UNIVERSITY OF CALIFORNIA,
IRVINE

Parcellating speech networks using connectivity and multivariate meta-analysis

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Sciences with a Concentration in Cognitive Neuroscience

by

Alex Teghipco

Dissertation Committee:
Professor Gregory Hickok, Chair
Associate Professor Alyssa Brewer
Professor Mark Steyvers

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would first like to recognize and thank the members of my committee, Alyssa Brewer and Mark Steyvers, for their guidance and support. I would especially like to thank the chair of my committee, Greg Hickok, for the innumerable ways in which his mentorship, advice and help–in the many forms that it has taken–has shaped both me and the research presented here.

I would also like to extend my gratitude to the many friends and colleagues who have helped me along the way, including all the wonderful members of the auditory and language neuroscience lab who have had the phenomenal patience to listen to too many of my presentations and help make them better every time. A very special thanks to Brandon Hackney and Sajjad Torabian for their dedication and commitment to science, it is no exaggeration that without their help, COVID-19 would have swallowed up a third of this dissertation.

Above all, I would like to thank Taylor McCabe, without whom none of this would have been remotely possible.

CURRICULUM VITA

Alexander Teghipco

EDUCATION

Doctor of Philosophy in Cognitive Neuroscience **2021**
University of California, Irvine *Irvine, CA*

Bachelor of Arts in Brain and Cognitive Science, and Philosophy **2014**
University of Rochester *Rochester, NY*

RESEARCH EXPERIENCE

Graduate Research Assistant/Fellow **2018-2021**
University of California, Irvine *Irvine, CA*

TEACHING EXPERIENCE

Teaching Assistant **2016-2018; Summer 2021**
University of California, Irvine *Irvine, CA*

REFEREED JOURNAL PUBLICATIONS

Chernoff, B. L., Teghipco, A., Garcea, F. E., Belkhir, R., Sims, M. H., Paul, D. A., ... Mahon, B. Z. (2020). Reorganized language network connectivity after left arcuate fasciculus resection: A case study. Cortex, 123, 173-184.

Chernoff, B. L., Teghipco, A., Garcea, F. E., Sims, M. H., Paul, D. A., Tivarus, M. E., ... Mahon, B. Z. (2018). A role for the frontal aslant tract in speech planning: A neurosurgical case study. (5), 752-769. Journal of Cognitive Neuroscience, 30.

Garcea, F. E., Chernoff, B. L., Diamond, B., Lewis, W., Sims, M. H., Tomlinson, S. B., Teghipco, A.,... B.Z., Mahon. (2017). Direct Electrical Stimulation in the Human Brain Disrupts Melody Processing. Current Biology, 27(17), 2684-2691.

Teghipco, A., Hussain, A., Tivarus, M. E. (2016). Disrupted functional connectivity affects resting state based language lateralization. NeuroImage: Clinical, 12, 910-927.

Stasenko, A., Bonn, C., Teghipco, A., Garcea, F. E., Sweet, C., Dombovy, M., ... Mahon, B. Z. (2015). A causal test of the motor theory of speech perception: A case of impaired speech production and spared speech perception. Cognitive neuropsychology, 32(2), 38-57.

REFEREED CONFERENCE PUBLICATIONS

Teghipco, A. Hickok, G. Data-driven meta-analysis of the structural-functional parcel- lation of temporal and temporoparietal cortex. Poster presentation at the Society for the Neurobiology of Language,Helsinki, Finland. August 2019. Abstract merit award winner.

Teghipco, A. Hickok, G. Evaluating the functional neuroanatomy of language represented across 11,406 neuroimaging studies: a multivariate meta-analysis. Poster presentation at the Society for the Neurobiology of Language, Montreal, Canada. August 2018.

Teghipco, A., Kundu, P., Buchsbaum, B., Bandettini, P. Hickok, G. Functional parcel- lation of the planum temporale. Poster presentation at the Cognitive Neuroscience Society Annual Meeting, San Francisco, CA. March 2017.

Diamond, B., Garcea, F.E, Chernoff, B.L., Belkhir, R., Teghipco, A., Smith, S., Navarrete, E., Pilcher, W., Mahon, B.Z. Connectivity of the language system revealed by direct current stimulation during awake neurosurgery. Poster presentation at the Cognitive Neuroscience Society Annual Meeting, San Francisco, CA. March 2017.

Diamond, B., Garcea, F.E., Chernoff, B.L., Teghipco, A., Smith, S.O., Pilcher, W.H., Mahon, B.Z. Language processing and cortical stimulation: A neurosurgical case study. Poster presentation at the Cognitive Neuroscience Society Annual Meeting, San Francisco, CA. March 2017.

Chernoff, B., Teghipco, A., Garcea, F.E., Smith, S.O., Pilcher, W.H., Mahon, B.Z. A cognitive impairment for sentence planning after focal damage to the frontal aslant tract. Poster presentation at the Cognitive Neuroscience Society Annual Meeting, San Francisco, CA. March 2017.

Garcea, F.E., Teghipco, A., Tivarus, M.E., Smith, S.O., Pilcher, W.H., and Mahon, B.Z. The role of the right hemisphere in recovery from conduction aphasia: A neurosurgical case study. Poster presentation at the Gordon Research Conference on the Neurobiology of Cognition. July 2016.

Chernoff, B., Garcea, F.E., Teghipco, A., Smith, S.O., Pilcher, W.H., and Mahon, B.Z. A temporal-parietal pathway for accessing action knowledge from visual input. Poster presen- tation at the Gordon Research Conference on the Neurobiology of Cognition. July 2016.

Teghipco, A., Garcea, F.E., Tivarus, M., Smith, S., Pilcher, W.H., and Mahon, B.Z. Recovery from conduction aphasia depends on contributions from the right hemisphere: A case study. Poster presentation at The Cognitive Neuroscience Society Annual Meeting, New York, NY. April 2016. 2016 CNS People's Choice Poster Award Winner.

Tivarus, M.E., Teghipco, A., Utz, M., Cole, D. Atlas based seed analysis of resting state fMRI for pre-surgical brain mapping. International Society for Magnetic Resonance in Medicine, Toronto, Ontario. June 2015.

Stasenko, A., Teghipco, A., Bonn, C., Garcea, F.E., and Mahon, B.Z. The limits of speech perception without a motor system. Poster presentation at The Cognitive Neuroscience Society Annual Meeting, Boston, MA. April 2014.

Teghipco, A., Navarette E., Mahon B.Z. Neural Correlates of Verbal Fluency. Poster presentation at The Cognitive Neuroscience Society Annual Meeting, Boston, MA. April 2014.

SOFTWARE

BrainSurfer https://github.com/alexteghipco/brainSurfer GUI-based MATLAB package that visualizes surface space data, projects volumes onto surface templates, and converts between surface formats.

PhysAlign https://github.com/alexteghipco/physalign Matlab package for aligning physiological data collected in MRI to dicoms.

SemanticSpace a PERL module for calculating semantic similarity between words using the WordNet database

ABSTRACT OF THE DISSERTATION

Parcellating speech networks using connectivity and multivariate meta-analysis

by

Alex Teghipco

Doctor of Philosophy in Cognitive Sciences with a Concentration in Cognitive Neuroscience

University of California, Irvine, 2020

Professor Gregory Hickok, Chair

A densely interconnected patchwork of cortical areas tiles the brain, balancing the integration and segregation of functions. The eminent strategy for charting the functions of this landscape has involved pinning tightly controlled behavioral tasks onto large and often isolated patches of cortex in piecemeal. Recent developments have made it possible to study the organization of this rich functional landscape from the bottom-up, by mapping the connectivity networks that underpin brain areas situated within larger functional regions. At the same time, the availability of large databases that aggregate published functional neuroimaging work make it possible to infer the functions of such connectivity-defined areas from a much broader behavioral space than can be accessed by any individual study. In the first study of the present work, this bottom-up approach was

leveraged to elucidate functional organization around the planum temporale (PT), an anatomical region that has been shown to be critical to speech production, but which has also been associated with a variety of other functions in functional neuroimaging studies. Functional connectivity estimated from multiecho resting state data (N =137) was used to map areas within the posterior perisylvian zone that functional neuroimaging studies have associated with the PT. A deliberate parcellation of this zone was produced by forming cluster ensembles from complementary unsupervised learning algorithms that produced reliable parcellations (12 clustering approaches were tested for their reliability). The resulting parcellation was well-aligned with observer-independent architectonic work, and even outperformed several popular atlases when used to classify external data. Registering the networks and areas within this parcellation to published functional neuroimaging work revealed that the functional zone associated with the PT was comprised of multiple speech areas, including an area involved in rehearsal and an area involved in production. Further, multiple areas were associated with receptive speech, auditory, and sensorimotor processing. At a broad level, a second study tested the extent to which there *can* be agreement between areas defined from the bottom-up based on connectivity data, and functional areas embedded in published neuroimaging work. That is, a parcellation of temporal and inferior parietal cortex based on structural connectivity (N = 70) was compared to a parcellation resulting from the same clustering approach adapted to brain-behavior associations (i.e., meta-analyses of behaviors). This functional parcellation carved out areas with similar behavioral profiles and was implemented in the service of localizing areas involved in speech perception. Prior meta-analytic studies with the same goal have used sample sizes now known to be insufficient, produced incompatible results, and failed

