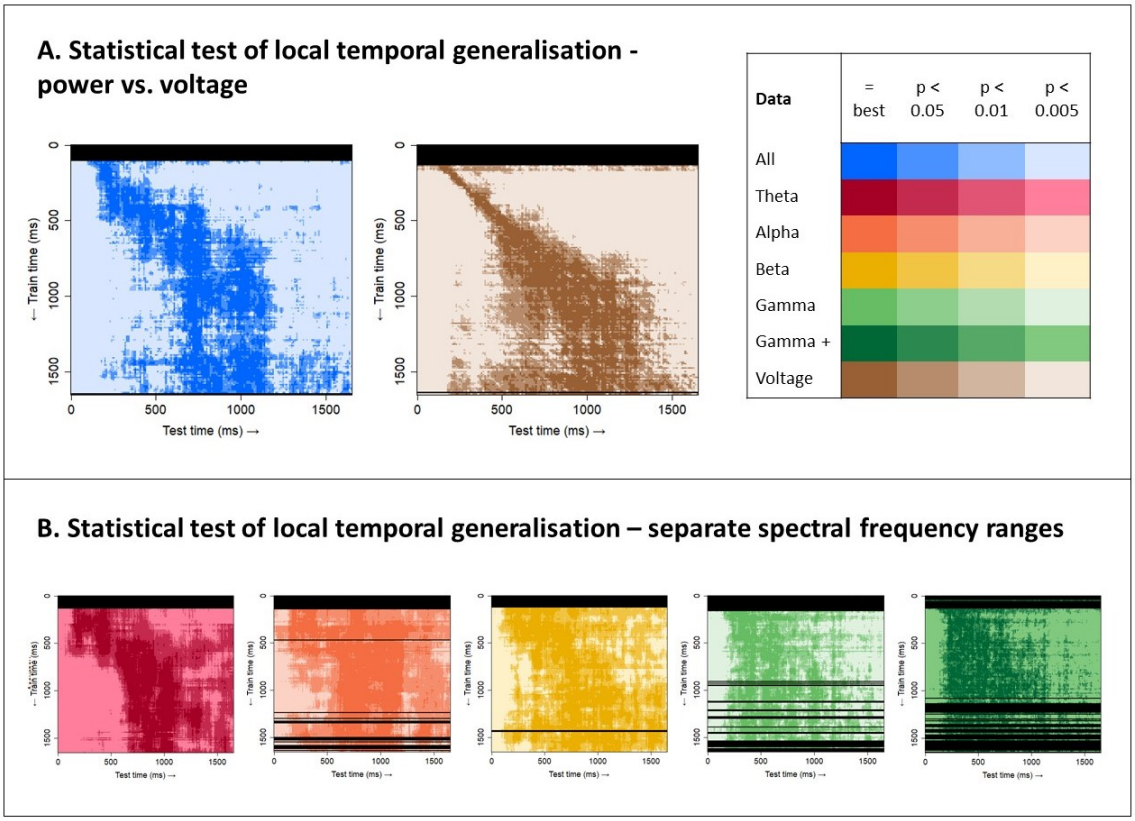

Figure B.3: Statistical test of local temporal generalisation. For each test time, the training time of the best classifier, plus the training times of other classifiers that perform statistically indistinguishably from the best ($p>0.05$, paired t-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$), are shown in the darkest colour. Progressively lighter colours show the training time of classifiers exhibiting a significant difference from the best classifier ($p<0.05$, $p<0.01$, $p<0.005$). Timepoints where classifiers trained at that timepoint do not perform significantly better than chance (0.5, one-sample t-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$) are shown in black. (A) Results for classifiers trained on power frequency features for all frequencies between 4 and 200 Hz (blue) or on voltage features (brown). (B) Results for classifiers trained on power frequency features from a single range – theta (4 – 7 Hz, red), alpha (12 – 18 Hz, orange), beta (13 – 30 Hz, yellow), gamma (30 – 60 Hz, light green) and high gamma (42 – 200 Hz, dark green).

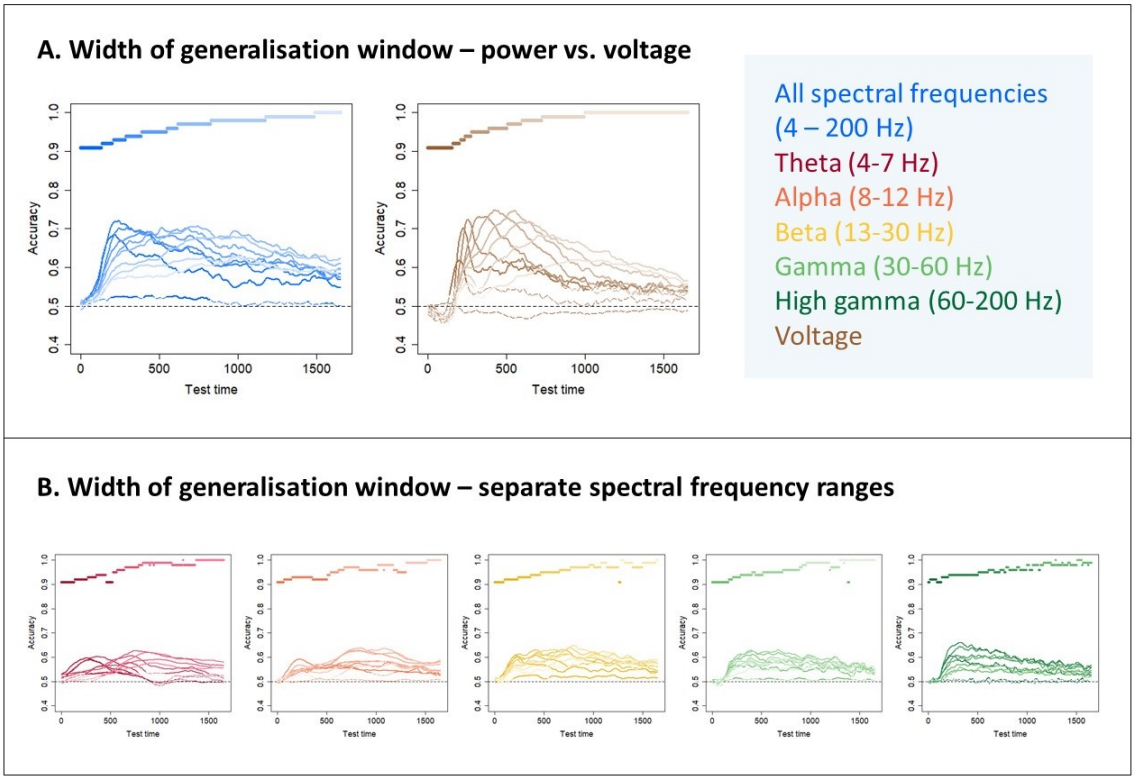

Figure B.4: Width of generalisation window for all frequency ranges. Classifiers are grouped into 10 clusters via agglomerative hierarchical clustering (see Methods). The "timecourses" show the mean hold-out accuracy for classifiers within each cluster at each timepoint. Lines are solid where there is a significant difference between classifier accuracy and chance (0.5, one-sample *t*-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$) and dashed where there is no significant difference. Coloured bars show the grouped timepoints in each cluster. (A) Results for classifiers trained on power frequency features for all frequencies between 4 and 200 Hz (shades of blue) or on voltage features (shades of brown). (B) Results for classifiers trained on power frequency features from a single range – theta (4 – 7 Hz, shades of red), alpha (12 – 18 Hz, shades of orange), beta (13 – 30 Hz, shades of yellow), gamma (30 – 60 Hz, shades of light green) and high gamma (42 – 200 Hz, shades of dark green).

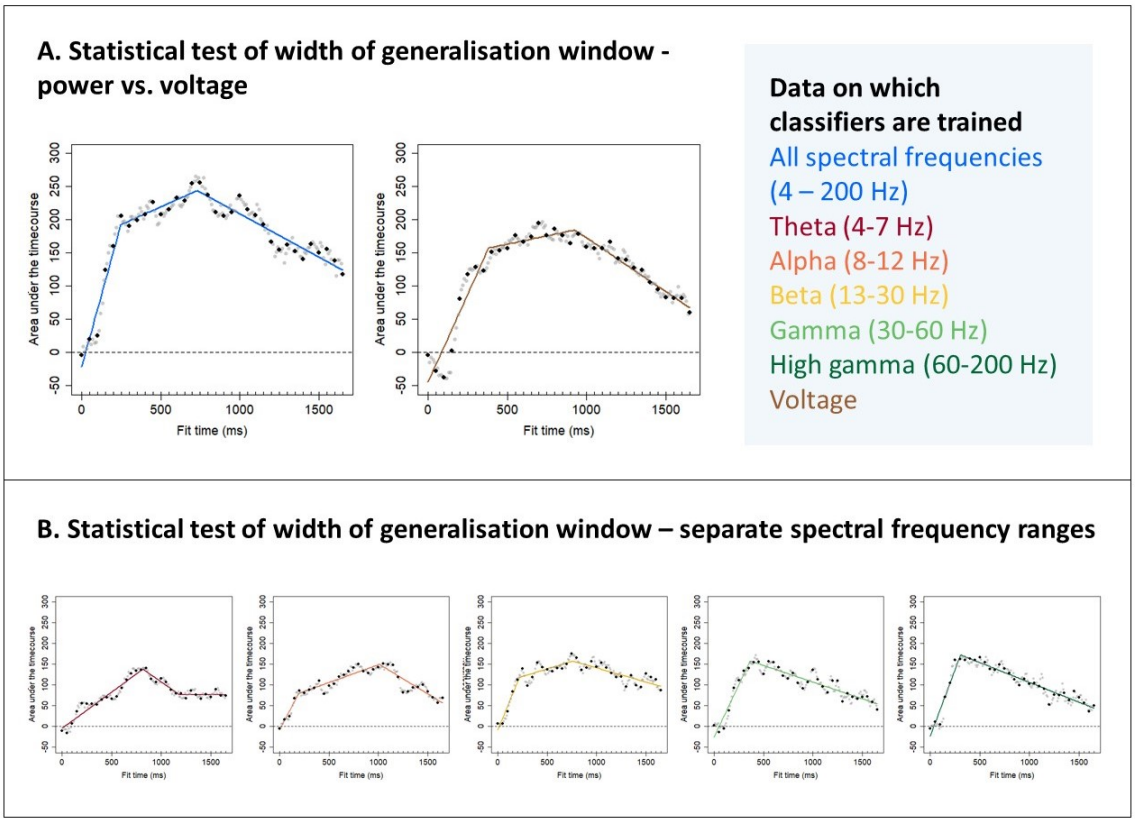

Figure B.5: Area under the curve between the timecourse of hold-out accuracy for each classifier and a horizontal line at chance (0.5). Lines show a piecewise linear model fit to timepoints 50 ms apart (0 ms, 50 ms, 100 ms, …; black dots). This ensures that voltage feature vectors from neighbouring timepoints do not contain overlapping data points and that power results are comparable with voltage results. Timepoints which were not used to fit the model are shown in grey. (A) Results for classifiers trained on power frequency features for all frequencies between 4 and 200 Hz (blue) or on voltage features (brown). (B) Results for classifiers trained on power frequency features from a single range – theta (4 – 7 Hz, red), alpha (12 – 18 Hz, orange), beta (13 – 30 Hz, yellow), gamma (30 – 60 Hz, light green) and high gamma (42 – 200 Hz, dark green).

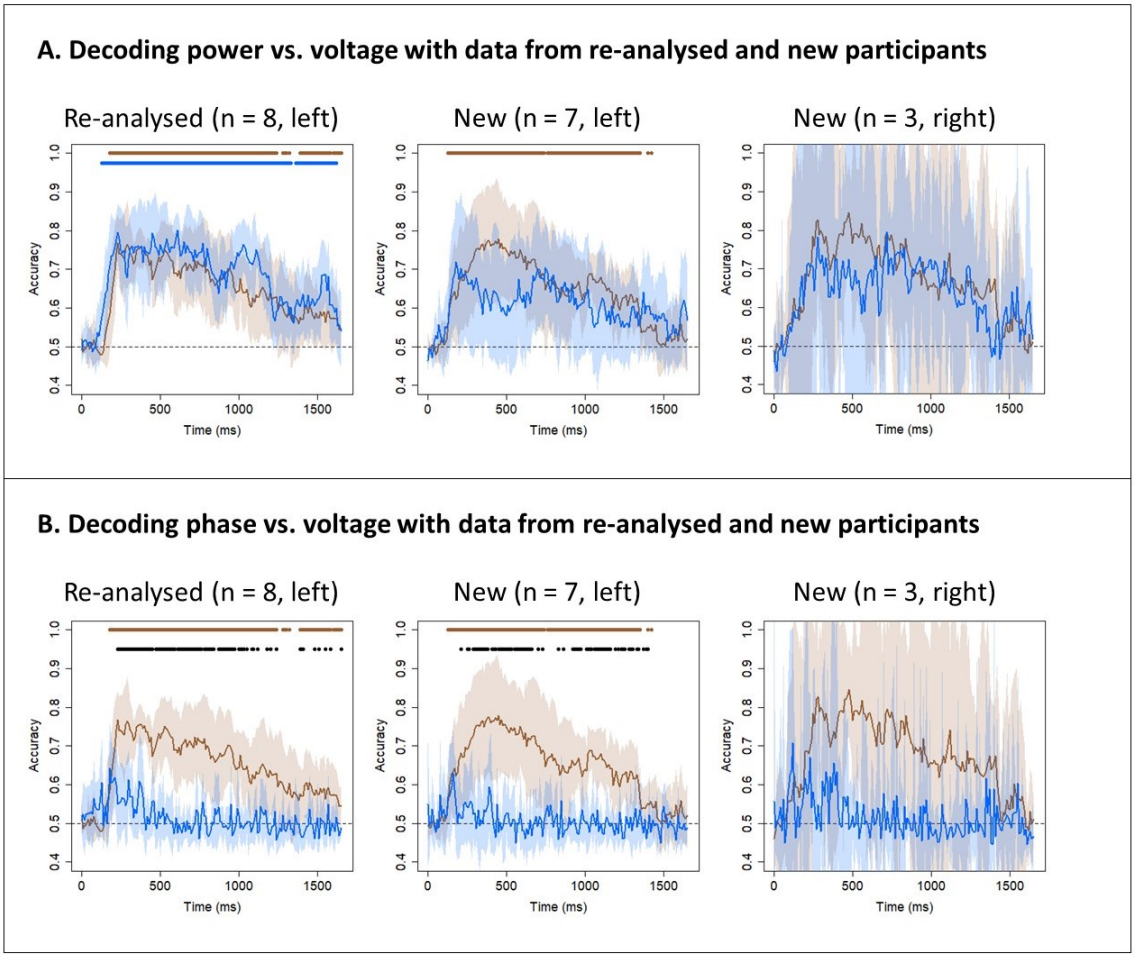

Figure B.6: Decoding subsamples of patients. (A) Mean and 95% confidence interval of the hold-out accuracy for classifiers trained on power frequency features for all frequencies between 4 and 200 Hz (blue) or on voltage features (brown) from three subsamples of patients – the re-analysed 8 patients (originally analysed by Rogers et al., 2021), the 7 left-hemisphere patients analysed for the first time in this work, and the 3 right-hemisphere patients analysed for the first time in this work. Coloured dots indicate a significant difference between classifier accuracy and chance (0.5, one-sample t-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Black dots indicate a significant difference between accuracies at a given timepoint (paired t-tests with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (B) Mean and 95% confidence interval of the hold-out accuracy for classifiers trained on phase frequency features (blue) or on voltage features (brown). Coloured dots indicate a significant difference between classifier accuracy and chance; black dots indicate a significant difference between accuracies.