to consider whether speech responsive areas are involved in other functions. Clustering was performed over lower-dimensional functional networks extracted from brain-behavior associations. Auditory and speech processes were robustly distinguished irrespective of how many functional networks were defined. A more optimal set of functional networks showed evidence of a posteriorly directed speech processing hierarchy that implicated an area in ventral superior temporal gyrus, posterior to auditory cortex, in speech sound analysis. The functional parcellation isolated areas along this speech hierarchy and captured the extent to which they were shared across functional networks. Overall, there was significant spatial correspondence between the functional and structural parcellations, and the parcel associated with speech sound processing showed the best fit to structural connectivity data. Structural data indicated that this parcel could be distinguished based on its strong connectivity along the middle longitudinal fasciculus. A third study investigated a more targeted hypothesis about the role of pMTG in error correction. The results of this study helped better characterize the function of one of two areas from the functional parcellation presented in the second study that were both associated with the pMTG, but difficult to functionally distinguish in a precise way. Previous work has demonstrated suggestive evidence of internal error detection and correction in left posterior middle temporal gyrus (pMTG) on the basis that this area tended toward showing a stronger response when *potential* speech errors (i.e., not realized) are biased towards nonwords compared to words. Building on this prior work, an experiment was designed introducing novel tongue twister stimuli that attempted to further tax internal error correction and detection mechanisms by biasing potential speech errors towards taboo words. Although the pMTG demonstrated significantly greater response to taboo words than neutral words

in a large sample of participants (N=40), other areas also showed an effect. Behavioral associations in the wider neuroimaging literature were used to better understand the functional significance of activity outside the pMTG, demonstrating that only activity within this area was likely to signal word-level processing, and therefore explain the internal error correction mechanism investigated. Taken together, these studies present a more nuanced description of the functional and structural networks underpinning the speech system and point to more specific functions that these networks may support.

Chapter 1

Introduction

A central aim of cognitive neuroscience is to understand how the brain enables mental processes that can be studied behaviorally. Functional neuroimaging has been paramount to this research program because it offers a powerful tool for linking mental constructs with brain structures. Indeed, the explosion of task-based functional magnetic resonance imaging (fMRI) studies over the last several decades has generated a wealth of brain-behavior relationships, affording an unparalleled window into functional organization in the brain (Derrfuss & Mar, 2009; Rosen & Savoy, 2012). At the same time, these efforts have underscored the complexity of establishing links between mental processes and brain regions, and the difficulty of building up satisfying accounts that can explain them. Not only are there numerous inferential challenges for individual task-fMRI studies to navigate when attempting to generate such links (e.g., Aue et al., 2009; Poldrack, 2006), but it's clear that many brain regions respond to a wider range of behavioral manipulations than anticipated (e.g., Costafreda et al., 2008; Hein & Knight, 2008). This has raised questions about the degree to which any region of the brain can be functionally specialized (e.g., Anderson, 2015; Hutto et al., 2017). This dissertation focuses on two complementary developments that have made it possible to begin tackling some of these challenges and extracting more detailed information about functional organization within brain regions.

Advances in imaging brain connectivity have made it possible to dissect larger functional regions into smaller brain areas that are embedded within different networks. At the same time, collecting information from tens of thousands of published functional neuroimaging studies has made it possible to probe for more nuanced differences in behavioral associations across larger regions, define networks that are shared amongst groups of different behaviors, and inform more targeted task-fMRI investigations by contextualizing their findings within a broader behavioral space. The current work adopts these developments to parcellate functional, structural, and task networks associated with speech processing.

The rapid growth of functional neuroimaging studies has galvanized efforts to aggregate published results. These efforts have cumulated in large databases that can be used to probe a wide range of behavioral associations across the entire brain (Laird et al., 2005; 2009; Yarkoni et al., 2011). The scale of these databases helps to address some of the inferential challenges in task-fMRI that can affect the localization of functions by smoothing out the idiosyncrasies of individual experiments. In addition, this scale permits a more complete characterization of the behavioral context in which a brain region responds, which can bootstrap hypotheses about the latent function it might support (e.g., Genon et al., 2018; Price & Friston, 2005; Van Essen, 2009). Access to this rich diversity of behavioral associations across the entire brain also enables meta-analyses that are more data-driven and less domain-specific, defining groups of behaviors that are likely to produce activity within the same networks of areas. The topography of these networks may provide a finer understanding of functional organization within larger regions. At the same time, databases that aggregate published results have made it possible for individual task-fMRI studies to

connect their results to a much wider literature, permitting better inferences about the functions that may be carried within different regions of activity. For example, more careful meta-analysis can describe the cognitive processes that may have been engaged as a function of observing activity in an unexpected area.

In principle, these large databases aggregating neuroimaging findings provide a radical alternative to the conventional strategy for mapping brain function, where some behavior is preselected, and the goal of the research is to infer which brain regions are associated with it. Instead, it is now possible to make inferences in the opposite direction—starting with brain regions and then working towards inferring their functions. This development aligns with a growing interest in bottom-up approaches for capturing representations of brain organization. That is, approaches that define brain areas as a function of change in some neurobiological property. Traditionally, these bottom-up approaches have been operationalized in the context of invasive methods. However, advances in imaging structural and intrinsic functional connectivity in vivo have demonstrated some sensitivity to cortical areas defined with invasive methods (e.g., Gao et al., 2018; Glasser et al., 2016). Because the function of any brain area is deeply constrained by its inputs and outputs (e.g., Amunts & Zilles, 2015; Passingham et al., 2002), evaluating subtle differences in connectivity patterns across the brain can reveal many brain areas presumed to carry out different functions based on their participation in different networks. Unlike task-fMRI studies that require piecemeal comparisons between behaviors to map typically larger brain regions, bottom-up methods can delineate many smaller brain areas at once. By doing this in a data-driven manner, they bypass some of the inferential challenges with task-fMRI that carry over into meta-analytic approaches. Moreover,

bottom-up methods support insights into functional organization by mapping the wider networks within which areas are embedded.

A major disadvantage to the bottom-up approach for mapping brain organization is that it cannot say much about the specific functions that areas carry out. For this reason, the availability of databases that aggregate functional neuroimaging findings has been incredibly timely and presents an opportunity to search for the correspondence between brain structure and function in novel ways. For example, bottom-up approaches can be especially helpful for clarifying organization within brain regions that have become associated with many different functions, or which have proven difficult to parcellate with task-fMRI (e.g., by providing some support for the number of functions a given brain region might be able to carry out, mapping the networks that might correspond to the different functions a region has been implicated in, and producing a model of the way these functions may be spatially configured in the brain). Aggregated functional neuroimaging data can then be used to test whether a bottom-up account of organization within a brain region corresponds to any variations in behavioral associations. In other words, the enhanced resolution at the level of neurobiology in bottom-up approaches can be used to organize brain-behavior associations mapped across many different task-fMRI studies.

In this research, both developments in connectivity mapping and synthesizing published functional neuroimaging work were leveraged to parcellate speech networks. The neurobiology of language, with longstanding debates about the extent to which the neural architectures for language and speech are shared with other cognitive domains (e.g., Buchsbaum & D'Esposito, 2019; Campbell & Tyler, 2018; Hickok & Poeppel, 2007; Zatorre & Gandour, 2008) presents a particularly salient target system for these approaches, both

of which can extract more detailed information about functional organization within brain regions. The first study of the current work aimed to map the intrinsic functional networks underpinning the planum temporale, an anatomical region that has been traditionally important to functional-anatomic models of speech, but which functional neuroimaging studies have implicated in a range of different functions. Networks within both this region and the larger zone that functional neuroimaging studies have associated with it were used to generate a parcellation describing the spatial arrangement of areas within this zone. The behavioral associations that have been reported within these areas and their corresponding networks were then examined using published functional neuroimaging work. In a second study, the way in which behavioral associations from published neuroimaging work are distributed across temporal and inferior parietal cortex was more carefully scrutinized to map speech networks. Instead of probing behavioral associations within predefined areas as performed in the first study, the behavioral associations themselves were used to carve up functional areas embedded within larger functional networks. First, behaviors were grouped together based on their pattern of association across the brain to form unique association patterns, or functional networks. These functional networks were then used to parcellate temporal and inferior parietal cortex such that each area showed a similar pattern of contribution to the functional networks. In this way, it was possible to define speech networks, and to investigate the extent to which the areas underpinning these speech networks were shared with other functional networks. The functional parcellation that resulted from this approach was validated by parcellating the same area using independent structural connectivity data. The impetus for this research was to provide a more robust localization of functional areas involved in

speech perception than prior meta-analytic studies. These prior studies have disagreed about the locus of speech perception, supporting different functional-anatomic models in the process, and have failed to describe the extent to which speech perception areas may be involved in other speech and non-speech functions. By considering many different behavioral associations when defining functional networks and functional areas, a less domain-specific and more data-driven meta-analysis could be carried out. In the final study of this dissertation, a more specific hypothesis about the involvement of the posterior middle temporal gyrus (pMTG) in error correction for speech was investigated with task-fMRI using a tongue twister paradigm. This study builds on prior work which has shown suggestive evidence that pMTG is involved in error correction by introducing a new class of tongue twister stimuli designed to induce more salient internal speech errors. In this study, published neuroimaging work was used to more carefully dissect the functional roles of areas within a larger speech network that was mapped. The results helped to functionally distinguish two parcels of the pMTG mapped in the second study that were both associated with lexical-semantics.