Appendix C

Supplementary methods for Chapter 3: The impact of preprocessing on decoding accuracy

C.1 Introduction

We formally checked two possible technical issues within the preprocessing pipeline. The first was that preprocessing may remove signal as well as noise (de Cheveigné, 2023; Delorme, 2023) and so may harm decoding performance. The second was that, since electrooculogram (EOG) data were available for only 5 patients, it was not possible to identify and correct for macrosaccades directly. This factor is important to examine given that saccadic activity may differ between conditions – the silhouettes of living things tend to be more similar to each other than the silhouettes of nonliving things are, which means that greater attention to visual detail is needed to identify living things (G. W. Humphreys & Forde, 2001). There is a notional possibility, therefore, that if microsaccades can influence the cortical electrode data, then this difference in eye movement might be enough to drive or contribute to decoding performance.

C.2 Methods

C.2.1 Data analysis

C.2.1.1 Preprocessing - ECoG

Preprocessing followed the pipeline described in the main text. To summarise, we implemented the following steps:

1. Filtering (CleanLine to remove line noise at 60 Hz and the harmonics 120 and 180 Hz, then filtering with low cutoff 0.5 Hz and high cutoff 300 Hz)
2. Rejection of channels below the seizure onset zone or with poor contact
3. Application of a common average reference

4. Rejection of trials that contained obvious interictal epileptiform activity, muscle activity, "electrode pop" or other artefacts

Raw data and data after each preprocessing step (to illustrate – data after filtering, data after filtering *and* channel rejection, etc.) were epoched between –1000 and 3000 ms relative to stimulus onset and baseline-corrected using the mean response across trials between –200 and –1 ms. Data from the 9 patients recorded at 2000 Hz was then downsampled to match the 10 patients recorded at 1000 Hz by boxcar-averaging pairs of neighbouring timepoints. For the five patients for whom we had EOG data, we also epoched, baseline-corrected and downsampled the EOG data (four patients had a single EOG channel and the other had two). No other preprocessing was performed on the EOG data.

Time-frequency power and phase were extracted from data at each preprocessing step using complex Morlet wavelet convolution with the same parameters used in the main analysis. Power was averaged across repeated presentations of the same stimulus, any missing trials were interpolated, and decibel normalisation was performed using the same parameters used in the main analysis. For phase, preprocessed trials were averaged across repeated presentations of the same stimulus before phase values were extracted. Power and phase were extracted from EOG data in the same way. As a comparison, preprocessed voltage was averaged over repeated presentations of the same stimulus.

C.2.1.2 Multivariate classification

C.2.1.2.1 Decoding approach

The decoding approach was identical to that described in the main text.

C.2.1.2.2 Experimental questions

To address the first query (whether preprocessing may remove signal as well as noise), we created frequency feature vectors (including all 60 frequencies) and voltage feature vectors from raw data and data after each step of preprocessing. We used these as input to classifiers and compared each group-average timecourse to chance (0.5) using one-tailed, one-sample *t*-tests with false discovery rate correction as described in the main text. Where there was a visible difference in accuracy before and after a preprocessing step, we compared the timecourse before that step to the timecourse after that step using paired *t*-tests with false discovery rate correction as described in the main text.

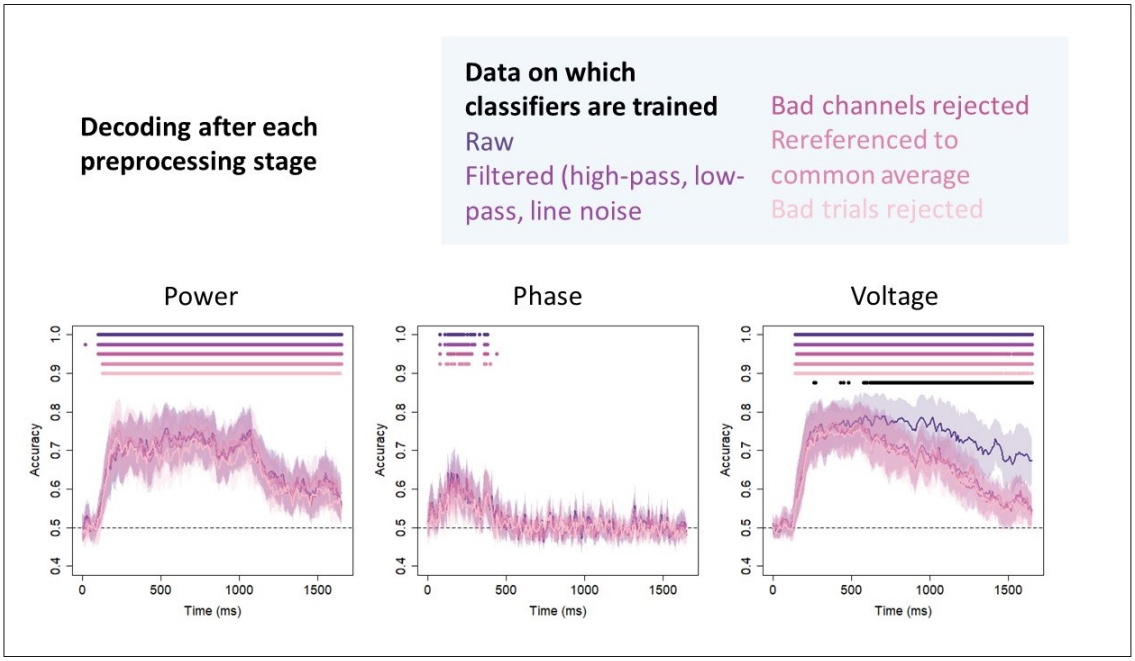

Figure C.1: Decoding during preprocessing. Mean and 95% confidence interval of the hold-out accuracy for classifiers trained on power or phase frequency features, including all 60 frequencies between 4 and 200 Hz, or on voltage features. Classifiers are trained after each preprocessing step – none (raw data; dark purple), after filtering (light purple), after bad channels were rejected (dark pink), after common average referencing (medium pink), and after bad trials were rejected (pale pink). Coloured dots indicate a significant difference between classifier accuracy and chance (0.5, one-sample t-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Black dots indicate a significant difference between accuracy for classifiers trained on voltage extracted from raw data (dark purple) and classifiers trained on voltage extracted from data after filtering (light purple) at a given timepoint (paired t-tests with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$).

To address the second query (whether saccadic activity may drive decoding performance), we created frequency feature vectors and voltage feature vectors from power, phase and voltage values extracted from EOG data. We hypothesised that eye movements may affect decoding results from different frequency ranges differently and so we created frequency vectors for all 60 frequencies and then for each range individually. We averaged the timecourses over the 5 participants with EOG data available and used paired t-tests with false discovery rate correction to compare the group-average timecourse of decoding EOG power, phase, or voltage to the average timecourse of decoding ECoG power, phase, or voltage.

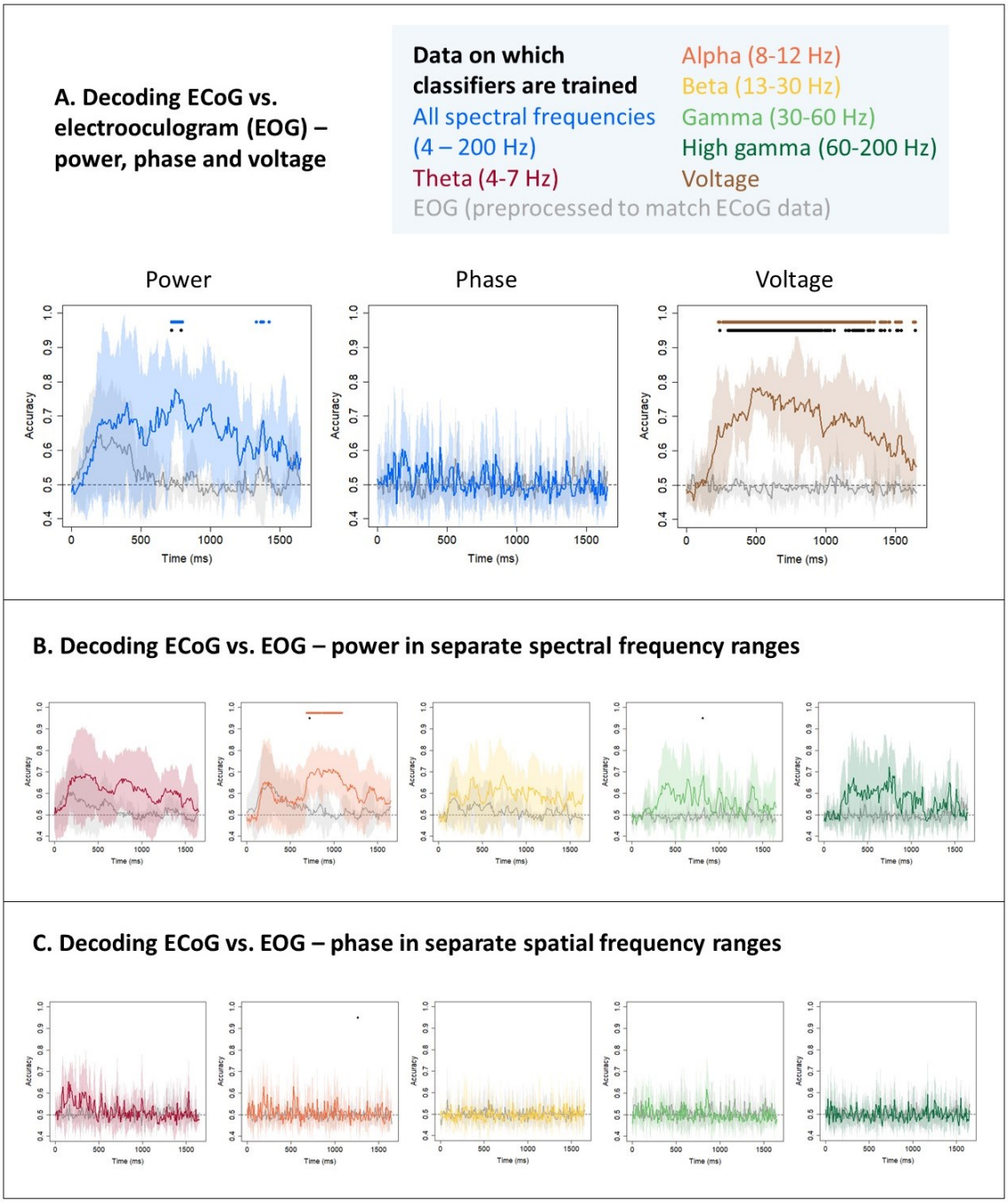

Figure C.2: Decoding electrooculogram (EOG) data. (A) Mean and 95% confidence interval of the hold-out accuracy for classifiers trained on power or phase frequency features for all frequencies between 4 and 200 Hz (blue) or on voltage features (brown), compared to classifiers trained on frequency features or on voltage features extracted from EOG data (grey). Coloured dots indicate a significant difference between classifier accuracy and chance (0.5, one-sample t-test with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). Black dots indicate a significant difference between accuracies at a given timepoint (paired t-tests with probabilities adjusted to control the false-discovery rate at $\alpha = 0.05$). (B) Mean and 95% confidence interval of the hold-out accuracy for classifiers trained on frequency features composed only of power values extracted from ECoG data within a given range – theta (4 – 7 Hz, red), alpha (12 – 18 Hz, orange), beta (13 – 30 Hz, yellow), gamma (30 – 60 Hz, light green)

Figure C.2: and high gamma (42 – 200 Hz, dark green) - compared to classifiers trained on frequency features composed of power or phase values extracted from EOG data within that same range (grey). Coloured dots indicate a significant difference between classifier accuracy and chance; black dots indicate a significant difference between accuracies. (C) Mean and 95% confidence interval of the hold-out accuracy for classifiers trained on frequency features composed only of phase values extracted from ECoG data within a given range – theta (4 – 7 Hz, red), alpha (12 – 18 Hz, orange), beta (13 – 30 Hz, yellow), gamma (30 – 60 Hz, light green) and high gamma (42 – 200 Hz, dark green) - compared to classifiers trained on frequency features composed of power or phase values extracted from EOG data within that same range (grey). Coloured dots indicate a significant difference between classifier accuracy and chance; black dots indicate a significant difference between accuracies.

C.3 Results

Figure C.1 shows the decoding profile of classifiers fitted to power, phase and voltage data – first to raw data, then to data after each preprocessing step (filtering, channel rejection, rereferencing and trial rejection). The decoding profile of time-frequency power and phase hardly changed when preprocessing steps were applied (a slight decrease in accuracy pushed phase decoding below the threshold for significance at some timepoints). By contrast, while voltage was significantly decodable throughout the time window of interest, the shape of the decoding profile changed after filtering – decoding accuracy reached the same maximum but declined more steeply from that point. After around 500 ms, accuracy for classifiers fitted to filtered data was significantly poorer than accuracy for classifiers fitted to raw data.

Figure C.2 shows the decoding profile of classifiers trained on power, phase and voltage extracted from EOG data. The only consistent significant differences were between voltage and the EOG analogue of voltage.

C.4 Discussion

We investigated two aspects of the preprocessing pipeline: the effect of removing noise and saccadic activity.

First we found that heavier cleaning hardly affected decoding with time-frequency power, subtly affected decoding with time-frequency phase, and had a significant effect on decoding with voltage. The absence of an influence of noise on decoding with time-frequency power was important to establish given that artefacts in the data could be “smeared” across (and therefore obscure) true signal during wavelet convolution. One possible reason for this result that we used

L1 (LASSO) regularisation (Tibshirani, 1996), which selects a sparse subset of the features offered to the classifier. Features obscured by artefacts may be removed by the classifier and decoding performance will be good as long as enough unobscured features remain. Regularisation methods that assume that the code is dense, such as L2 (ridge) regularisation (Hoerl & Kennard, 1970), are forced to place weights on all features and so the presence of artefacts may be deleterious to decoding when a different type of classifier is used.

The significant difference for voltage appeared after filtering: before filtering, the decoding profile resembled that obtained by Rogers et al. (2021). Three filters were applied to the data – CleanLine filtering (Mitra & Bokil, 2007) for line noise at 60 Hz and the harmonics, high-pass filtering at 0.5 Hz, and low-pass filtering at 300 Hz. For 9 patients whose data were recorded at 1000 Hz, the low-pass filter was imposed by the recording equipment – since this was the same for Rogers et al. (2021), who used no low-pass filter, the low-pass filter is unlikely to be responsible for the decrease in accuracy. Line noise is constant over time and so does not differ between living and nonliving trials. Therefore, the culprit is likely to be the high-pass filter – filtering very low-frequency activity out may change the shape of the voltage deflection over time and reduce its correlation with animacy.

We found that there were few significant differences between the decoding profiles of power and phase extracted from ECoG data and the decoding profiles of power and phase extracted from EOG channels. It should be noted that it is difficult to draw firm conclusions from this analysis because the sample size was small – only 5 of our 19 patients had EOG data. A larger sample size may reinforce the conclusion that there is no significant difference. Alternatively, on visual inspection, the profiles of EOG decoding and ECoG decoding appear to differ and this difference may solidify with a larger number of participants. Until we have clarified the extent to which eye movements may contain information about semantic category, we should take care to correct for eye movements as best we can. Future ECoG data collection projects should take care to record EOG data so that they can be subtracted from the ECoG signal.

To summarise, preprocessing aims to eliminate the possibility of false-positive results (for example, significant decoding driven by a difference in eye movement between categories) and false-negative results (for example, artefacts destroying signal during time-frequency decomposition). We found evidence that most stages of preprocessing make little, if any, difference to time-frequency power and phase decoding with logistic regression with L1

(LASSO) regularisation. Filtering can remove signal as well as noise from voltage data; future work should avoid, or exercise caution when, applying heavy preprocessing to voltage data (Delorme, 2023). Evidence for significant ECoG decoding above that which can be achieved with EOG is inconclusive; future work should seek to clarify the extent to which EOG can cause false positive decoding.

Appendix D

Supplementary results for Chapter 4

Table D.1: Significant cluster and peak information for main effects, ANOVA effects of interest and directed effects of interest when comparing activation magnitude (contrast betas) and activation precision (statistical *t*-values). All coordinates are given in MNI space and all anatomical labels are extracted from the Harvard-Oxford cortical and subcortical structural atlases. SESB = single-echo single band, pTx = parallel transmit, SEMB = single echo multiband, MESB = multi-echo single band, MEMB = multi-echo multiband, ME = multi-echo without ICA denoising, SE = single-echo, MB = multiband, SB = single band, MEdn = multi-echo with ICA denoising, MBodd = multiband downsampled to match the number of volumes in single band.

	Cluster extent (voxels)	z-value	x	y	z	Anatomical label
Activation magnitude						
Comparing pTx and SESB						
SESB	8941	7.55	-29	-42	-13	Left temporal fusiform cortex pos
		7.39	34	-47	-20	Right temporal occipital fusiform cortex
		7.16	-31	-34	-23	Left temporal fusiform cortex pos
		6.93	-39	-52	-18	Left temporal occipital fusiform cortex
		6.83	31	-57	-18	Right temporal occipital fusiform cortex
		6.76	26	-42	-18	Right temporal occipital fusiform cortex
		6.71	-49	-77	17	Left lateral occipital cortex sup
		6.57	-41	-74	4	Left lateral occipital cortex inf
	2294	6.75	-46	43	-3	Left frontal pole
		6.73	-54	40	-8	Left frontal pole

		6.72	-49	33	10	Left inferior frontal gyrus p tri
		6.43	-44	20	20	Left inferior frontal gyrus p ope
		6.18	-51	36	24	Left frontal pole
		6.13	-39	6	34	Left middle frontal gyrus
		5.94	-36	16	24	Left inferior frontal gyrus p ope
		5.5	-41	8	22	Left inferior frontal gyrus p ope
	232	5.91	-6	33	44	Left superior frontal gyrus
	182	5.59	-26	16	42	Left middle frontal gyrus
		3.6	-39	20	57	Left middle frontal gyrus
	160	5.01	41	23	20	Right middle frontal gyrus
		4.27	36	16	30	Right middle frontal gyrus
	66	4.3	41	33	-8	Right frontal orbital cortex
	110	4.21	-6	-60	17	Left precuneous cortex
		4.1	-4	-52	12	Left precuneous cortex
		3.89	-11	-57	7	Left precuneous cortex
pTx	11682	7.79	-29	-42	-13	Left temporal fusiform cortex pos
		7.69	34	-47	-20	Right temporal occipital fusiform cortex
		7.52	36	-67	-16	Right occipital fusiform gyrus
		7.52	-39	-52	-18	Left temporal occipital fusiform cortex
		7.43	-36	-42	-20	Left temporal fusiform cortex pos

	7.34	41	-42	-18	Right temporal occipital fusiform cortex
	7.3	-29	-32	-20	Left parahippocampal gyrus pos
	7.17	46	-77	-8	Right lateral occipital cortex inf
3242	6.79	-54	40	-8	Left frontal pole
	6.72	-51	36	24	Left frontal pole
	6.67	-46	43	-3	Left frontal pole
	6.67	-49	33	10	Left inferior frontal gyrus p tri
	6.63	-39	6	34	Left middle frontal gyrus
	6.59	-44	20	20	Left inferior frontal gyrus p ope
	6.22	-51	20	30	Left middle frontal gyrus
	5.83	-29	18	62	Left superior frontal gyrus
501	6.1	14	-74	-30	Right cerebellum
	6.06	9	-82	-30	Right cerebellum
	5.34	34	-74	-48	Right cerebellum
296	5.57	44	26	20	Right inferior frontal gyrus p tri
	4.59	36	16	30	Right middle frontal gyrus
	3.28	49	33	34	Right middle frontal gyrus
97	4.61	24	-60	20	Right precuneous cortex
	4.03	19	-50	10	Right precuneous cortex
	4.01	11	-52	14	Right precuneous cortex
146	4.4	34	38	-13	Right frontal pole
	4.2	31	33	-20	Right frontal pole
	3.63	34	30	0	Right frontal orbital cortex
	3.53	51	33	-16	Right frontal orbital cortex
	3.43	46	26	-8	Right frontal orbital cortex

ANOVA	126	4.62	36	-67	-16	Right occipital fusiform gyrus
		4.08	39	-77	-10	Right lateral occipital cortex inf
		3.9	31	-80	-13	Right occipital fusiform gyrus
		3.39	51	-72	0	Right lateral occipital cortex inf
pTx>SESB	172	4.76	36	-67	-16	Right occipital fusiform gyrus
		4.24	39	-77	-10	Right lateral occipital cortex inf
		4.07	31	-80	-13	Right occipital fusiform gyrus
		3.57	51	-72	0	Right lateral occipital cortex inf
		3.18	54	-70	7	Right lateral occipital cortex inf
2 × 2 factorial design – echo and band						
SESB	11459	Inf	-31	-44	-20	Left temporal fusiform cortex pos
		Inf	34	-47	-20	Right temporal occipital fusiform cortex
		Inf	36	-40	-23	Right temporal occipital fusiform cortex
		Inf	-39	-57	-16	Left temporal occipital fusiform cortex
		Inf	46	-80	0	Right lateral occipital cortex inf
		Inf	-44	-77	4	Left lateral occipital cortex inf
		Inf	49	-67	12	Right lateral occipital cortex inf
		Inf	31	-44	-13	Right temporal occipital fusiform cortex

	2394	Inf	-49	33	10	Left inferior frontal gyrus p tri
		7.68	-39	10	24	Left inferior frontal gyrus p ope
		7.25	-46	20	22	Left inferior frontal gyrus p ope
		7.2	-46	43	-3	Left frontal pole
		7.12	-49	33	-3	Left inferior frontal gyrus p tri
		6.84	-44	40	-10	Left frontal pole
		6.66	-56	18	27	Left inferior frontal gyrus p ope
		6.62	-59	23	14	Left inferior frontal gyrus p tri
	304	7.68	39	20	20	Right inferior frontal gyrus p ope
		4.78	54	30	12	Right inferior frontal gyrus p tri
	276	6.44	-6	23	50	Left superior frontal gyrus
		5.12	-6	33	44	Left superior frontal gyrus
	206	5.89	-29	20	52	Left middle frontal gyrus
		4.09	-39	20	57	Left middle frontal gyrus
		3.78	-34	10	64	Left middle frontal gyrus
	140	5.4	34	-70	-46	Right cerebellum
		3.49	24	-77	-43	Right cerebellum
	109	4.95	11	-77	-28	Right cerebellum
		4.06	14	-80	-36	Right cerebellum
	113	4.4	-9	-52	14	Left precuneous cortex
		4.33	-11	-54	4	Left precuneous cortex
SEMB	20421	Inf	-31	-44	-20	Left temporal fusiform cortex pos
		Inf	34	-47	-20	Right temporal occipital fusiform cortex

		Inf	49	-67	12	Right lateral occipital cortex inf
		Inf	44	-77	-3	Right lateral occipital cortex inf
		Inf	-39	-57	-16	Left temporal occipital fusiform cortex
		Inf	31	-34	-20	Right temporal fusiform cortex pos
		Inf	39	-40	-23	Right temporal occipital fusiform cortex
		Inf	36	-64	-16	Right occipital fusiform gyrus
	568	7.7	39	20	17	Right inferior frontal gyrus p ope
		7.46	49	28	17	Right inferior frontal gyrus p tri
		6.37	39	16	30	Right middle frontal gyrus
		3.95	19	10	32	Right lateral ventricle
	225	5.46	44	30	-6	Right frontal orbital cortex
		4.57	34	33	-18	Right frontal pole
	203	5.09	-14	3	10	left anterior thalamic radiation
		5.05	-9	-12	4	left thalamus
MESB	22950	Inf	-31	-44	-20	Left temporal fusiform cortex pos
		Inf	34	-47	-20	Right temporal occipital fusiform cortex
		Inf	-44	-50	-16	Left inferior temporal gyrus temocc
		Inf	-39	-57	-18	Left temporal occipital fusiform cortex
		Inf	36	-67	-16	Right occipital fusiform gyrus
		Inf	34	-60	-18	Right temporal occipital fusiform cortex

		Inf	-41	-40	-18	Left temporal fusiform cortex pos
		Inf	-34	-70	-16	Left occipital fusiform gyrus
	733	Inf	41	20	22	Right inferior frontal gyrus p ope
		Inf	49	28	20	Right inferior frontal gyrus p tri
		6.23	54	33	14	Right inferior frontal gyrus p tri
		4.57	54	30	4	Right inferior frontal gyrus p tri
	468	7.75	31	36	-13	Right frontal pole
		5.91	39	28	-3	Right frontal orbital cortex
	212	6.95	-11	-82	-33	Left cerebellum
MEMB	20648	Inf	-34	-44	-23	Left temporal fusiform cortex pos
		Inf	36	-42	-23	Right temporal occipital fusiform cortex
		Inf	-36	-57	-16	Left temporal occipital fusiform cortex
		Inf	-44	-50	-16	Left inferior temporal gyrus temocc
		Inf	34	-60	-18	Right temporal occipital fusiform cortex
		Inf	36	-67	-16	Right occipital fusiform gyrus
		Inf	-34	-70	-16	Left occipital fusiform gyrus
		Inf	-44	-40	-18	Left temporal fusiform cortex pos
	461	7.65	41	23	20	Right middle frontal gyrus
		7.24	49	28	20	Right inferior frontal gyrus p tri
	234	6.64	31	36	-13	Right frontal pole
		4.67	41	30	-6	Right frontal orbital cortex
	167	6.09	-9	-84	-33	Left cerebellum

ANOVA	268	5.3	-44	-40	-18	Left temporal fusiform cortex pos
		4.71	-51	-57	-20	Left inferior temporal gyrus temocc
		4.39	-59	-50	-16	Left inferior temporal gyrus temocc
		4.09	-36	-44	-26	Left temporal occipital fusiform cortex
		3.77	-41	-47	-8	left inferior longitudinal fas
		3.63	-66	-54	-6	Left middle temporal gyrus temocc
	108	4.44	-34	18	20	Left inferior frontal gyrus p ope
		4.09	-36	18	32	Left middle frontal gyrus
		3.61	-44	26	22	Left inferior frontal gyrus p tri
		3.33	-39	26	14	Left inferior frontal gyrus p tri
	80	4.31	41	-54	17	Right angular gyrus
		4.08	46	-67	10	Right lateral occipital cortex inf
		3.64	29	-60	0	Right lingual gyrus
		3.62	34	-70	4	right inferior longitudinal fas
	123	4.27	26	-47	-3	Right lingual gyrus
		4.18	34	-44	-16	Right temporal occipital fusiform cortex
		3.97	24	-32	-13	right cingulum hipp
		3.86	31	-24	-20	Right parahippocampal gyrus pos

		3.5	34	-34	-20	Right temporal fusiform cortex pos
ME>SE	620	5.98	-44	-40	-18	Left temporal fusiform cortex pos
		5.27	-51	-57	-20	Left inferior temporal gyrus temocc
		5.12	-59	-50	-16	Left inferior temporal gyrus temocc
		4.6	-36	-44	-26	Left temporal occipital fusiform cortex
		4.51	-41	-47	-10	left inferior longitudinal fas
		4.43	-66	-54	-6	Left middle temporal gyrus temocc
		4.18	-44	-24	-20	Left inferior temporal gyrus post
		3.53	-36	-24	-26	Left temporal fusiform cortex pos
	101	5.44	34	-40	-28	Right temporal fusiform cortex pos
		3.55	46	-40	-20	Right inferior temporal gyrus temocc
	117	5.11	-21	36	-10	Left frontal orbital cortex
		3.68	-34	36	-10	Left frontal orbital cortex
		3.43	-9	23	-13	Left subcallosal cortex
2 × 2 factorial design – denoising and band						
MEdn>ME	323	4.62	-31	-47	2	Left lateral ventricle
		4.22	-14	-34	4	left thalamus
		4.13	-29	-37	14	left anterior thalamic radiation
		4.08	4	-32	12	3rd ventricle
		4.08	-16	-42	17	Left lateral ventricle
		4.05	11	-32	4	right thalamus