Chapter 2

Extracting more detailed information about functional organization within brain regions through multivariate meta-analysis and connectivity-based parcellations

The following sections provide a more detailed description of how analysis of aggregate functional neuroimaging data and connectivity can afford a more detailed account of functional organization within brain regions. First, meta-analysis is presented as a useful tool for navigating many inferential challenges when generating brain-behavior associations using task-fMRI. The spectrum of inferential challenges in task-fMRI is explored and specific methods and tools for meta-analysis of task-fMRI data are considered. Further, approaches for multivariate meta-analysis are motivated. Next, mapping brain connectivity is explored as a complementary method for investigating functional organization. The ways in which connectivity can be mapped are discussed, as well as the different ways in which connectivity can be modeled. Evidence that connectivity may be sensitive to cortical areas is surveyed. Finally, the ways in which connectivity data, meta-analysis, and targeted task-fMRI approaches for mapping functional organization can complement each other are discussed in relation to the studies of this dissertation.

2.1 Core inferential challenges in task-fMRI

Functional neuroimaging with task-fMRI has continuously evolved over the last nearly three decades of use in human neuroscience research. This progress has not always advanced without challenges, many of which have been either addressed or assuaged (2020; Kriegeskorte et al., 2009; Logothetis, 2008). Nevertheless, there remain a number of issues, some which are more fundamental than others, that complicate inferences made about brain-behavior relationships with task-fMRI.

One core issue in task-fMRI research is that designing experimental manipulations that are able to purely isolate mental or cognitive processes (used interchangeably here) is remarkably difficult. Tacit in many task-fMRI study designs is the idea that a cognitive process within some task can be altered without disturbing the other processes that make up the task. Under this assumption of pure insertion, the neural correlates of a cognitive process can be mapped by subtracting brain activity during some control task from activity elicited by a version of the same task manipulated to engage that particular cognitive process of interest (Friston et al., 1996; Logothetis et al., 2001; Klein, 2010; Poldrack & Yarkoni, 2016). One complication in the interpretation of cognitive subtraction is that the control task is often baseline brain activity (i.e., in the absence of task). The findings that the brain at rest not only exhibits a rich functional structure (Raichle, 2010), but can also interact with certain task-evoked activity patterns in a complex way undermines straightforward interpretation of many task-fMRI designs relying on this logic (Kubit & Jack, 2013; Northoff et al., 2010; Stark & Squire, 2001).

Moreover, the practice of tinkering with the cognitive components of a task often has unintended and complex consequences (Binder, 2009; Friston et al., 1996; Jennings et al., 1997; Kubit & Jack, 2013; Poldrack & Yarkoni, 2016), and although the assumption of pure insertion can be valid (Zhang et al., 2018), it can also fail (Friston et al., 1996; Jennings et al., 1997). This poses a broader challenge for task manipulations. For instance, ensuring that changes to a task do not place even slightly greater demands on domain-general systems (e.g., sustained or selective attentional, error-monitoring, or response inhibitory systems), or increase task difficulty relative to the control task is non-trivial. Indeed, attentional components in tasks designed to isolate other cognitive processes are a common source of dispute about which functional labels should be assigned to brain regions (e.g., Binder, 2009; Campbell & Tyler, 2018; Hickok & Poeppel, 2000; Kubit & Jack, 2013). Although task-fMRI study designs have developed to allow for alternatives to this subtraction paradigm (Price et al., 1997), which nevertheless remains popular (Poldrack & Yarkoni, 2016), other designs often make the similar assumption that BOLD response for different cognitive components sums linearly (Amaro Jr. & Barker, 2005).

A separate issue for the prospect of mapping functional organization in the brain with task-fMRI is that it necessitates a divide-and-conquer strategy. The cost associated with data acquisition makes it such that a limited number of tasks can be compared in any individual task-fMRI study. Consequently, results must be registered across studies to achieve a more complete description of which functions or behaviors are associated with each brain region.

2.2 Practical challenges in task-fMRI

There are also more pragmatic challenges to inference with task-fMRI. Evidence has been slowly mounting that underpowered task-fMRI studies are endemic, undoubtedly due to data collection costs (e.g., Carp, 2012; Gordon et al., 2017; Button et al., 2013; Cunningham & Koscik, 2017; Dernez et al., 2014; Poldrack et al., 2012; 2017; Thirion et al., 2007). Further, not only does fMRI data have a low signal-to-noise ratio (Parish et al., 2000; Welvaert & Rosseel, 2013), but the predominant mode of analysis is at the group-level, despite demonstrably high individual variability in functional neuroanatomy (e.g., Amunts et al., 1999; Galaburda et al., 1990; Gordon et al., 2017; Mueller et al., 2013; Ojemann, 1979; Steinmetz & Seitz, 1991). As a result, small sample sizes (and short scan times) further reduce the spatial resolution available at the group-level (Costafreda et al., 2009). This effect is independent of interstudy variability stemming from low signal-to-noise (Ioannidis, 2005; 2005). Further, small sample sizes (stemming from low power) are associated with higher rates of both false positives and false negatives (e.g., Ioannidis, 2005; 2005). There are other independent factors too that contribute to interstudy variability in functional localization even in large sample sizes, including equipment (i.e., scanner), decisions about scanner parameters, task selection, study design, preprocessing, and analysis pipelines (e.g., Bennett & Miller, 2010; 2013; Botvinik-Nezer et al., 2020; Carp, 2012; Elliot et al., 2020).

These findings shore up support for increased standardization of fMRI methods (Brown & Behrmann, 2017; Poldrack & Yarkoni, 2016), but issues identified with current standard practices (Eklund et al., 2016) suggest more emphasis should also be placed on

validating functional neuroimaging results using multiple complementary and multidisciplinary approaches. At the same time, it should be qualified that issues of sample size persist in other areas of functional neuroimaging, including brain morphometry and structural imaging (e.g., as often related to behavioral measures; Boekel et al., 2015; Button et al., 2013; Genon et al., 2017; Masouleh et al., 2019), as well as voxel-based lesion-symptom mapping (Lorca-Puls, et al., 2018) and clinical fMRI (Szucs & Ioannidis, 2020).

2.3 Meta-analysis for navigating interstudy variability in task-fMRI

The practical issues that surround research relying on task-fMRI, arguably chief of which is small sample sizes, are not insurmountable. Indeed, there is evidence that neuroimaging studies are slowly moving towards sample sizes that can achieve appropriate power for typical effect sizes in the field (Poldrack et al., 2017; Szucs & Ioannidis, 2020). As such, it's important to note that not all task-fMRI studies suffer from these problems, and it appears likely that fewer future studies will. However, progress has been slow and much of what we know about functional organization in the brain either hinges on or has been informed by results from past studies more likely to exhibit low statistical power. Recent estimates submit that some of the most highly cited neuroimaging studies have a median sample size of 12 (Szucs & Ioannidis, 2020). In stark contrast, multiple meta-studies have indicated that even sample sizes of 20 may be too low to produce reliable results (c.f., Thirion et al., 2007; Turner et al., 2018).

One intuitive approach for tackling many of the practical problems associated with individual neuroimaging studies is to consider the consensus across many studies in tandem, thereby evaluating interstudy variability. Pooling studies in this way increases sample size (e.g., Costafreda, 2009; Eickhoff et al., 2012; Møller & Myles, 2016; Walker et al., 2008) and helps address interstudy variability rooted in methods as detailed in the previous section (e.g., equipment, task or contrast variations, study design choices, preprocessing and analysis pipelines, etc.; Samartsidis et al., 2017). Under the assumption that false-positives in task-fMRI research are not spatially structured, meta-analysis can also provide greater confidence that effects should not have been rejected (e.g., Fox et al., 1998; Costafreda, 2009; Samartsidis et al., 2017).

2.4 Meta-analysis of brain data requires unique frameworks and databases

A number of tools for performing meta-analysis on task-fMRI data have emerged in response to the massive growth in published neuroimaging work over the last decades (Wager et al., 2007). Many of these tools provide statistical frameworks for meta-analysis adapted to the peculiarities of fMRI data, where effects are spatially complex but reduced to single stereotaxic coordinates. A common approach behind these methods is to model foci across selected experiments as a probability distribution (e.g., smoothing foci using a kernel), evaluating the spatial locations where probabilities are most dense (Tukeltaub et al., 2002). These frameworks have become more sophisticated over time—for example, the most common framework of Activation Likelihood Estimation (ALE) now uses interstudy

and interlaboratory variability to represent the spatial uncertainty of individual foci, minimizes the ability of a single study or experiment to dominate effects (e.g., because multiple experiments were performed with the same group or multiple neighboring foci were reported in a single experiment), and has more robust correction for multiple comparisons (Eickhoff et al., 2009; 2012; 2012; Laird et al., 2005). As a result, these frameworks can now provide a more rigorous understanding of the consistency with which task-fMRI effects are reported across studies.

In parallel with the development of tools for performing meta-analysis, many efforts have been aimed at building and maintaining databases that catalogue published neuroimaging work (Dickson et al., 2001; Laird et al., 2005; Nielsen et al., 2004). Although not all of these databases are currently being maintained, the BrainMap database (Laird et al., 2005) continues to be meticulously updated (now encompassing 3,861 studies) according to a specific taxonomy (e.g., covering domains, experimental paradigms, directions, contrast conditions, population groups, handedness, etc.; Fox et al., 2014) and continues to receive substantial attention from the neuroimaging community (e.g., Bottenhorn & Laird, 2021; Hansen et al., 2021; Sheets et al., 2020).