	104	4.55	-1	-52	-56	Left cerebellum
		3.75	-11	-40	-50	brain stem
		3.59	4	-54	-66	brain stem
	135	3.94	29	-34	4	Right lateral ventricle
		3.81	29	-34	17	Right lateral ventricle
		3.79	21	-24	22	right corticospinal tract
		3.64	31	-44	0	Right lateral ventricle
2 × 2 factorial design – echo and (downsampled) band						
ANOVA	224	4.77	-49	-52	-20	Left inferior temporal gyrus temocc
		4.72	-44	-40	-18	Left temporal fusiform cortex pos
		4.02	-41	-47	-8	left inferior longitudinal fas
		4.02	-59	-50	-16	Left inferior temporal gyrus temocc
		3.68	-39	-44	-26	Left temporal fusiform cortex pos
	85	4.34	41	-57	17	Right angular gyrus
		4.18	44	-67	10	Right lateral occipital cortex inf
	89	4.19	-34	13	20	Left inferior frontal gyrus p ope
		4.04	-36	18	32	Left middle frontal gyrus
		3.56	-44	28	22	Left middle frontal gyrus
Activation precision						
Comparing pTx and SESB						
SESB	9906	7.62	-29	-42	-13	Left temporal fusiform cortex pos
		7.35	39	-44	-18	Right temporal occipital fusiform cortex

	7.29	34	-52	-20	Right temporal occipital fusiform cortex
	7.18	-34	-50	-20	Left temporal occipital fusiform cortex
	7.18	34	-67	-16	Right occipital fusiform gyrus
	7.06	24	-42	-18	Right temporal occipital fusiform cortex
	7.01	36	-60	-18	Right temporal occipital fusiform cortex
	7	51	-70	7	Right lateral occipital cortex inf
2162	6.27	-54	40	-8	Left frontal pole
	5.97	-51	36	10	Left inferior frontal gyrus p tri
	5.91	-41	13	22	Left inferior frontal gyrus p ope
	5.89	-54	30	17	Left inferior frontal gyrus p tri
	5.68	-36	36	-13	Left frontal orbital cortex
	5.56	-51	20	22	Left inferior frontal gyrus p ope
	5.36	-46	36	-16	Left frontal pole
	5.3	-39	6	34	Left middle frontal gyrus
383	5.47	-24	16	44	Left superior frontal gyrus
	5.14	-9	23	47	Left superior frontal gyrus
	4.97	-29	23	57	Left superior frontal gyrus
	4.68	-6	30	44	Left superior frontal gyrus
	3.81	-14	18	42	Left paracingulate gyrus
165	5.43	31	0	-36	Right temporal fusiform cortex ant

		3.69	31	-10	-36	Right temporal fusiform cortex pos
	185	5.19	44	26	17	Right inferior frontal gyrus p tri
	129	4.83	14	-74	-30	Right cerebellum
		3.73	14	-90	-30	Right cerebellum
	121	4.59	-6	-60	17	Left precuneous cortex
		3.7	-16	-57	10	Left precuneous cortex
	91	4.44	36	-64	-38	Right cerebellum
		3.9	34	-67	-48	Right cerebellum
		3.8	29	-72	-40	Right cerebellum
pTx	12474	7.73	34	-67	-16	Right occipital fusiform gyrus
		7.7	34	-52	-20	Right temporal occipital fusiform cortex
		7.7	-29	-42	-13	Left temporal fusiform cortex pos
		7.68	39	-44	-18	Right temporal occipital fusiform cortex
		7.53	24	-42	-18	Right temporal occipital fusiform cortex
		7.32	36	-60	-18	Right temporal occipital fusiform cortex
		7.26	-36	-50	-20	Left temporal occipital fusiform cortex
		7.21	34	-37	-20	Right temporal fusiform cortex pos
	2923	6.34	-54	40	-8	Left frontal pole
		6.22	-54	30	17	Left inferior frontal gyrus p tri
		6	-41	16	22	Left inferior frontal gyrus p ope
		5.97	-39	6	34	Left middle frontal gyrus
		5.95	-51	40	7	Left frontal pole
		5.92	-46	43	-3	Left frontal pole

		5.86	-46	36	-16	Left frontal pole
		5.81	-51	38	24	Left frontal pole
	333	5.82	44	26	17	Right inferior frontal gyrus p tri
		4.17	36	16	30	Right middle frontal gyrus
	637	5.74	9	-84	-33	Right cerebellum
		5.44	14	-77	-30	Right cerebellum
		5.26	34	-74	-53	Right cerebellum
		4.69	36	-70	-38	Right cerebellum
		4.26	24	-82	-48	Right cerebellum
	336	5.21	-6	30	44	Left superior frontal gyrus
		4.4	-6	43	30	Left paracingulate gyrus
		4.36	-1	28	52	Left superior frontal gyrus
		3.47	-11	53	37	Left frontal pole
	179	4.7	24	-60	17	Right supracalcarine cortex
		4.41	21	-57	10	Right precuneous cortex
		4.27	9	-52	10	Right precuneous cortex
	159	4.43	34	38	-16	Right frontal pole
		4.11	51	33	-16	Right frontal orbital cortex
		4.06	34	30	0	Right frontal orbital cortex
		3.4	44	38	-23	Right frontal pole
2 × 2 factorial design – echo and band						
SESB	9047	Inf	-29	-44	-18	Left temporal occipital fusiform cortex
		Inf	36	-42	-20	Right temporal occipital fusiform cortex
		Inf	-29	-52	-10	Left temporal occipital fusiform cortex
		Inf	49	-67	12	Right lateral occipital cortex inf
		Inf	31	-60	-16	Right temporal occipital fusiform cortex

		7.7	26	-50	-13	Right temporal occipital fusiform cortex
		7.69	-46	-74	12	Left lateral occipital cortex inf
		7.68	34	-34	-23	Right temporal fusiform cortex pos
	1738	6.46	-49	33	7	Left inferior frontal gyrus p tri
		6.37	-39	13	22	Left inferior frontal gyrus p ope
		6.19	-44	26	12	Left inferior frontal gyrus p tri
		6.18	-46	20	20	Left inferior frontal gyrus p ope
		5.13	-41	40	-10	Left frontal pole
		5.12	-44	33	-13	Left frontal orbital cortex
		5.11	-46	46	0	Left frontal pole
		4.85	-49	38	-18	Left frontal pole
	213	6.28	36	20	20	Right inferior frontal gyrus p ope
		5.53	49	28	17	Right inferior frontal gyrus p tri
	136	5.16	-6	23	47	Left superior frontal gyrus
		4.04	-9	33	44	Left superior frontal gyrus
	123	4.62	-29	20	54	Left middle frontal gyrus
SEMB	22882	Inf	-29	-47	-18	Left temporal occipital fusiform cortex
		Inf	46	-67	12	Right lateral occipital cortex inf
		Inf	34	-34	-23	Right temporal fusiform cortex pos
		Inf	-46	-77	10	Left lateral occipital cortex inf

		Inf	29	-54	-13	Right temporal occipital fusiform cortex
		Inf	36	-42	-20	Right temporal occipital fusiform cortex
		Inf	-26	-52	-10	Left temporal occipital fusiform cortex
		Inf	-41	-80	-10	Left lateral occipital cortex inf
	452	Inf	51	28	17	Right inferior frontal gyrus p tri
		4.7	39	16	30	Right middle frontal gyrus
	248	5.42	34	33	-20	Right frontal pole
		5.22	44	33	-8	Right frontal pole
	104	5.36	6	-62	-50	Right cerebellum
		4.9	9	-54	-43	Right cerebellum
	198	5.35	-6	20	-28	Left subcallosal cortex
		3.55	-4	43	-30	Left frontal medial cortex
		3.37	-11	10	-26	Left frontal orbital cortex
		3.32	6	23	-26	Right subcallosal cortex
	210	5.35	-9	-12	7	left thalamus
		5.03	-14	3	4	left anterior thalamic radiation
	141	5.16	-9	-82	-33	Right cerebellum
	104	5.1	-9	63	-18	Left frontal pole
MESB	23065	Inf	-31	-44	-18	Left temporal occipital fusiform cortex
		Inf	36	-42	-20	Right temporal occipital fusiform cortex
		Inf	-39	-42	-16	Left temporal fusiform cortex pos
		Inf	-46	-74	10	Left lateral occipital cortex inf
		Inf	34	-34	-23	Right temporal fusiform cortex pos

		Inf	-26	-52	-13	Left temporal occipital fusiform cortex
		Inf	34	-54	-18	Right temporal occipital fusiform cortex
		Inf	-44	-50	-13	Left inferior temporal gyrus temocc
	1169	7.74	36	20	20	Right inferior frontal gyrus p ope
		7.72	49	28	17	Right inferior frontal gyrus p tri
		6.92	34	36	-13	Right frontal pole
		6.03	31	10	30	Right middle frontal gyrus
		5.87	41	28	-3	Right frontal orbital cortex
	216	5.84	-9	-82	-33	Left cerebellum
	130	5.24	6	-54	-43	Right cerebellum
		5.01	4	-62	-50	Right cerebellum
		3.45	-6	-60	-48	Left cerebellum
MEMB	23862	Inf	-31	-44	-18	Left temporal occipital fusiform cortex
		Inf	-36	-52	-18	Left temporal occipital fusiform cortex
		Inf	-44	-77	2	Left lateral occipital cortex inf
		Inf	-49	-74	10	Left lateral occipital cortex inf
		Inf	34	-44	-20	Right temporal occipital fusiform cortex
		Inf	-41	-42	-18	Left temporal fusiform cortex pos
		Inf	-31	-64	-13	Left occipital fusiform gyrus
		Inf	-44	-50	-13	Left inferior temporal gyrus temocc
	1016	Inf	49	28	17	Right inferior frontal gyrus p tri

		7.76	39	23	17	Right inferior frontal gyrus p tri
		6.83	34	36	-13	Right frontal pole
		6.42	36	13	32	Right middle frontal gyrus
		5.55	41	28	-3	Right frontal orbital cortex
		4.66	49	33	7	Right frontal pole
	223	6.3	-9	-84	-36	Left cerebellum
	143	5.46	4	-57	-48	Right cerebellum
		5.3	6	-54	-40	Right cerebellum
		3.66	-6	-57	-48	Left cerebellum
	90	5.28	-21	68	17	Left frontal pole
ANOVA	2729	6.34	-34	-44	-20	Left temporal fusiform cortex pos
		6.15	-44	-40	-18	Left temporal fusiform cortex pos
		6.14	-44	-77	2	Left lateral occipital cortex inf
		6.07	-49	-77	10	Left lateral occipital cortex inf
		5.99	-44	-50	-16	Left inferior temporal gyrus temocc
		5.88	-36	-72	-16	Left occipital fusiform gyrus
		5.69	-41	-80	-10	Left lateral occipital cortex inf
		5.35	-51	-44	-13	left superior longitudinal fas
	2159	5.73	46	-67	12	Right lateral occipital cortex inf
		5.64	36	-40	-28	Right temporal fusiform cortex pos
		5.6	36	-74	-13	Right occipital fusiform gyrus
		5.44	46	-77	-3	Right lateral occipital cortex inf

	5.19	29	-60	-13	Right temporal occipital fusiform cortex
	5.17	44	-32	-23	Right inferior temporal gyrus post
	5.05	46	-77	7	Right lateral occipital cortex inf
	4.96	36	-52	-20	Right temporal occipital fusiform cortex
188	5.01	-21	36	-10	Left frontal orbital cortex
	4.06	6	6	-16	Right subcallosal cortex
	3.84	4	13	-13	Right subcallosal cortex
	3.81	-9	26	-13	Left subcallosal cortex
	3.5	-34	33	-13	Left frontal orbital cortex
90	5	34	3	-38	Right temporal pole
	4.13	34	-12	-30	Right parahippocampal gyrus ant
	4.12	34	8	-48	Right temporal pole
501	4.89	-34	13	20	Left inferior frontal gyrus p ope
	4.68	-36	18	30	Left middle frontal gyrus
	4.56	-46	28	22	Left inferior frontal gyrus p tri
	4.31	-46	33	7	Left inferior frontal gyrus p tri
	4.17	-54	23	30	Left middle frontal gyrus
	4.06	-44	26	10	Left inferior frontal gyrus p tri
	3.29	-56	33	14	Left inferior frontal gyrus p tri

	97	4.87	-31	-74	47	Left lateral occipital cortex sup
	613	4.79	-24	-94	10	Left occipital pole
		4.56	-21	-94	20	Left occipital pole
		3.83	-9	-90	17	Left occipital pole
		3.81	-21	-90	32	Left occipital pole
	109	4.74	31	36	-10	Right frontal pole
		4.07	14	30	-10	right uncinate fas
	80	4.36	14	-82	-46	Right cerebellum
		3.93	31	-70	-43	Right cerebellum
		3.48	36	-77	-46	Right cerebellum
ME>SE	1309	6.42	-44	-40	-18	Left temporal fusiform cortex pos
		5.93	-51	-44	-13	left superior longitudinal fas
		5.69	-41	-47	-8	left inferior longitudinal fas
		5.14	-54	-57	-20	Left inferior temporal gyrus temocc
		5.01	-46	-52	-20	Left inferior temporal gyrus temocc
		4.9	-56	-77	14	Left lateral occipital cortex inf
		4.71	-46	-42	10	Left superior temporal gyurs pos
		4.58	-39	-24	-20	Left temporal fusiform cortex pos
	348	6.13	36	-40	-28	Right temporal fusiform cortex pos
		4.35	24	-52	-18	Right temporal occipital fusiform cortex
		4.19	34	-22	-36	Right temporal fusiform cortex pos
		3.98	44	-32	-23	Right inferior temporal gyrus post

		3.9	34	-12	-33	Right temporal fusiform cortex pos
		3.73	46	-44	-16	Right inferior temporal gyrus temocc
		3.27	24	-44	-23	Right temporal occipital fusiform cortex
	562	5.51	-19	36	-13	Left frontal orbital cortex
		5.12	31	38	-10	Right frontal pole
		4.84	9	8	-16	Right subcallosal cortex
		4.66	14	30	-10	right uncinate fas
		4.44	-9	23	-13	Left subcallosal cortex
		3.89	-34	36	-10	Left frontal orbital cortex
		3.77	1	18	-18	Right subcallosal cortex
		3.76	11	20	-13	Right subcallosal cortex
	134	4.95	14	-82	-46	Right cerebellum
		3.64	26	-77	-46	Right cerebellum
		3.53	11	-82	-36	Right cerebellum
		3.4	14	-72	-36	Right cerebellum
	92	4.34	54	-74	-10	Right lateral occipital cortex inf
		4.32	49	-84	2	Right lateral occipital cortex inf
		4.11	54	-77	10	Right lateral occipital cortex inf
SE>ME	93	4.89	-36	10	17	Left frontal operculum cortex
		3.71	-34	-2	32	Left precentral gyrus
MB>SB	3959	6.7	-44	-77	2	Left lateral occipital cortex inf
		6.58	-49	-77	10	Left lateral occipital cortex inf

	6.53	-34	-47	-20	Left temporal occipital fusiform cortex
	6.43	-36	-72	-16	Left occipital fusiform gyrus
	6.28	-41	-80	-10	Left lateral occipital cortex inf
	6.26	-36	-54	-18	Left temporal occipital fusiform cortex
	5.95	-44	-50	-16	Left inferior temporal gyrus temocc
	5.7	-29	-54	-13	Left temporal occipital fusiform cortex
2569	6.18	46	-67	12	Right lateral occipital cortex inf
	6.1	46	-77	-3	Right lateral occipital cortex inf
	6.01	36	-74	-13	Right occipital fusiform gyrus
	5.65	31	-62	-16	Right occipital fusiform gyrus
	5.62	46	-77	7	Right lateral occipital cortex inf
	5.49	29	-90	14	Right occipital pole
	5.35	36	-52	-20	Right temporal occipital fusiform cortex
	5.3	34	-34	-23	Right temporal fusiform cortex pos
828	4.68	-54	23	30	Left middle frontal gyrus
	4.62	-46	33	12	Left inferior frontal gyrus p tri
	4.47	-46	26	22	Left inferior frontal gyrus p tri
	4.39	-39	16	30	Left middle frontal gyrus

		4.12	-51	20	42	Left middle frontal gyrus
		3.59	-44	38	-6	Left frontal pole
		3.51	-46	48	7	Left frontal pole
		3.44	-39	23	52	Left middle frontal gyrus
SB>MB	427	4.36	44	-44	57	Right superior parietal lobule
		4.2	46	-40	42	Right supramarginal gyrus pos
		3.65	56	-32	50	Right supramarginal gyrus ant
		3.64	31	-47	64	Right superior parietal lobule
		3.53	59	-40	52	Right supramarginal gyrus pos
		3.49	26	-44	57	Right superior parietal lobule
		3.35	61	-47	37	Right supramarginal gyrus pos
	130	3.95	-4	-60	64	Left precuneous cortex
		3.59	-11	-57	74	Left superior parietal lobule
		3.56	1	-52	67	Right precuneous cortex
		3.47	-21	-57	62	Left superior parietal lobule
		3.45	-4	-54	54	Left precuneous cortex
		3.37	-11	-64	62	Left lateral occipital cortex sup
2 × 2 factorial design – denoising and band						
ANOVA	6202	Inf	36	-32	-26	Right temporal fusiform cortex pos
		6.88	39	-40	-20	Right temporal occipital fusiform cortex
		6.64	26	-37	-23	Right temporal fusiform cortex pos

	6.51	-36	-44	-16	Left temporal occipital fusiform cortex
	6.45	-39	-34	-28	Left temporal fusiform cortex pos
	6.28	-44	-37	-20	Left inferior temporal gyrus post
	6.15	41	-77	-18	Right lateral occipital cortex inf
	6.03	-39	-20	-26	Left temporal fusiform cortex pos
630	4.77	-44	30	4	Left inferior frontal gyrus p tri
	4.68	-39	13	22	Left inferior frontal gyrus p ope
	4.56	-56	40	2	Left frontal pole
	4.56	-44	23	14	Left inferior frontal gyrus p tri
	4.32	-36	40	-6	Left frontal pole
	4.3	-56	30	-3	Left inferior frontal gyrus p tri
	4.12	-39	13	34	Left middle frontal gyrus
	3.92	-54	38	17	Left frontal pole
286	4.59	14	-97	27	Right occipital pole
	4.51	21	-94	20	Right occipital pole
	4.35	11	-90	22	Right occipital pole
	4	9	-92	32	Right occipital pole
	3.95	31	-100	12	Right occipital pole
118	4.55	-16	-97	0	Left occipital pole
	3.96	-26	-92	10	Left occipital pole
413	4.42	44	-40	42	Right supramarginal gyrus pos

		4.28	54	-37	42	Right supramarginal gyrus pos
		4.23	49	-37	32	right superior longitudinal fas
		3.98	44	-42	52	Right supramarginal gyrus pos
		3.86	29	-47	42	Right superior parietal lobule
		3.61	61	-47	44	Right angular gyrus
		3.5	59	-37	34	Right supramarginal gyrus pos
	91	4.39	-41	36	-20	Left frontal pole
	93	4.17	64	-24	20	Right parietal operculum cortex
		3.99	66	-27	27	Right supramarginal gyrus ant
	196	4.08	-16	-90	27	Left occipital pole
		4.03	-1	-94	22	Left occipital pole
		3.97	-4	-87	24	Left cuneal cortex
		3.91	-26	-94	30	Left occipital pole
		3.88	-29	-87	27	Left lateral occipital cortex sup
	85	3.94	-1	-50	52	Left precuneous cortex
		3.7	11	-60	54	Right precuneous cortex
		3.26	-4	-62	57	Left precuneous cortex
MEdn>ME	8383	7.46	36	-32	-23	Right temporal fusiform cortex pos
		7.22	26	-37	-23	Right temporal fusiform cortex pos
		6.76	34	-40	-26	Right temporal occipital fusiform cortex
		6.61	41	-77	-18	Right lateral occipital cortex inf

		6.54	-39	-37	-26	Left temporal fusiform cortex pos
		6.53	-36	-44	-16	Left temporal occipital fusiform cortex
		6.51	21	-47	-18	Right temporal occipital fusiform cortex
		6.45	29	-52	-16	Right temporal occipital fusiform cortex
	263	4.93	14	-97	27	Right occipital pole
		4.37	4	-92	32	Right occipital pole
		4.32	-1	-94	24	Left occipital pole
		4.22	31	-100	12	Right occipital pole
		4.03	-19	-100	27	Left occipital pole
		3.78	26	-97	22	Right occipital pole
		3.75	-19	-92	34	Left occipital pole
		3.6	-4	-87	40	Left cuneal cortex
	102	4.92	-11	18	50	Left superior frontal gyrus
		3.64	-6	20	42	Left paracingulate gyrus
		3.56	-1	23	50	Left superior frontal gyrus
	129	4.74	39	23	14	Right inferior frontal gyrus p tri
		3.71	59	30	10	Right inferior frontal gyrus p tri
		3.67	56	28	20	Right inferior frontal gyrus p tri
	88	4.25	29	-37	14	Right lateral ventricle
		3.98	21	-27	20	Right lateral ventricle
		3.48	26	-32	7	Right lateral ventricle
		3.4	34	-44	10	right inferior frontal occipital fas
ME>MEdn	80	4.54	-71	-20	2	Left superior temporal gyurs pos

		3.76	-61	-24	7	Left planum temporale
		3.23	-71	-30	7	Left superior temporal gyurs pos
	112	4.42	1	-14	27	Right cingulate gyrus ant
		3.73	9	-24	34	right cingulum cingulate
		3.69	6	-22	24	Right lateral ventricle
	163	4.22	-1	30	14	Left cingulate gyrus ant
		4.09	1	38	-6	Right cingulate gyrus ant
		3.81	-4	36	7	Left cingulate gyrus ant
		3.65	6	46	0	Right paracingulate gyrus
		3.36	16	48	2	forceps minor
	134	4.03	21	-42	-30	Right corticospinal tract
		4.01	16	-57	-48	Right cerebellum
		3.59	14	-47	-26	Right cerebellum
		3.4	19	-52	-38	Right corticospinal tract
2 × 2 factorial design – echo and (downsampled) band						
ANOVA	486	5.4	-44	-40	-18	Left temporal fusiform cortex pos
		4.85	-41	-47	-10	left inferior longitudinal fas
		4.79	-51	-44	-13	left superior longitudinal fas
		4.58	-46	-52	-20	Left inferior temporal gyrus temocc
		4.44	-34	-44	-20	Left temporal fusiform cortex pos
		4.16	-39	-24	-20	Left temporal fusiform cortex pos
		4.02	-36	-52	-18	Left temporal occipital fusiform cortex

	3.82	-59	-40	-10	Left middle temporal gyrus pos
281	5.39	44	-34	-23	Right inferior temporal gyrus post
	5.19	36	-37	-28	Right temporal fusiform cortex pos
	3.77	49	-47	-16	Right inferior temporal gyrus temocc
	3.74	39	-24	-18	Right temporal fusiform cortex pos
	3.7	34	-40	-18	Right temporal fusiform cortex pos
	3.28	36	-54	-20	Right temporal occipital fusiform cortex
	3.21	26	-42	-16	Right temporal occipital fusiform cortex
78	4.63	-19	36	-13	Left frontal orbital cortex
	3.5	-19	28	-18	Left frontal orbital cortex
	3.35	-9	23	-13	Left subcallosal cortex
57	4.39	31	36	-10	Right frontal pole
	3.22	21	38	-13	Right frontal pole
116	4.37	-36	18	30	Left middle frontal gyrus
	4.13	-36	13	17	Left frontal operculum cortex
	3.88	-34	20	20	Left inferior frontal gyrus p ope
	3.83	-46	26	24	Left middle frontal gyrus
64	4.19	41	-70	10	Right lateral occipital cortex inf
	4.01	44	-60	12	Right lateral occipital cortex inf

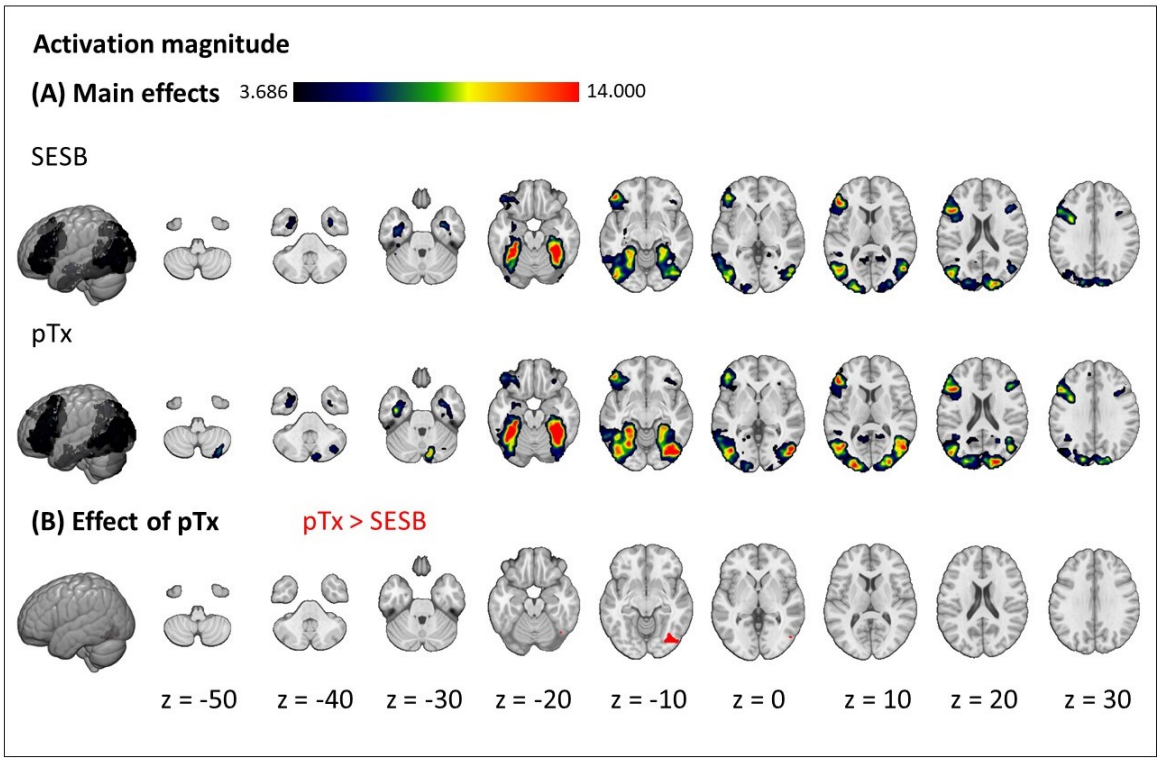

Figure D.1: Effects of parallel transmit on activation magnitude. Group maps of activation magnitude (contrast betas) for the contrast of interest (semantic>control): (A) main effect of the standard (SESB) sequence and the parallel transmit (pTx) sequence; (B) effect of parallel transmit (pTx>SESB, red; SESB>pTx, no significant clusters). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

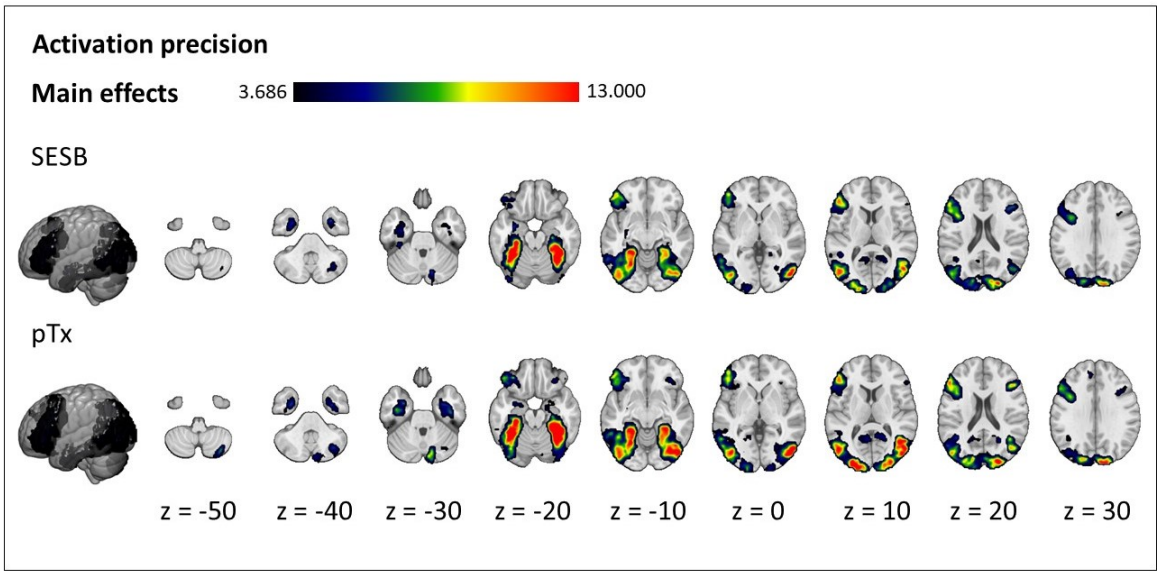

Figure D.2: Effects of parallel transmit on activation precision: main effects for group maps of activation precision (statistical t-values) for the contrast of interest (semantic>control) for the standard (SESB) sequence and the parallel transmit (pTx) sequence. Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