2.5 Meta-analysis can address some deeper challenges in task-fMRI

The availability of both large databases aggregating functional neuroimaging results and tools for analyzing them has presented new opportunities to probe the functional organization of the brain. For example, one implication of task-fMRI studies at large that

has received significant attention is that most brain regions appear to be associated with multiple behaviors, undermining the hope many researchers have harbored that elements of existing psychological and cognitive theories might neatly divide up the brain into functionally specialized regions (e.g., Anderson, 2015; Buzsáki, 2020; Embick & Poeppel, 2005; 2015; Friston & Price, 2005; 2011; Hutto et al., 2017; Klein, 2012; Krakauer et al., 2017; Poldrack, 2010). Although there are outstanding questions surrounding how overlapping brain-behavior relations should be interpreted, it is clear that a more systematic mapping between brain regions and behavioral interventions is necessary for making progress in linking functions to the brain (e.g., Anderson, 2015; Friston & Price, 2005; Hutto et al., 2017; Klein, 2012). Meta-analytic tools provide the means to begin systematically compiling and quantifying the diversity of functions and tasks that have been associated with different parts of the brain, providing a more rigorous approach for synthesizing neuroimaging research than traditionally more qualitative literature reviews. As reviewed earlier, an example of this rigor is that meta-analysis can more critically consider whether the association between some brain regions and behaviors might be the result of any number of factors that tend to increase interstudy variability. Indeed, many studies have already employed these tools to more robustly map the neural correlates of various tasks and behaviors (e.g., swallowing: Soros et al., 2008; Wisconsin Card-Sorting Task: Buchsbaum et al., 2005; viewing paintings: Vartanian & Skov, 2014; etc.).

There are multiple ways in which a more comprehensive understanding of brain-behavior relationships has the potential to enrich accounts of functional organization. First, purely understanding the wider behavioral context in which a brain region responds may bootstrap hypotheses about its latent function (Friston & Price, 2005; Genon et al., 2018;

Klein, 2012; Van Essen, 2009). In other words, it is possible that multiple behavioral interventions engage the same brain region because it performs some elemental computation that these behaviors all involve. Understanding this more elemental operation can guide the development of new behavioral interventions and hypotheses. Further, hypotheses about a region's latent function may benefit not just from the understanding of which behavioral interventions produce a response within it, but also the magnitude of each of these associations. Some behaviors may be more strongly associated with a brain region than others and it may be reasonable to give these behaviors more credit towards the function. Thus, meta-analysis can provide a shift towards focusing on regions and inferring their functions.

A more systematic mapping between brain regions and behavioral interventions also presents the opportunity to leverage neural data for organizing mental constructs. For example, it is not uncommon for researchers to present differing brain response patterns for tasks as some evidence that the mental constructs presumed to underlie these tasks are dissociable, and meta-analysis is often used precisely in this way (e.g., Costafreda, 2009; Hung et al., 2017). In the context of the delicate assumption of pure insertion described previously (see section 2.1), it is possible that the neural differences between some tasks may not represent the exact mental constructs they are thought to index. However, having access to a wider range of brain-behavior associations can help provide an understanding of which mental processes *do* differ between the tasks being compared. Consider the example provided earlier that altering a task to isolate some cognitive process can introduce unforeseen consequences, like increasing task difficulty or placing greater demands on a domain-general system like attention or response inhibition. If the neural

difference between two tasks is concentrated in a brain region that is overwhelmingly associated with a domain-general system, this may indicate that the two tasks were not well-matched, and that the observed neural response may reflect a different mental process than the one under study. By comparing brain response patterns across many behaviors, neural data can be used to inform behavioral models. Notwithstanding debates about whether neurobiological or behavioral data should get privilege in the effort to link behaviors to the brain, it's relatively uncontroversial that neurobiological data is at least relevant to behavioral models and therefore that it can be used to inform them without necessarily being given more privilege (e.g., Anderson, 2015; Buzsáki, 2020; Embick & Poeppel, 2005; 2015; Friston & Price, 2011; Hutto et al., 2017; Klein, 2012; Krakauer et al., 2017; Small & Hickok, 2016). At the very least, understanding which behaviors are neurally indistinguishable provides targets for further investigation.

2.6 The need for data-driven multivariate meta-analysis and larger scales

An issue with conventional meta-analysis of brain data that was hinted at in the preceding section is that it works in a similar piecemeal fashion to individual task-fMRI studies. Meta-analytic tools hold the promise of synthesizing the divide-and-conquer strategy of task-fMRI, but in practice researchers focus on relatively domain-specific meta-analyses to define functional brain units or networks (i.e., what are the neural correlates of *this* behavior?). By fixating on specific domains, researchers overlook massive amounts of potentially useful data, failing to link their area of interest to a rich and broad functional

neuroimaging literature. Thus, the kind of systematic mapping of brain-behavior associations that can produce the insights into functional organization described in the preceding section requires a different analytic strategy. This motivates approaches for meta-analysis which are data-driven and multivariate. Databases aggregating functional neuroimaging results can already facilitate this kind of work by reducing the effort involved in combing through the tens of thousands of studies published each year that mention fMRI (e.g., PubMed trends retrieved ~41,000 studies for the key phrase "fMRI" in 2020).

If the goal of meta-analysis is to piece together the body of functional neuroimaging work across different domains, it is critical for databases aggregating results to keep up with the pace of progress in the field. Moreover, these databases need to remain as current as possible to take advantage of inferential improvements in task-fMRI afforded by slowly increasing sample sizes, standardized and improved preprocessing pipelines, etc. (see section 2.2). Unfortunately, many databases (e.g., BrainMap; see section 2.4) have been slow to adapt, particularly because they have placed a premium on carefully verifying and annotating published neuroimaging work (c.f., Derrfuss & Mar, 2009; Laird et al., 2009; Van Essen, 2009). This has led to the development of approaches for maintaining neuroimaging databases that take advantage of automated text-mining techniques (Nielsen et al., 2003; 2004; Yarkoni et al., 2011). Although this approach produces information about individual studies that is much noisier than the manually annotated approach (e.g., studies are queried based on the frequency with which they use certain behaviorally relevant terms in their abstracts and the retrieved coordinates are not distinguished by contrast), the vastly

greater amount of data that it can harness produces meta-analytic results that are exceedingly similar (Yarkoni, 2011; Poldrack & Yarkoni, 2016).

One advantage to the most comprehensive text-mined database that is unmatched is that it provides a framework for evaluating the specificity of brain-behavior relationships (Yarkoni, 2011). Understanding the behaviors that are more specific to brain areas has the potential to better characterize functions. Conventional meta-analysis evaluates the consistency of reported effects across studies (e.g., Eickhoff, 2012). These consistencies are informative but fundamentally do not reflect specificity, or the extent to which response in the brain implies a behavior is being engaged (e.g., Poldrack, 2006; 2008; 2011; Hutzler, 2014). Consequently, even an exhaustive understanding of associations between a brain region and behaviors based in consistency of activity cannot provide clear insight into functional specificity. For example, if the temporoparietal junction is most consistently activated across studies employing theory of mind tasks, but most other tasks produce only modestly less consistent activity in this region, it would be hard to conclude that activity in temporoparietal junction is indicative of engagement in theory of mind. Yet the success of this kind of reverse inference (so-called because the object of inference is reversed relative to conventional fMRI logic which seeks to infer which brain regions respond to a behavior) is precisely what is necessary to achieve an understanding of which behaviors are more indicative of the functions brain areas carry out. Put differently, a more formal framework for evaluating brain-behavior consistencies relative to each other can generate more focal brain-behavior associations.

The Neurosynth database provides a framework for defining brain-behavior relationships grounded in specificity by quantifying the probability that a term is used in an

article given that activity is present within a brain region (Yarkoni, 2011). Note, that in this case terms are treated as an index for behavior. This approach demonstrably produces more focal brain-behavior associations compared to evaluations of consistency in activity (Poldrack & Yarkoni, 2016). The accuracy of this reverse inference partly contingent on the scope of the data used. This is because specificity is defined in relation to other behaviors. Larger databases are more likely to capture behavioral heterogeneity (Poldrack, 2011). For example, the range of behaviors available in BrainMap are limited to a predefined taxonomy and may be biased as a result of allowing researchers to nominate studies into the database.

2.7 Complementary approaches for mapping functional organization are necessary

Although the previous sections have outlined how meta-analysis can help address inferential issues in task-fMRI, there are still complexities to interpreting both task-fMRI and meta-analytic data that warrant studying the functional architecture of the brain from the perspective of multiple complementary modalities. One major challenge to the interpretation of results from task-fMRI is that using behavioral tasks to isolate mental constructs is subject to a number of delicate assumptions that are often difficult to confirm (e.g., pure insertion; see section 2.1). Further, while meta-analysis can help address issues in task-fMRI, it is not a panacea and the quality of the results is contingent on the quality of the data analyzed (e.g., Button et al., 2013; Van Erp et al., 2017). Consequently, combining task-fMRI with methods that can probe brain function while making fewer assumptions, or

while avoiding some of the practical constraints of task-fMRI, like the difficulty of scaling up sample sizes or even task comparisons, can help produce a more balanced account of brain organization. The following sections argue that one candidate approach that can complement task-fMRI data in this way—as well as others—is connectomics.