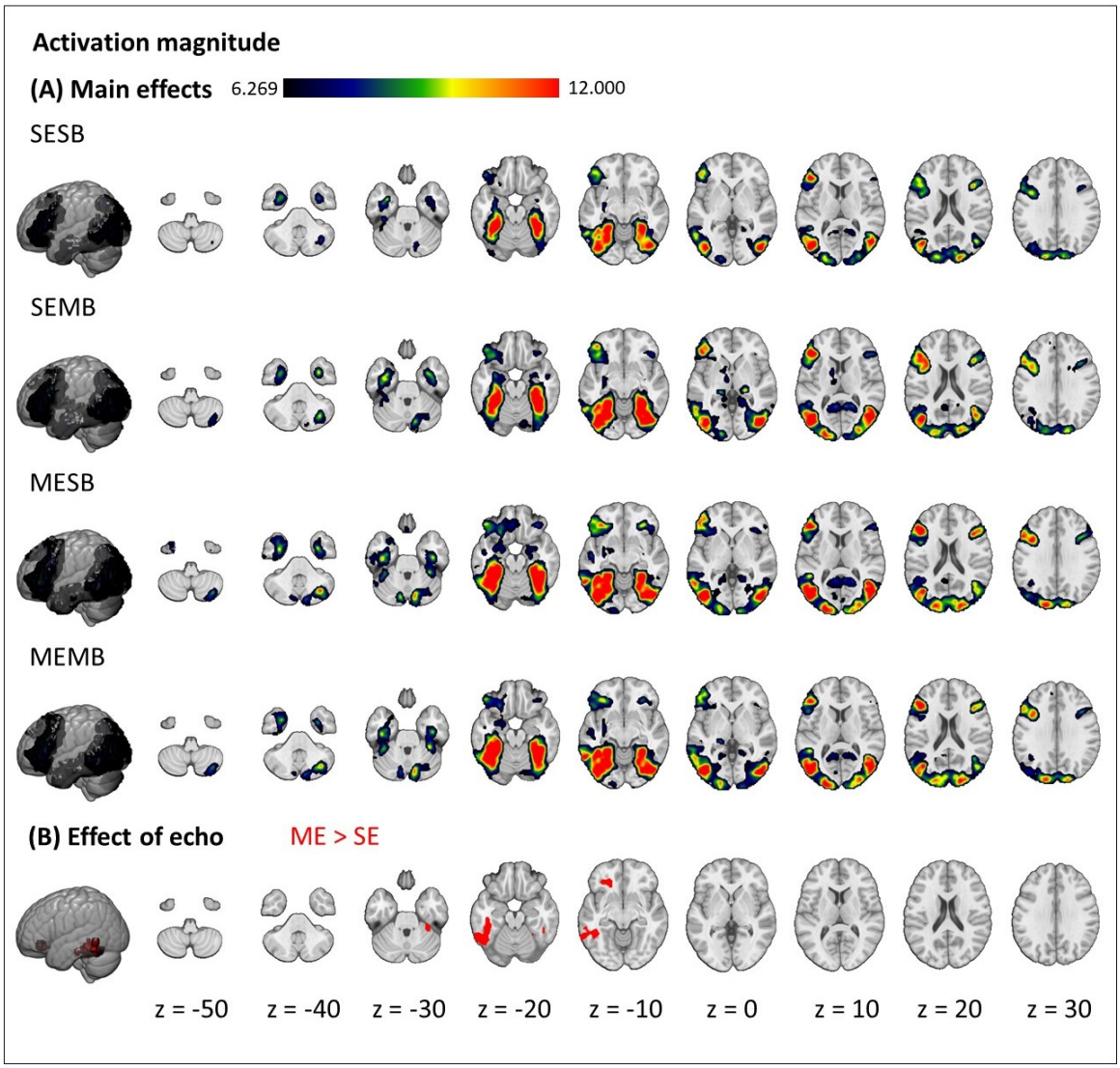

Figure D.3: Effects of multi-echo and multiband on activation magnitude. Group maps of activation magnitude (contrast betas) for the contrast of interest (semantic>control) in a 2 × 2 ANOVA manipulating echo and band: (A) main effect of each sequence (SESB = single-echo single band (standard), SEMB = single-echo multiband, MESB = multi-echo single band, MEMB = multi-echo multiband); (B) effect of echo (ME>SE, red; SE>ME, no significant clusters). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

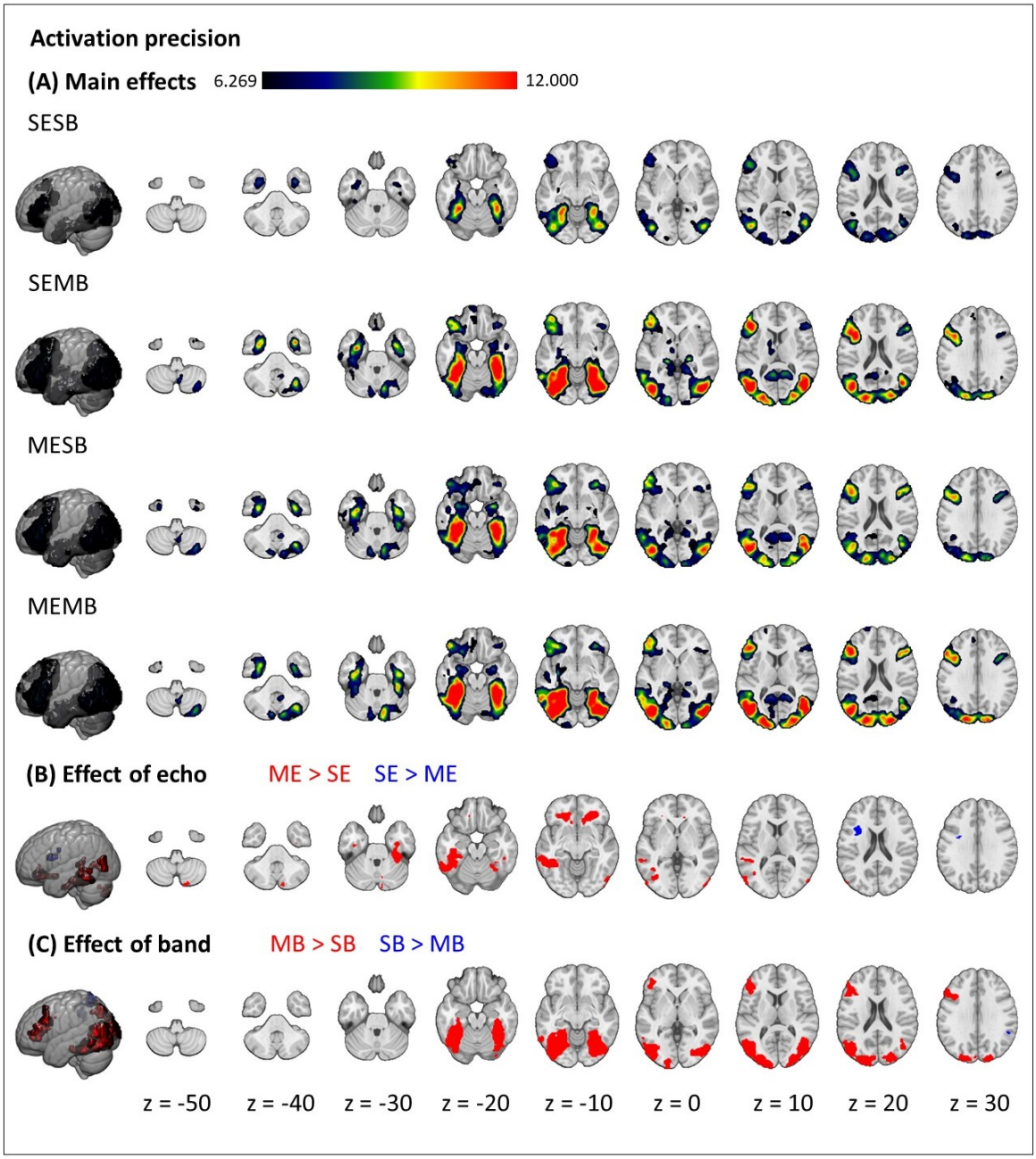

Figure D.4: Effects of multi-echo and multiband on activation precision. Group maps of activation precision (statistical *t*-values) for the contrast of interest (semantic>control) in a 2 × 2 ANOVA manipulating echo and band: (A) main effect of each sequence (SESB = single-echo single band (standard), SEMB = single-echo multiband, MESB = multi-echo single band, MEMB = multi-echo multiband); (B) effect of echo (ME>SE, red; SE>ME, blue); (C) effect of band (MB>SB, red; SB>MB, blue). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

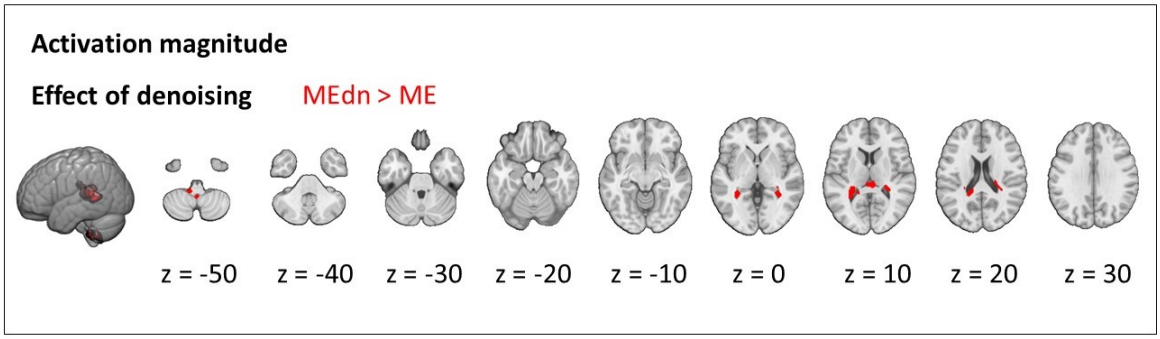

Figure D.5: Effect of ME-ICA denoising on activation magnitude: effect of ME-ICA denoising (MEdn>ME, red; ME>MEdn, no significant clusters). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

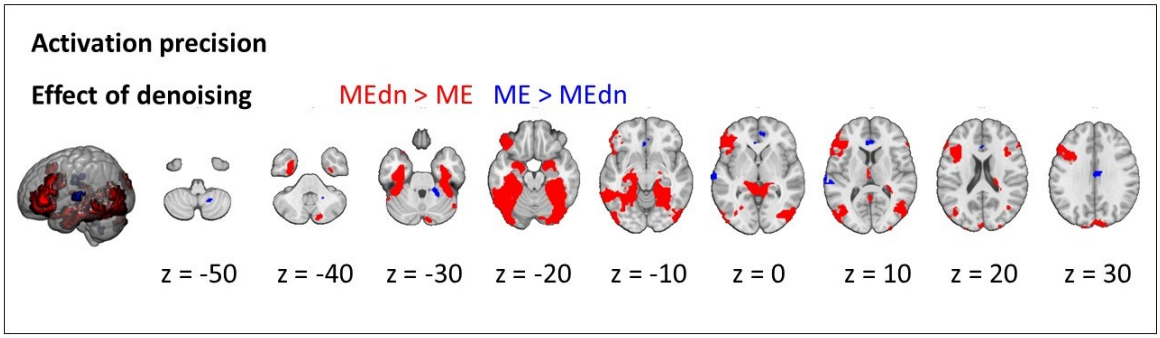

Figure D.6: Effect of ME-ICA denoising on activation precision: effect of ME-ICA denoising (MEdn>ME, red; ME>MEdn, blue). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$ and are overlaid on the MNI template to which all images were coregistered during preprocessing (MNI152NLin2009cAsym).

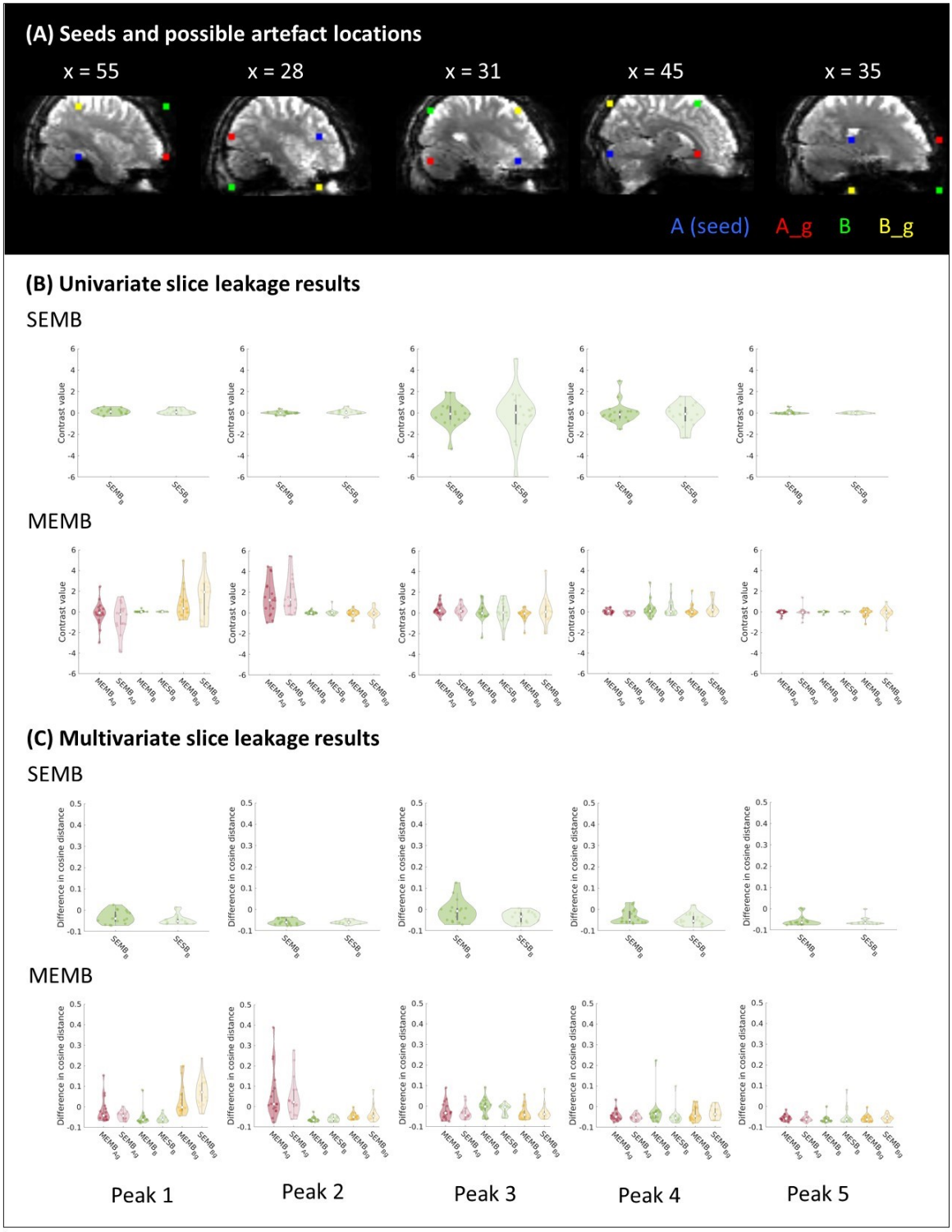

Figure D.7: Slice leakage analysis (all peaks). (A) Seed and possible artefact locations for a single participant. A (blue) is the seed location, B (green) is the possible artefact location based on phase shift, Ag (red) is the possible artefact location in the same slice as A based on GRAPPA, and Bg (yellow) is the possible artefact location in the same slice as B based on GRAPPA. All spheres have a radius of 4 voxels and are overlaid on one volume of multi-echo multiband (MEMB) data for a single participant in that participant's native space. Informed consent was obtained from the participant for this image to be published. (B) Mean activation magnitude (contrast betas) within each sphere for each MB sequence and the corresponding control

Figure D.7: sequence for each seed location. (C) Mean MVPA performance within each sphere for each MB sequence and the corresponding control sequence. The MVPA metric is the mean between-task cosine dissimilarity minus mean within-task cosine dissimilarity over all possible pairs of task blocks.