A fundamental principle of brain organization is the integration and segregation of functions. On the one hand, the brain is organized into a mosaic of areas with typically sharp boundaries defined by converging variations along different neurobiological properties (e.g., Amunts et al., 2020; Eickhoff et al., 2018). On the other hand, brain areas are densely interconnected along both short and long-range pathways that embed them within larger networks (e.g., Passingham et al., 2002; Van Essen, 2013). Thus, the fact that neurobiological properties can distinguish areas that are likely to be functionally specialized does not diminish the importance of interaction between brain areas for functional organization. In this context, it may be less surprising that behaviors haven't always mapped so neatly onto the brain in task-fMRI studies (i.e., in a one-to-one fashion). It may be the case that many behaviors are better explained by the coordination of activity across wide networks rather than activity within a specific area (e.g., Poldrack, 2011; Pessoa, 2014; Klein, 2012). As a result, understanding how the brain is interconnected is a critical component to mapping functional organization.

2.8 Mapping connections

Several modalities can be used to investigate connectivity in the brain. Diffusion tractography takes advantage of the fact that water diffuses more readily along fiber

bundles to reconstruct white matter pathways from diffusion weighted imaging data, thus providing information about how brain regions are physically embedded within networks. Seeding from some brain region, diffusion tractography involves propagating streamlines from that seed through neighboring voxels based on the direction of diffusion that is modeled within those voxels (e.g., Jbadi & Johansen-Berg, 2011). When the direction of diffusion is represented as a distribution, many streamlines can be repeatedly propagated by sampling randomly from these distributions to estimate the probability of a connection (e.g., Behrens et al., 2007). Consequently, structural connectivity between areas does not have to be represented in binary terms and several variations of the probabilistic approach to propagating streamlines, as well as post-processing techniques, can give other information about the structural connections between brain areas (e.g., connectional densities; Smith et al., 2015).

Brain connectivity can also be conceptualized in terms of functional relationships. That is, areas within different structural networks may still interact, forming polysynaptic connections, even though they are not directly connected by monosynaptic pathways (e.g., Honey et al., 2009). Functional connectivity is typically inferred from the statistical dependence between BOLD response in different brain areas (e.g., Bastos & Schoffelen, 2015). Thus, just like structural connectivity, functional connectivity at least requires a seed to which BOLD response across the brain can be related. Although functional connectivity can be computed during behavioral interventions, it is typically studied in the absence of task, as participants are instructed to "lie still, think of nothing in particular, and do not fall asleep" (Smith et al., 2013). Brain activity during this resting state is highly structured and forms a set of organized networks that appear to be hierarchically arranged

and can be distinguished based on their respective timecourses (e.g., Beckmann et al., 2005; Doucet et al., 2011; Fox et al., 2005; Smith et al., 2009). These intrinsic functional networks are highly consistent across subjects (e.g., Van Dijk et al., 2010; Gratton et al., 2018), but do exhibit some intersubject variability that is both persistent across sessions and predictive of individuals (Shehzad et al., 2009; Gratton et al., 2018; Finn et al., 2015). Further, many of these intrinsic networks are comprised of areas that have been previously shown to coactivate during various behavioral interventions (Cole et al., 2014; Fox & Raichle, 2007; Gratton et al., 2018; Hermundstad et al., 2013; Krienen et al., 2014; Mennes et al., 2010; Smith et al., 2009). Indeed, intrinsic networks can successfully predict activation during task-fMRI (Cohen et al., 2019; Hermundstad et al., 2013; Lacosse et al., 2021) and individual differences in intrinsic networks carry over into task networks (Shah et al., 2016). Thus, it appears that task and resting state fMRI reflect the same functional architecture of the brain, which is both relatively stable, and unique to individuals. Curiously, a number of studies have also highlighted strong similarities between intrinsic functional networks and structural networks, even though the areas that comprise these networks are not always shared (e.g., Greicius et al., 2009; Hermundstad et al., 2013; Honey et al., 2009; Liegeois et al., 2020; Skudlarski et al., 2008).

2.9 Network properties as a window into functional organization

The fact that brain areas do not act in isolation has driven a strong interest in modeling brain networks as graphs whose properties can be further investigated through the

framework of graph theory (e.g., Park & Friston, 2013; Van Den Heuvel & Pol, 2010). Overall, graph theoretic approaches provide a unique perspective on functional organization from the viewpoint of interacting brain networks. For example, although behavioral interventions clearly generate greater *activity* within specific areas compared to activity in the brain at rest, studies (mostly) focusing on graph theoretic properties such as modularity and efficiency have shown that the transition from rest into a behavioral state tends to correspond to *functional connectivity* changes that reflect increased integration of information between networks and decreased integration of information within networks (Di et al., 2013; Gao et al., 2013; Spadone et al., 2015; Wig, 2017; Wang et al., 2021). Nevertheless, whether connectivity is measured at rest or during some behavior, networks appear to retain their small-world architecture, and are characterized by high local clustering coefficients and short path lengths (e.g., Di et al., 2013). Thus, brain organization is always characterized by many local networks that efficiently communicate by way of highly interconnected hubs (e.g., Bullmore & Sporns, 2012). Going further, some studies have indicated that the delicate balance between integration and segregation in the brain at rest may be a predictor of the success of many different kinds of behavioral interventions (e.g., Gallen & D'Esposito, 2019). For instance, brains that exhibit higher modularity at rest appear to show stronger functional recovery from damage (e.g., in stroke; Siegel et al., 2018) and are associated with larger cognitive improvements with training (e.g., in older populations; Gallen et al., 2016).

2.10 Defining networks

A critical issue that has been so far ignored is that understanding how brain networks interact requires defining networks in the first place. In the context of a graph, nodes (i.e., brain areas) need to be defined over which edges (i.e., connectivity) can be computed. For that matter, estimating connectivity at all requires some seed (e.g., section 2.8). Consequently, studies have to rely on brain atlases generated from other groups of subjects and often using other modalities than the one under investigation. For example, it's not uncommon for resting state studies to use anatomical atlases (e.g., Cohen & D'Esposito, 2016), even though they are based on macroscopic landmarks that have been known not to align with cortical areas since Brodmann's early histological work (Amunts & Zilles, 2015). Considering that the brain at rest reflects its functional architecture (e.g., section 2.8), an anatomical atlas may poorly represent the functional areas that are embedded in resting state data. Indeed, evidence is accumulating that network graphs and their properties are strongly impacted by atlas choice (e.g., de Reus & Van den Heuvel, 2013; Luo et al., 2021; Messé et al., 2019; Park et al., 2013; Wang et al., 2009). Many differences between network properties that occur as a function of behavioral intervention or group can change in significant and unpredictable ways when brain areas are defined based on a participant's unique data, and when the definition of brain areas is tailored to the specific modality being investigated (Luo et al., 2021; Salehi et al., 2020).

The impact of atlas choice on connectivity affects other studies relying on connectivity data as well. For instance, atlas choice substantially impacts relationships between connectivity and behavioral measures reflecting cognitive abilities (Bryce et al.,

2021). Thus, defining areas over which connectivity should be investigated is a fundamental and non-trivial problem that is largely underappreciated (e.g., Bijsterbosch et al., 2020; Bohland et al., 2009; de Reus & Van den Heuvel, 2013; Eickhoff et al., 2015; 2018; 2018). Fortunately, connectivity data can also be leveraged to define brain areas. Although this kind of work provides less insight into how networks interact, it is valuable for addressing the more fundamental and open question of how the brain is interconnected in the first place.

2.11 Mapping brain areas using connectivity

The gold standard for mapping brain areas relies on invasive methods. Cortical areas can be distinguished based on their unique microstructural properties and their pattern of connectivity, with both features shaping areal functions (e.g., section 2.7; Eickhoff et al., 2018; Passingham et al., 2002; Van Essen, 2013). Classically, delineating these areas has involved tracking changes in microstructural properties assessed post-mortem using histological data. For example, cytoarchitecture, myeloarchitecture and chemoarchitecture have all generated (Amunts & Ziles, 2015), and continue to generate (Amunts et al., 2020), incredibly detailed maps of cortical areas. Yet, many cortical areas show a substantial degree of variability across individuals (Spocter et al., 2010; Zachlod et al., 2020) and registration between invasive and noninvasive methods is made imprecise by the typically small sample sizes and limited brain coverage that is characteristic of invasive studies (Dell'Acqua et al., 2013; Glasser & Van Essen, 2011). As a result, it's unclear how well histological atlases can generalize to in vivo imaging data. However, the fact that the microstructural properties of cortical areas appear to be inextricably related to their

connectivity patterns (Goulas et al., 2018; Passingham et al., 2002) can be leveraged to map brain areas that are likely to be functionally distinct in vivo (e.g., Eickhoff et al., 2015).

Studying connectivity data at a level that is less coarse than discussed in previous sections can provide a different kind of insight into functional organization in the brain, demonstrating how larger regions are configured based on putatively functionally distinct areas that participate in different networks. This kind of analysis is typically referred to as a connectivity-based parcellation, and usually involves grouping imaging units like voxels together into parcels, or areas, based on the similarity of their connectivity profiles (Eickhoff et al., 2015). Connectivity-based parcellations can be applied to both structural and functional connectivity to find representations of areal organization that are more or less influenced by polysynaptic pathways. These parcellations typically rely on unsupervised machine learning techniques that can help evaluate subtle patterns in the data and model areas with homogenous connectivity profiles (Bijsterbosch et al., 2020; Eickhoff et al., 2015; 2018). It is worth noting that this method does not consider the spatial distance between the brain units being grouped together. That is, an 'area' defined on a specific connectivity pattern can span multiple non-adjacent brain units. While this differs from the focus on local changes in neurobiological properties in histological work, it is worth noting that local approaches also end up delineating areas that are overall homogenous in terms of neurobiological features (e.g., connectivity; Gordon et al., 2016). However, local approaches can require manual intervention based on anatomical priors to define areas (Glasser et al., 2016). Further, the approach of maximizing homogeneity within areas ensures the resulting parcellations can be better used as a form of neurobiologically informed dimensionality reduction in future datasets (Eickhoff et al.,

2018). For example, maximally homogenous areas defined on functional connectivity will better reflect unique timecourse signatures in resting state data.

2.12 Connectivity-based parcellations may be sensitive to microstructural organization

Although connectivity-based parcellations can distinguish smaller areas within larger brain regions, the extent to which they can approximate cortical areas mapped invasively is uncertain. Clearly neuroimaging cannot provide the kind of microscopic resolution of histological studies. However, inferences about microstructure can be made at this scale on the basis that the data being collected reflects a combination of signals generated at the microscopic scale (e.g., Weiskopf, et al., 2015). Some evidence suggests that in vivo imaging can be sensitive to cortical areas typically accessed by invasive methods. For example, studies have indicated that the lamination pattern on T1-weighted images reflects histological properties (Eickhoff et al., 2005). More recently, studies have used T1-weighted and/or T2-weighted structural scans to estimate myelin content with success (e.g., Edwards et al., 2018). A number of these studies have reported correspondence between such myelin maps and myeloarchitectonic boundaries, particularly at the group-level and closer to sensory cortex (Glasser & Van Essen, 2011; Van Essen et al., 2019). Further, these myelin maps have been shown to closely correspond to resting state functional connectivity patterns in sensory cortex but have also been shown to deviate when compared to connectivity patterns in other areas (c.f., Glasser et al., 2016; Huntenburg et al., 2017). More directly, some studies have reported significant correlations

between functional connectivity at rest and tracer-based cellular connectivity estimated across brain areas (Hori et al., 2020).

Some evidence also exists that structural connectivity measured in vivo captures features of connectivity mapped invasively. For example, ex vivo tractography produces end points that are largely in agreement with tracer studies (Gao et al., 2018), irrespective of which tractography algorithm is used (roughly 76% agreement across different tractography methods; Girard et al., 2020). Further, there are some reports that ex-vivo tractography shows good agreement with chemoarchitectural features in the cerebellum (Dell'Acqua et al., 2013). In a more direct test of the sensitivity of ex-vivo tractography to cortical areas, one study has shown that structural connectivity patterns could be classified with reasonable success into cytoarchitectonic areas mapped in the same brains (Gao et al., 2018). In addition, there have been a number of studies that have indicated correspondence between microdissections and in vivo tractography (Benedictis et al., 2016; Maffei et al., 2018; Meola et al., 2015). One caveat that is worth highlighting about the work comparing tractograms directly to structural networks mapped invasively is that generating tractograms requires making a number of subjective decisions independent of imaging parameters or tractography algorithms (e.g., streamline thresholds) that can impact the sensitivity and specificity of the resulting comparisons (Girard et al., 2020).

Connectivity-based parcellations themselves are often compared to architectonic parcellations, with many studies reporting generally strong qualitative agreement between features, and often strong spatial overlap between parcellations when architectonic atlases are available for the region (e.g., Anwander et al., 2007; Beckmann et al., 2009; Buckner & Yeo, 2014; Cloutman & Lambon Ralph, 2012; Cha et al., 2017; Chao et al., 2009; Chang et al.,

2013; Fan et al., 2013; Genon et al., 2017; Gordon et al., 2016; Glasser et al., 2016; Kahnt et al., 2012; Li et al., 2013; Liu et al., 2013; Long et al., 2014; Mars et al., 2011; Wang et al., 2015; 2017; Wu et al., 2009; Zhang et al., 2014). Critically, correspondence between architectonic and connectivity-based parcellations can be present even in the presence of inaccuracies caused by the imaging method used to define networks in vivo. This is because it is possible for the imaging data to remain sensitive to the fact that areas lie on *different* networks, even if the definition of these networks is subject to some error. Put more concretely, seeding from voxels belonging to two different cortical areas may produce inaccurate tractograms (e.g., because streamlines eventually get knocked off course when passing through voxels with crossing fibers/complex geometry, or when succumbing to the many biases exhibited by tractography) that are nonetheless consistently different such that these areas can be distinguished. It's worth mentioning in the context of this specific example that some of the previously discussed studies have demonstrated that network definition can remain reasonably accurate despite errors incurred during streamline propagation (including false-positive connections; Gao et al., 2018; Girard et al., 2020). The same potential for distinguishing cortical areas despite inaccuracies in network definition is applicable to functional connectivity data.

This is not to say, however, that agreement is always seamless between connectivity-based and architectonic parcellations. Indeed, some areas within connectivity-based parcellations can show nearly perfect spatial overlap, while others may show less agreement (e.g., Buckner & Yeo, 2014; Wang et al., 2015). There are many reasons for which correspondence cannot be perfect, and as such it is remarkable that there *is* strong correspondence at all. Foremost, neuroimaging provides indirect information about

neurobiological features, so any correspondence should be approximate at best. One issue mentioned previously is that it is unclear how well architectonic parcellations should generalize to imaging data given the small sample sizes in invasive work (see section 2.11). Another issue that has been brought up earlier is that different imaging modalities, reflecting different neurobiological properties, can also produce areas with slightly different boundaries (see section 2.10). This is relatively unsurprising as different imaging modalities will have different biases. Moreover, although so far the discussion about cortical areas has been couched in sharp boundaries that converge along many neurobiological properties, this is essentially an oversimplification. In reality, boundaries occasionally shift depending on the neurobiological property investigated (Eickhoff et al., 2018).

Occasionally, connectivity-based parcellations can contain more areas than architectonic work has mapped. Reassuringly, these additional areas are usually found to correspond to functional distinctions probed with task-fMRI and therefore may represent subareal features (e.g., Genon et al., 2017; 2018; Glasser et al., 2016). For example, some studies have mapped areas that correspond to topographic organization within motor cortex (Long et al., 2014; Gordon et al., 2014; Yeo et al., 2011) or visual eccentricities (c.f., Buckner & Yeo, 2014; Haak et al., 2018). However, not all differences between parcellations may be rooted in such functional distinctions, and it is worth pointing out that many connectivity-based parcellations are motivated specifically to parcellate larger brain regions that task-fMRI studies have difficulty dissecting (e.g., Anwander et al., 2007). A potential source of discrepancy between connectivity-based parcellations and architectonic work (at the level of areas) is that it is unclear how to establish the number of unique

networks that can be distinguished by the neuroimaging data. That is, the kinds of unsupervised learning algorithms that are typically used to define parcellations require an a priori decision about the number of areas contained in a region (sometimes indirectly through another parameter that must be chosen in the same way). A number of heuristics can be used to make a choice about the parcellation complexity supported by the data (i.e., number of areas), but these heuristics can produce widely different results and some are better suited than others for evaluating different methods. Thus, a major obstacle to this kind of research is a lack of comparative studies that can facilitate standardization (e.g., Eickhoff, 2015).

2.13 Meta-analysis as a complement to connectivity based parcellations and targeted task-fMRI studies

Connectivity-based parcellations provide a different approach to task-fMRI for mapping functional organization in the brain by focusing on networks and considering subtle differences in connectivity patterns between voxels. Connectivity data has several attractive properties that make it well-suited for defining functional areas. Foremost, structural and functional connectivity data permits the definition of multiple networks, and thereby areas, in a single scan. In task-fMRI, these networks can only be accessed through many different behavioral interventions, which is an expensive proposition. As a consequence of requiring fewer resources for mapping multiple areas, it is possible to collect connectivity data in much larger sample sizes relative to task-fMRI. While meta-analysis can help address existing issues of sample size in task-fMRI (see section 2.3), the

higher heterogeneity of results that characterizes small studies, and the existence of publication bias can affect the results (e.g., Acar et al., 2018; Button, 2013; Dubois, 2016; Turner et al., 2018). By the same token, multivariate meta-analysis can define networks that span many different behavioral interventions (like connectivity data), but the results are subject to the same concerns. Perhaps the most attractive property of connectivity data is that it can be used to define areas in a data-driven way that sidesteps some of the core issues that cloud inference in task-fMRI. Because no behavior is explicitly studied, there are no concerns associated with the assumption of pure insertion, or the difficulty of establishing and isolating the latent constructs being studied in a behavioral intervention (see section 2.1). However, connectivity approaches are fundamentally complementary because they cannot provide the kind of rich information about brain function that is available to task-fMRI.

One approach for unifying task-fMRI and connectivity data is to capitalize on the fact that meta-analysis permits a region-focused perspective that facilitates inferences about the behaviors in which brain regions are involved (see section 2.5). In other words, connectivity data can be used to define many functional areas with potentially more precision, and meta-analytic approaches can be used to interpret the behavioral profiles of these areas, hinting at the computations they may be carrying out (Genon et al., 2018). In this way, behavioral differences between connectivity-defined areas can be explored and the unique combinations of behaviors associated with each area can be used to help understand the underlying function it may be performing. This approach is typical in connectivity-based parcellation studies (e.g., Wang et al., 2015), although it is less common to define brain-behavior associations using the more specific reverse-inference strategy

discussed previously (e.g., Chang et al., 2012; Genon et al., 2018; see section 2.6 for more information about reverse inference).