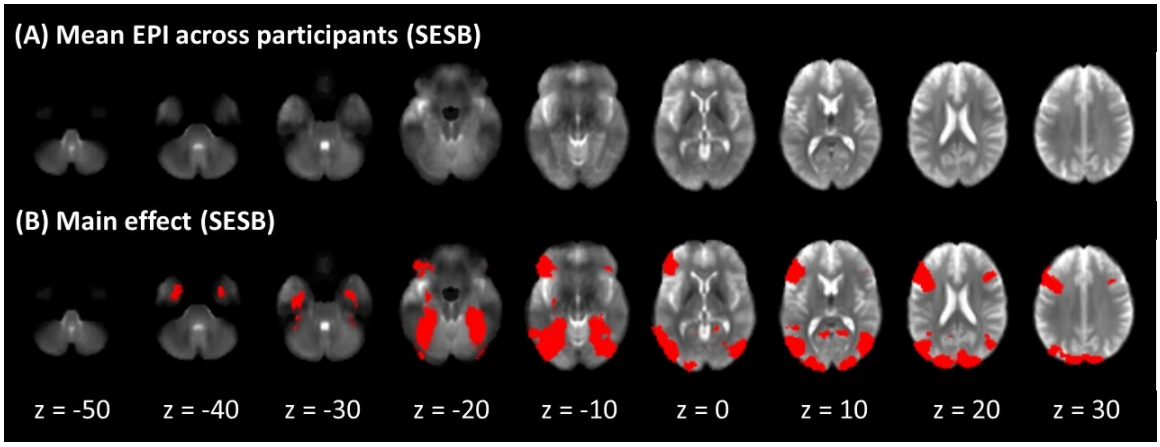

Figure D.8: Contrast despite signal dropout and distortions. (A) EPI data from the standard sequence (SESB = single-echo single band) coregistered to MNI space (template: MNI152NLin2009cAsym) and averaged over all volumes and all participants. (B) Main effect for the group map of activation magnitude (contrast betas) for the contrast of interest (semantic>control) for the SESB sequence overlaid on the mean EPI data in (A). Results are cluster-corrected at $p<0.05$ based on an uncorrected voxel threshold of $p<0.001$.

	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
Activation magnitude					
$SEMB_B$	0.3228	0.8153	0.3865	0.2161	0.6244
$MEMB_{Ag}$	0.0539	0.9614	0.3099	0.0068*	0.5378
$MEMB_B$	0.2499	0.5327	0.3019	0.7382	0.5362
$MEMB_{Bg}$	0.9647	0.3988	0.8770	0.9917	0.5296
MVPA					
$SEMB_B$	0.0596	0.3208	0.0107*	0.1079	0.3597
$MEMB_{Ag}$	0.0974	0.1860	0.2596	0.0959	0.3434
$MEMB_B$	0.2528	0.6570	0.0354*	0.0083*	0.8442
$MEMB_{Bg}$	0.9408	0.8805	0.5601	0.8335	0.5945

Table D.2: *p*-values for all slice leakage tests. A is the seed location, B is the possible artefact location based on phase shift, Ag is the possible artefact location in the same slice as A based on GRAPPA, and Bg is the possible artefact location in the same slice as B based on GRAPPA. * = $p<0.05$, ** = $p<0.05$ (Bonferroni-corrected for the number of peaks and possible artefact regions), SEMB = single-esho multiband, multiband, MEMB = multi-echo multiband.

Appendix E

Supplementary methods for Chapter 5: An overview of sparse decoding methods

E.1 Classification with regularised logistic regression

Logistic regression models the probability of a target condition (in this case, "the stimulus is inanimate") as a linear combination of features (in this case, each feature is a beta value for one voxel), each of which is assigned a coefficient (also called a weight). In the following, the features are denoted as $x_{1...m}$ and the coefficients are denoted as $\beta_{0...m}$. The weighted sum of features, z, is projected onto a sigmoid curve bound between zero and one – a value that is interpreted as the estimated probability of the target condition ($y = 1$):

$$P(y = 1|x) = \frac{e^z}{1 + e^z}\ , \text{ where } z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m$$

The optimal set of coefficients is the set that minimises the *loss function*. For logistic regression, the loss function is defined (where the total loss is obtained by summing over n stimuli, each of which has a condition label $y_i \in \{0, 1\}$ and a vector of features x_i):

$$\text{logistic loss} = f(y, x) = \sum_{i=1}^{n} -\Big(y_i \log P(y_i = 1|x_i\ ) + (1 - y_i) \log P(y_i = 0|x_i)\Big)$$

Once the optimal vector of coefficients has been determined, the coefficients can be multiplied by their corresponding features (in this case, beta values for each voxel) to obtain an estimated probability that the vector of features corresponds to the target condition (in this case, "the stimulus is inanimate"). The objective of minimising the logistic loss function with respect to the vector of coefficients can be expressed as:

$$\min_{\beta} f(y, x)$$

A logistic regression model becomes a classifier by applying a threshold. Conventionally, if the estimated probability is greater than 0.5, we say that the model has classified that vector of features as belonging to the target condition.

When there are many more features than stimuli, there are infinitely many vectors of coefficients for which the logistic loss is zero. A method for distinguishing between solutions must be selected. One strategy is to *regularise* the coefficients – rather than allowing the vector of coefficients to take any set of values, regularised logistic regression prefers some coefficient vectors over others even when the logistic loss is equal. This preference is implemented by penalising the classifier for adopting combinations of coefficients that the regularisation "dislikes". In general, the *objective function* (the function that, when minimised, gives the preferred solution) for regularised logistic regression is (where $h(\beta)$ is the regularisation penalty):

$$\min_{\beta} \left(f(y, x) + \lambda h(\beta) \right)$$

Here, λ is a positive scalar that controls the magnitude of the regularisation penalty and thus how the optimisation should prioritise the preferences of the regularisation penalty compared to the logistic loss (when λ is low, it is more important to prioritise minimising the logistic loss, and when λ is high, it is more important to prioritise finding a vector of coefficients that the model "likes"). The optimum value of λ differs for each dataset – it is a hyperparameter that must be learnt. In this study, the best value for λ was determined via ten-fold nested cross-validation. Data were divided into folds (in this case, ten folds, each containing ten of the 100 stimuli). One fold was defined as the "outer-loop holdout set" and was not used in the search. A second fold was defined as the "inner-loop holdout set". Models using different values of λ were trained on the remaining eight folds and were tested on the inner-loop holdout set. Then the folds were reassigned so that one of the other training folds became the inner-loop holdout set and the previous inner-loop holdout set was added to the training data. Once all combinations of inner-loop holdout set and training data had been explored, the value of λ that minimised the objective function across inner-loop holdout sets was identified. This value of λ, along with all the training data, was used to find a final vector of coefficients and the final model was assessed on the outer-loop holdout set. The procedure was repeated with the other 9 sets being the final holdout set (note that different values of λ can be selected for each fold).

E.1.1 Logistic regression with LASSO regularisation

In logistic regression with LASSO (L1) regularisation (Tibshirani, 1996), the regularisation penalty $h(\beta)$ is the sum of the absolute values of the coefficients:

$$h(\beta) = \sum_{j=1}^{m} |\beta_j|$$

The LASSO penalty increases as the sum of the absolute values of the coefficients increases. The more features that receive a coefficient of zero, the smaller $h(\beta)$ and thus the lower the objective function. This means that the LASSO penalty prefers solutions that are sparse (most coefficients are set to zero). LASSO also prefers solutions in which features are uncorrelated - if there are several units that predict category membership well and their activity is correlated, $h(\beta)$ will be minimised by placing a large coefficient on the feature that happens to be the best predictor and coefficients of zero on the others.

Note also that, in this study, logistic regression classifiers with LASSO regularisation were trained on data from each participant individually.

E.1.2 Logistic regression with SOSLASSO regularisation

Our knowledge of the brain tells us that, during representation, neural activity is unlikely to be as sparse and uncorrelated as the LASSO regularisation penalty would assume. Voxels are a property of data acquisition, not of the brain itself, and so the neural populations participating in representation may straddle voxel boundaries. This means that activity in groups of neighbouring voxels may be correlated yet important. Additionally, since a separate LASSO model is fitted for each participant, a completely different set of features may be selected in each brain, making group conclusions difficult. Since people share the same neuroanatomical macrostructure but differ in microstructure, a gentle yet well-justified assumption is that important voxels are in roughly (though not exactly) the same locations within a single brain and across multiple brains in the sample.

The sparse-overlapping-sets LASSO (SOSLASSO) regularisation penalty (Cox & Rogers, 2021; Rao et al., 2013, 2016) incorporates this assumption. Before training, features are assigned to overlapping sets. In this case, each voxel (in native space) was assigned MNI coordinates based on the location that the voxel *would* have *were* the images to be transformed into MNI space (actually normalising the data would cause spatial blurring and limit detection of

"heterogeneous" representations in which neural populations encode the same information via different responses; Frisby et al., Chapter 2). Based on these coordinates, voxels were assigned to region-of-interest-like spheres that were of a fixed radius and overlapped by a fixed amount. The sets were defined to correspond across individuals, so voxels that were in the same location in different participants were assigned to the same set.

The SOSLASSO penalty has two components. The first is the LASSO penalty, $h(\beta) = \sum_{j=1}^{m} |\beta_j|$. The second is called the grouping penalty:

$$\sqrt{\sum_{j=1}^{m} \beta^2}$$

An increase in the magnitude of individual coefficients causes a steeper increase in the grouping penalty than increasing the sum of the coefficients does. This means that the grouping penalty prefers solutions in which correlated features are each assigned a small coefficient, rather than a large coefficient being assigned to one and zero to the others.

The LASSO penalty and the grouping penalty are combined into the SOSLASSO penalty (α is a learnt hyperparameter that controls the relative importance of each penalty; note also that the SOSLASSO penalty is calculated for each set individually and summed over all sets S):

$$h(S, \beta) = \sum_{S} \left((1 - \alpha) \sum_{j=1}^{m} |\beta_j| + \alpha \sqrt{\sum_{j=1}^{m} \beta^2} \right)$$

In this case, the optimum value of α (like λ) differs for each dataset and must be learnt via cross-validation. Both the LASSO penalty and the grouping penalty increase as the number of nonzero coefficients increase, and so SOSLASSO produces solutions that are sparse. However, because of the grouping penalty, the magnitude of the SOSLASSO penalty is also determined by the set assignment of the features that receive nonzero coefficients. Features that are in the same set will make a smaller contribution to the SOSLASSO penalty than the same number of features (with the same coefficients) split into different sets. By preferring solutions in which voxels are in the same sets, SOSLASSO implements the assumption that informative features are found in similar locations within and across participants.

Although data from all participants is used in the inner loop of cross-validation, the outer loop is run for all participants individually. This means that each participant receives their own set of final coefficients, which is constrained by α and λ but need not be identical to the set

that other participants receive. This is how SOSLASSO implements the assumption that informative voxels will be roughly, but need not be exactly, in the same location within and across participants.

E.2 Representational similarity learning (RSL)

Semantic representations express conceptual similarity structure – an overall similarity that goes beyond individual sensory modalities (for example, knowledge that an orange and a banana are more similar to each other than either of them is to a basketball, even though the orange resembles the basketball and the banana and the basketball both begin with "ba"; Devereux et al., 2018). This fact can guide our search for semantic representation in the brain: a region of the brain that represents semantic information should encode semantically similar stimuli using similar activity patterns (Edelman, 1998). RSL is one method of probing the similarity structure in neural activity (C. R. Cox et al., 2024; Oswal et al., 2016).

The first stage in RSL is the creation of a target representational similarity matrix (RSM). The target RSM is an $n \times n$ matrix (where n is the number of stimuli – in this case, $n = 100$). Each cell expresses the degree of similarity that would be predicted between that pair of stimuli under a hypothesis - in this case, the target RSM expresses the semantic similarity between each pair. There are many ways of estimating semantic similarity (Frisby et al., Chapter 2) - in this study we modelled semantic similarity using feature verification norms (Dilkina & Lambon Ralph, 2012). Participants were asked to list features that were true of a given concept; a second set of participants was then given the list of concepts and the entire list of features (generated for any concept) and filled in a *concept-by-feature matrix* (in which rows are concepts and columns are features) to indicate which feature belonged to each concept. The concept-by-feature matrix was then normalised – the average value for each column (feature) was subtracted from the column, which emphasised the values of features that differed between concepts. Each row of the matrix then formed a vector – one for each concept. Semantic similarity between two concepts was operationalised as the cosine distance between their pair of vectors. In the second stage of RSL, singular value decomposition (SVD) is used to define a multidimensional target space with r dimensions (in this case, $r = 3$) based on the target RSM ($\boldsymbol{S}$). Stimuli that are more similar to one another are closer together in this space. SVD works by decomposing a matrix into orthogonal components – the first component explains the most variance, the second explains

the second most variance, and so on. SVD yields three products:

1. An $n \times r$ matrix ($\boldsymbol{U}$) of components, also known as singular vectors.
2. The transpose of $\boldsymbol{U}$ ($\boldsymbol{U^T}$).
3. An $r \times r$ matrix ($\boldsymbol{D}$) of singular values.