The strength of meta-analysis for complementing connectivity data is grounded in the quality of the individual studies that can be pooled together. Consequently, the fact that task-fMRI studies can be complicated by the various issues that have discussed throughout this chapter does not diminish their importance for mapping functional organization. Indeed, task-fMRI remains arguably the most favorable method for testing new theories about brain function. For example, while understanding the range of behaviors that produce response within a brain area can spur hypotheses about the more elemental computation it may be performing (Genon et al., 2018), testing these theories requires a more targeted functional neuroimaging approach. It is worth mentioning too that meta-analysis can complement individual task-fMRI studies. The same region-focused perspective that permits inferences about the functions of brain areas delineated with connectivity data can be used to make inferences about the functions that brain areas perform within a task network. For example, as discussed earlier in this chapter, it can be difficult to isolate cognitive processes cleanly with behavioral tasks (e.g., the assumption of pure insertion, inadvertent task confounds like attention). This implies that many behavioral interventions may observe patterns of brain activity associated with additional processes that are not of interest. Decoding brain response mapped in individual task-fMRI studies can highlight these processes. That meta-analysis can also complement individual task-fMRI studies highlights that all of these methods for studying functional organization are truly complementary to each other and can provide unique insights based on their individual strengths.

2.14 Current work

The first study of the current dissertation combined a connectivity-based parcellation and meta-analysis of published neuroimaging work exactly as described above, and in line with prior studies. The goal of this work was to parcellate the planum temporale, a region that has traditionally played an important role in functional-anatomic models of language, but which has been associated with many different behavioral interventions with task-fMRI. Although a few whole-brain functional connectivity-based parcellations have addressed the networks underpinning this region, no region-based parcellations have. Crucially, focusing connectivity-based parcellations on regions provides better sensitivity to differences in connectivity patterns (Eickhoff, 2015; 2018), and whole-brain parcellations of this area based on functional connectivity have not been able to show much functional diversity in this region (Joliot et al., 2015). In the service of producing a reliable parcellation, this study focused on many issues described elsewhere in this chapter. For instance, large sample sizes and state-of-the-art resting state acquisition and preprocessing methods were used in an effort to improve the reliability of the results. In addition, a framework for measuring reliability was implemented to test which parcellation complexities (i.e., number of clusters) and methods could produce the most reliable representation of the areas within the region. Finally, the results were directly benchmarked against existing atlases using independent data.

Mapping brain-behavior associations from task-fMRI onto areas defined on connectivity privileges connectivity data for mapping brain areas. Although there are reasons to think that connectivity data may define areas with more precision (see section

2.13), the extent to which aggregated task-fMRI data may be sensitive to the same areas is an open question. The second study of the current research broadly addressed this question. As described previously, multivariate meta-analysis can be used to define patterns of activity that cross many different behaviors, effectively establishing unique functional networks (see section 2.6). In turn, these activation patterns, or networks, can be treated much like connectivity data to produce a functional parcellation that defines areas based on homogenous network association patterns. In other words, each area can be defined such that its degree of involvement in the functional networks that carve up the behavioral space is similar. This procedure was used in the second study of the current work to generate a functional parcellation, which was compared to an independently produced structural connectivity-based parcellation of the same region. Although understanding whether aggregated published task-fMRI data could be used to produce the same functional areas as connectivity data was of general interest to this study, it was not its chief focus. Instead, the main objective of the study was to localize speech perception processes using a multivariate meta-analytic approach deployed over a large sample of studies. This aim was motivated on the basis that previous meta-analyses attempting to localize speech perception have used sample sizes that are insufficient, produced conflicting results, and relied on relatively narrow definitions of which tasks index speech perception processes. By mapping the networks that are shared across different behaviors and tasks, this study was able to examine which aspects of speech perception could be dissociated based on the brain response patterns they elicited. The functional parcellation based on these networks helped to describe the extent to which the individual areas that comprised these speech networks were associated with other behaviors. The structural

parcellation not only helped to validate these areas but indicated which structural networks supported them.

In the speech networks that were mapped as part of the second study, the posterior middle temporal gyrus (pMTG) emerged as an important area for speech processing. However, a portion of the pMTG exhibited much greater functional complexity, activating under a broader range of conditions. The third study demonstrated that this more functionally complex portion of pMTG may be involved in internal error correction for speech. This study was motivated by prior work that has used a tongue twister paradigm to suggest that internal error detection and correction of word-level speech errors involves pMTG on the basis that this area tends toward showing a stronger response when *potential* speech errors are biased towards nonwords compared to words. That is, this prior work found response in pMTG despite the fact that both of the conditions being compared involved reciting the same sets of words, and the fact that no overt errors occurred during recitation. Nevertheless, this effect did not survive cluster correction. The third study of this dissertation attempted to replicate the effect reported by prior work in a larger sample of participants. In addition, novel taboo tongue twister stimuli were introduced to induce the potential for errors that are more salient to the internal error detection and correction mechanism. To foreshadow the results, although the pMTG presented a robust taboo effect, a larger network of areas were unexpectedly involved as well. Brain-behavior associations across task-fMRI studies were used to understand the roles of these areas within the larger network that was mapped.

Chapter 3

Functional parcellation of the posterior perisylvian zone

3.1 Introduction

3.1.1 The Planum Temporale (PT)

The posterior perisylvian cortex has historically played a central role in functional-anatomical models of language. Carl Wernicke famously argued for a two-stage model of comprehension that implicated the entire superior temporal gyrus in the process of recognizing speech sounds and guiding speech production (Wernicke, 1874/1997). Although assessments of lesions primarily in patients with fluent aphasia have expanded the territory of so-called Wernicke's area (e.g., Bates et al., 2003; Damasio & Geschwind, 1984; Naeser et al., 1987; Selnes et al., 1983), they have overwhelmingly emphasized the importance of more posterior portions of the superior temporal gyrus (e.g., Jacobs & Scheibel, 1993; Luria, 1958; Masland, 1968; Naeser et al., 1987; Penfield & Roberts, 1959; Selnes et al., 1983; for reviews see: Binder, 2015; 2017; Bogen & Bogen, 1976; Damasio & Geschwind, 1984; Dick & Tremblay, 2012). This has helped to develop the view that a strongly leftward asymmetric anatomical area within auditory association cortex, called

the planum temporale (PT), constitutes the functional core of Wernicke's area (Geschwind & Levitsky, 1968).

A number of findings have since upset the theory that the PT subserves speech recognition. Chimpanzees (Gannon et al., 1998) and other great apes (Hopkins et al., 1998; Marie et al., 2017; Yeni-Komshian & Benson, 1976) show asymmetry of the PT. In humans, functional neuroimaging has both challenged the association between language lateralization and morphology of the PT (e.g., Eckert et al., 2006; Keller et al., 2011), and repeatedly demonstrated that the perception of speech and non-speech sounds activates the PT just as well (e.g., Belin et al., 1998; Binder et al., 2000; Hickok et al., 2003; Specht & Reul, 2003; Wise et al., 2001). The lack of selectivity for speech sounds in the PT has naturally pointed to a role for this area in auditory analysis. Indeed, studies have linked it to the perception of complex melodies and music (Callan et al., 2006; Gaab et al., 2003; Griffiths et al., 2000; Patterson et al., 2002; Tervaniemi et al., 2000), analysis of pitch (Griffiths et al., 1999; Warren & Griffiths, 2003; Warren et al., 2005; Xu et al., 2006), perception of rapidly changing acoustic cues (Elmer et al., 2011; Jancke et al., 2002), tracking of statistical properties embedded in sounds (Overath et al., 2007; Tremblay, 2013b), discrimination of novel sounds (Jääskeläinen et al., 2004), and spatial hearing (Alain et al., 2005; Barrett & Hall, 2006; Deouell et al., 2007; Krumbholz et al., 2005; Patterson et al., 2002; Pavani et al., 2002; Smith et al., 2004; 2007; 2010; Warren et al., 2002; Zatorre et al., 2002; Zundorf et al., 2013).

3.1.2 A single function in the PT

One view that has attempted to reconcile the range of auditory stimuli to which the PT responds has proposed that this region serves a singular function: categorizing sounds by matching them to stored templates of auditory representations (Griffiths & Warren, 2002) using spectrotemporal analysis (Warren et al., 2005). Because this operation is thought to occur over both shorter and longer timescales, it is speculated to be able to support spatial analysis of acoustic information, separating it from information about sound identity in the auditory signal. Support for this view comes from studies that have reported spectrotemporally modulated response in the PT over different timescales (Meyer et al., 2005; Overath et al., 2008; Warren et al., 2005; but see Langers et al., 2003; Schonwiesner & Zatorre, 2009).

An alternative account of a single function within the PT sees its response pattern as being consistent with a region that contains auditory representations and assigns it the functional role of supporting auditory imagery (Price, 2012). This is consistent with the findings that response in the PT to speech is not purely driven by external auditory stimulation. Both overt (e.g., Bohland & Guenther, 2006; Peschke et al., 2009; Sörös et al., 2006) and covert (e.g., Graves et al., 2008; Hickok et al., 2003; Okada et al., 2018; Price et al., 2011; Wise et al., 2001) production of speech has reliably demonstrated activity in this area, with some studies reporting stronger activity during covert production tasks (Andreatta et al., 2010; Kleber et al., 2007). However, studies have also shown activity in the PT during imagined speech, including in the second and third person (Aleman et al., 2005; Jancke & Shah, 2004; Shergill et al., 2001). Further, the PT activates during the silent

watching of environmental sounds being produced (Bunzeck et al., 2005), or a finger tapping on a piano (Hasegawa et al., 2004), and during the process of completing familiar music in silence (Kraemer et al., 2005; but see Zatorre & Halpern, 2005).