To summarise:

$$\boldsymbol{S} \cong \boldsymbol{U}\boldsymbol{D}\,\boldsymbol{U^T}$$

The matrix $\boldsymbol{U}$ gives the target coordinates of each stimulus in a multidimensional space. An alternative set of coordinates $\boldsymbol{C}$, in which each column of $\boldsymbol{U}$ is weighted by the square root of the corresponding singular values, can be calculated:

$$\boldsymbol{C} = \boldsymbol{U}\sqrt{\boldsymbol{D}}$$

Either $\boldsymbol{U}$ or $\boldsymbol{C}$ can be used as target coordinates for RSL – following the approach of Cox et al. (2024), we used $\boldsymbol{C}$. Note that $\boldsymbol{S}$ can be approximated by multiplying $\boldsymbol{C}$ by its transpose $\boldsymbol{C^T}$:

$$\boldsymbol{S} \cong \boldsymbol{C}\,\boldsymbol{C^T}$$

This operation rarely calculates $\boldsymbol{S}$ exactly because r dimensions are rarely enough to explain 100 % of the variance in $\boldsymbol{S}$. In this study, the three dimensions accounted for a total of 89.5 % of the variance (81.1 %, 4.4 %, and 4.0 %, respectively).

In the third stage of RSL, models are trained on an $n \times m$ feature matrix $\boldsymbol{X}$, where m is the number of features (in this case, the number of voxels in each participant's data). The model learns to predict coordinates of each stimulus on multiple dimensions, so each stimulus receives multiple coefficients. These are stored in an $m \times r$ matrix called a decoding matrix ($\boldsymbol{\beta}$). One training strategy is to take each dimension one by one – first to find the vector of coefficients that minimises the sum of squared errors between predicted and target values on dimension 1, then to find the vector of coefficients that minimises the sum of squared errors between predicted and target values on dimension 2, and so on – and fill in each column of $\boldsymbol{\beta}$ individually.

However, it is possible to obtain the same value of $\boldsymbol{\beta}$ with another objective function - the Frobenius norm of the differences between $\boldsymbol{C}$ and $\boldsymbol{\beta}$:

$$\min_{\beta} ||\boldsymbol{C} - \boldsymbol{X}\boldsymbol{\beta}||_F$$

This defines $\boldsymbol{\beta}$ for all dimensions at once.

As for logistic regression, there are many vectors of coefficients for which the Frobenius norm is zero. A regularisation penalty, $H(\beta)$, can be implemented to distinguish between solutions. The objective function becomes:

$$\min_{\beta} (||\boldsymbol{C} - \boldsymbol{X}\boldsymbol{\beta}||_F + \alpha H(\beta))$$

Again, the optimum value of α differs for each dataset and must be learnt, in this case by ten-fold nested cross-validation.

E.2.1 RSL with LASSO regularisation

In RSL, the LASSO regularisation penalty is computed for each dimension independently – for each dimension, it is the sum of the absolute coefficients in the corresponding column of $\boldsymbol{\beta}$. Again, it produces solutions in which features are sparse and uncorrelated.

E.2.2 RSL with group-ordered-weighted LASSO (grOWL) regularisation

Stimuli that are more semantically similar should be represented by more similar patterns of neural activity. However, it is highly unlikely that those patterns vary between stimuli along exactly r dimensions and that those dimensions are exactly the ones produced by SVD of a particular target RSM – that is to say, the target similarity space is unlikely to be "axis-aligned" with the real similarity space encoded by the brain. Therefore, the activity of a neural population that encodes just one dimension of the brain's real similarity space is likely to correlate with multiple dimensions of the target space.

The group-ordered-weighted LASSO (grOWL; C. R. Cox et al., 2024; Oswal et al., 2016) incorporates this and two other assumptions:

1. *Spanning assumption*: features that predict one dimension of the target similarity space should predict all other dimensions of the target similarity space.

2. *Sparsity assumption*: only a small subset of the total number of features carries important information.

3. *Redundancy assumption*: features that do carry important information are correlated in their activity patterns over stimuli.

grOWL works by creating a vector (w) of non-negative, non-increasing values – for example, a vector of positive values where the first element is the largest and each successive element is smaller than or equal to the previous one. w is of length m and therefore has the same number of elements as there are rows in $\boldsymbol{\beta}$. A correspondence is learnt between elements of w and rows of $\boldsymbol{\beta}$. The total regularisation penalty is calculated by multiplying the row of $\boldsymbol{\beta}$ that contains the biggest coefficients by the biggest value in w, multiplying the row of $\boldsymbol{\beta}$ that contains the next biggest coefficients by next biggest value in w, and so on and then summing the results. Since there are multiple elements in each row, coefficient size is defined by the Euclidean norm or 2-norm, which is the length of the vector that those coefficients subtend in r-dimensional space. Alternatively, the Euclidean norm can be thought of as the square root of the sum of squares of the coefficients in that row (similar to the grouping penalty in SOSLASSO).

The grOWL penalty is formalised as the sum of the Euclidean norms of the rows, each multiplied by its value in w ($\boldsymbol{\beta}_{[i]\cdot}$ is the row of $\boldsymbol{\beta}$ with the i-th largest 2-norm and w_i is the value in w applied to $\boldsymbol{\beta}_{[i]\cdot}$):

$$H(\beta) = \sum_{i=1}^{m} w_i ||\boldsymbol{\beta}_{[i]\cdot}||_2$$

The largest value of w_i is assigned to the row with the largest coefficient size. In order to make the large value of w_i "worth" the large penalty, it should be applied to the feature (in this case, voxel) that is the most informative (and so will keep $||\boldsymbol{C} - \boldsymbol{X}\boldsymbol{\beta}||_F$ low, keeping the objective function low overall). Note that grOWL prefers solutions in which all r coefficients in a row are similar in size to solutions in which one of the coefficients is very large and the others are zero. This is because, within each row, an increase in the magnitude of individual coefficients causes a steeper increase in $||\boldsymbol{\beta}_{[i]\cdot}||_2$ than increasing the sum of the coefficients does. Therefore, features that predict multiple dimensions contribute less to the total regularisation penalty than features that predict one dimension excellently and others poorly. This is how grOWL implements the spanning assumption.

Rows that contain only small coefficients receive small values of w_i; however, if that row reduces $||\boldsymbol{C} - \boldsymbol{X}\boldsymbol{\beta}||_F$ very little (or not at all), even a small value of w_i will increase the objective function. The only way to avoid an increase is to set that row of $\boldsymbol{\beta}$ to zeros. This is how grOWL implements the sparsity assumption.

If there are multiple correlated rows, the value of w_i that is allocated to the first of those rows is allocated to all the others as well. This means that features that are correlated in their activity (and therefore receive similar coefficients) are likely to receive an identical value of w_i, so they will either all receive nonzero coefficients or all receive coefficients of zero. This is how grOWL implements the redundancy assumption.

Appendix F

Supplementary results for Chapter 5

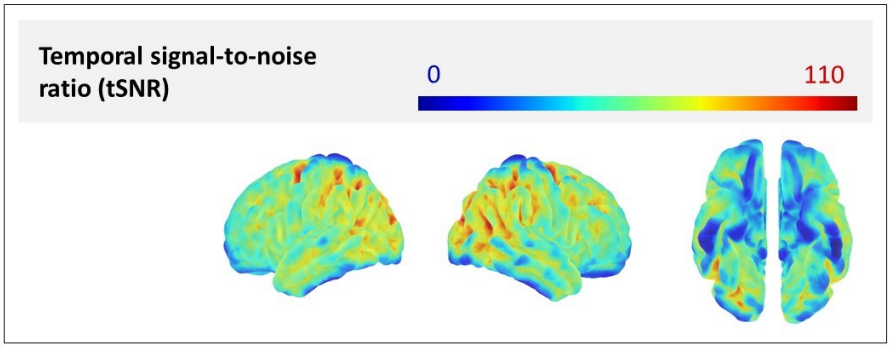

Figure F.1: Mean temporal signal-to-noise ratio. Values are shown projected to the surface (fsaverage template).

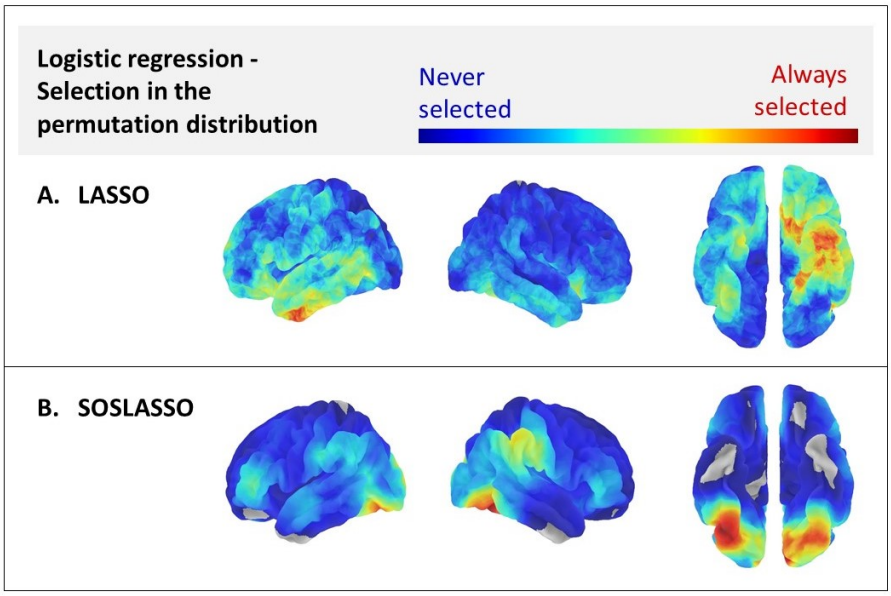

Figure F.2: Selection in the permutation distribution for logistic regression classifiers. (A) Proportion of participants in which each vertex is assigned a non-zero coefficient by logistic regression classifiers trained with LASSO regularisation on *permuted* contrast beta values to discriminate animate from inanimate stimuli. Warm colours indicate that voxels in the region are selected more frequently. (B) Proportion of participants in which each vertex is assigned a non-zero coefficient by logistic regression classifiers trained with SOSLASSO regularisation on *permuted* contrast beta values to discriminate animate from inanimate stimuli. The colour scaling is the same as in (A).

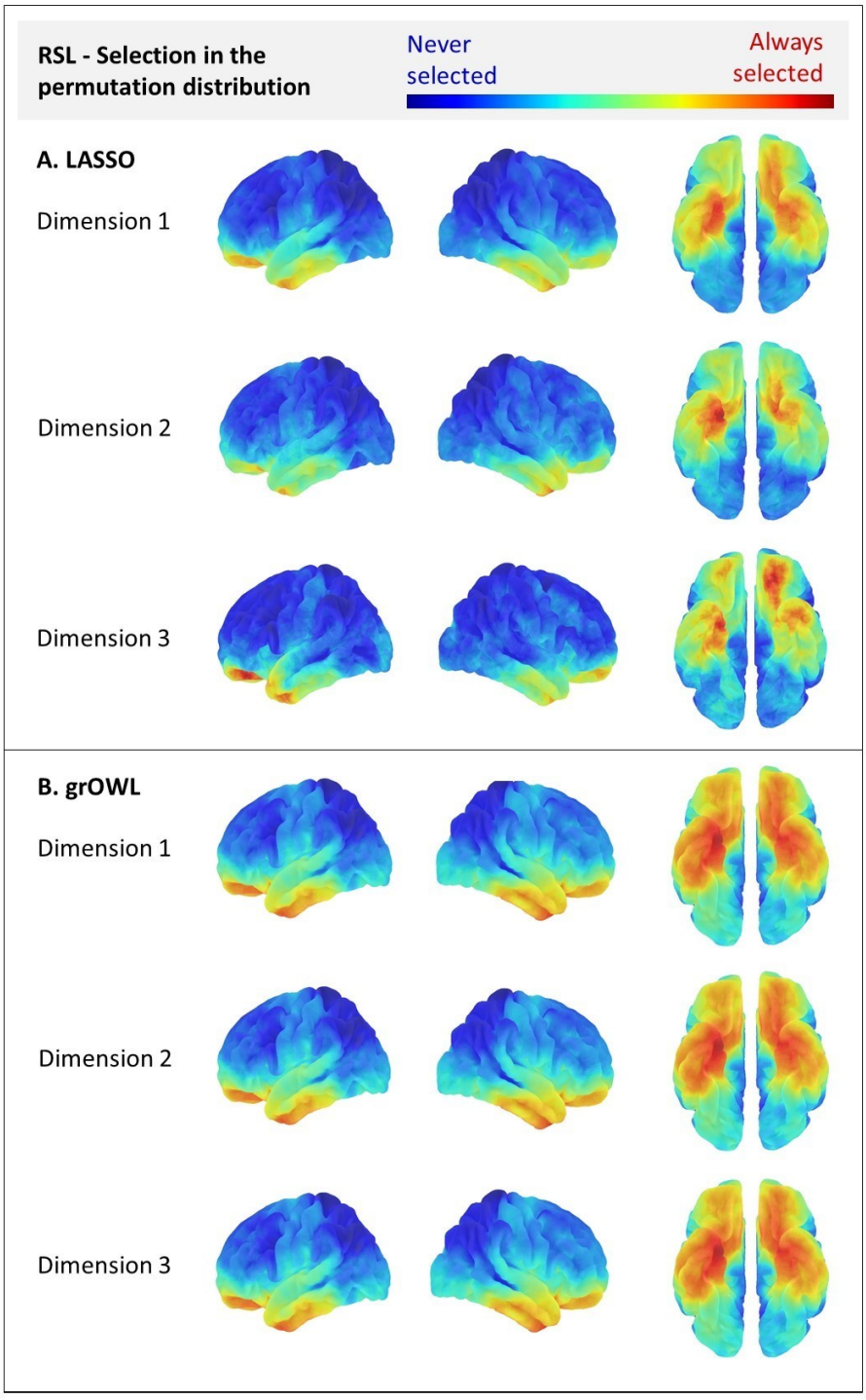

Figure F.3: Selection in the permutation distribution for RSL models. (A) Proportion of participants in which each vertex is assigned a non-zero coefficient by RSL models trained with LASSO regularisation on *permuted* contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions. Warm colours indicate that voxels in the region are selected more frequently. (B) Proportion of participants in which each vertex is assigned a non-zero coefficient by RSL models trained with grOWL regularisation on *permuted* contrast beta values to predict the coordinates of held-out stimuli on three target semantic dimensions. The colour scaling is the same as in (A).