3.1.3 The structure of the PT complicates investigations into its functions

The work of mapping out what functions may be supported by the PT faces serious challenges. Foremost, the PT lacks straightforward definition, especially around its anterior and posterior borders (e.g., Leonard et al., 1998; Tzourio-Mazoyer et al., 2017; Tzourio-Mazoyer & Mazoyer, 2017; Westbury et al., 1999; Zetzsche et al., 2001; see Shapleske et al., 1999 for review). As a result, foci of activity reported in studies may be misclassified as belonging to the PT or neighboring anatomical areas, depending on which definition of the PT is operationalized (e.g., inclusion/exclusion of second Heschl's gyrus or vertical segment of the Sylvian fissure). Further, the PT shows high variability across individuals (e.g., Knaus et al., 2006; Shapleske et al., 1999; Tzourio-Mazoyer et al., 2017; Tzourio-Mazoyer & Mazoyer, 2017), making it difficult to localize effects to this area in groups (note also the previous discussion about small sample sizes being endemic in groups; Chapter 1). When spatial smoothing is applied to improve localization, specificity is lost, making it more likely that one larger region is associated with many different tasks. For example, effects caused by smoothing can demonstrably complicate localization of auditory response to either auditory cortex and/or PT (Tahmasebi et al., 2009), and can entirely remove relatively focal patterns of activity that have been observed within the PT (Buchsbaum et al., 2005). At the same time, the cortical fields underlying the PT extend beyond its macroscopic

boundaries, which makes it possible that functional response just outside of the PT is misattributed to some other anatomically defined area. In general, macroscopic landmarks are poor markers of architectonic borders (e.g., Morosan et al., 2001). Cortical fields within the PT can stretch into the superior temporal sulcus, and the opercular and inferior parietal cortices (Fullerton & Pandya, 2007; Galaburda & Sanides, 1980; Rademacher et al., 1996; Sweet et al., 2005). Thus, although a diverse range of response properties have been linked to the PT in task-fMRI, it is difficult to know whether these response properties truly reflect computations carried out in the PT.

3.1.4 Microstructure of the PT implies multiple functions

An issue for accounts that unite response properties of the PT under a single function is that architectonic studies have consistently distinguished three or more cortical fields in this area, suggesting that it is comprised of multiple functionally dissociable regions (for reviews see: Baumann et al., 2013; Morosan et al., 2005b; Clarke & Morosan, 2012; Hickok & Saberi, 2012). One line of research focusing on cyto- and myeloarchitecture has delineated: i) an anteromedial area of the PT that is located mostly within the lateral fissure, but which extends ventrally to include a slice of the superior temporal gyrus, ii) a posterolateral area abutting it that spans a large portion of the superior temporal gyrus and borders the ventrolateral extension of the anteromedial cortical area, and iii) a dorsomedial area linked to the auditory core and encroaching on the parietal operculum (Fullerton & Pandya, 2007; Galaburda & Sanides, 1980). Histochemical methods (Rivier & Clarke, 1997; Wallace et al., 2002) have largely upheld this parcellation (Fullerton & Pandya, 2007), describing medial and lateral areas on the temporal plane, and in some

cases an area on the lateral superior temporal gyrus. In addition, the same studies have isolated a smaller cortical field just posterior to Heschl's gyrus that has been associated with the auditory core (e.g., Morosan, 2005a,b; Wallace et al., 2002).

Observer-independent methods for delineating cortical fields using cyto- and receptorarchitecture have produced a similar parcellation scheme overall and have helped clarify the organization of areas in superior temporal gyrus (STG) and superior temporal sulcus (STS). This research has provided support for distinguishing areas of the superior temporal plane from what other work has described as their extensions into STG and STS. This work has mapped a lateral STG area, as well as two more ventral areas. Both of these ventral areas cover portions of STS, but one has greater overlap with the ventral bank of the STG (Zachlod et al., 2020).

In the most recent line of observer-independent architectonic work, some portions of cortex near posterior PT remain unmapped (Amunts et al., 2020; Gulban et al., 2020). Within this unmapped space, older studies have consistently demonstrated the existence of an additional cortical area, often called Tpt, that is situated within posterior PT, but extends into STG, STS, and inferior parietal cortex (e.g., Economo, 1925; Fullerton & Pandya, 2007; Galaburda & Sanides, 1980; Sweet et al., 2005). Area Tpt marks the transition between temporal and parietal cortex, and is typically described as auditory-related, yet noticeably lacking many features associated with auditory cortex (Clarke & Morosan, 2012; Fullerton & Pandya, 2007; Galaburda & Sanides, 1980; Hackett & Kaas, 2004; Smiley & Falchier, 2009; Spocter et al., 2010; Sweet et al., 2005).

In summary, slight variations on a relatively consistent parcellation of the PT have been reported by histological studies over the years, many of which have focused on

different microstructural properties. This is unsurprising given that architectonic borders generally mark a transition of many microstructural features of the brain, but not always all of them (Amunts & Zilles, 2015). A common pattern we can extract from histological work is that there appear to be at least two nonprimary auditory areas that lie on the lateral fissure, just posterior to HG. In addition, there is strong evidence of a third area in posterior PT that fails to resemble most of auditory cortex and marks the transition into parietal cortex, showing intriguing signs characteristic of multisensory areas. Finally, it's possible that one other area in lateral STG overlaps with the PT.

Figure 1: Cytoarchitectonic and histochemical parcellations of auditory cortex and surrounding posterior perisylvian zone

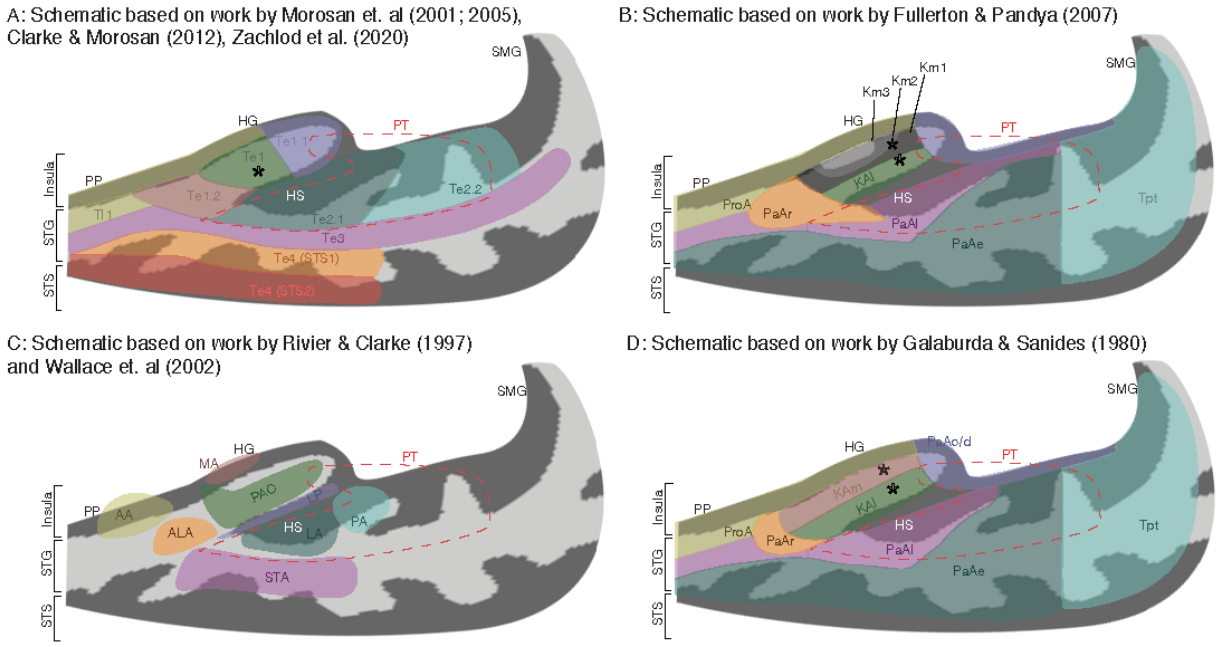

Figure 1: Cytoarchitectonic and histochemical parcellations of auditory cortex and surrounding posterior perisylvian zone. Schematics of various parcellations of auditory cortex and the planum temporale. Anatomically defined planum temporale (PT) is superimposed in red (definition based on maximal overlap across participants in anatomically defined PT from the Harvard-Oxford atlas). Stars represent auditory core where specified. Several other anatomical regions are labeled in grey or white, including the planum polare (PP), Heschl's sulcus (HS), Heschl's gyrus (HG), and the supramarginal gyrus (SMG). Labels are colored to correspond to the cortical fields they refer to.

3.1.5 Evidence for speech specific computations in an area of the PT that shares features with Tpt

Other evidence that the PT may be comprised of multiple functional areas comes from task-fMRI studies that have demonstrated a speech-specific pattern of response in a small portion of the PT that shares similarities with area Tpt mapped in architectonic work. This functional area, which is often referred to as Spt (Sylvian-posterior-temporal) responds during both speech perception and speech production tasks (e.g., Blank et al., 2002; Buchsbaum et al., 2001; Hickok et al., 2003; Wise et al., 2001). Further, it shares many of the properties associated with other sensorimotor integration areas probed within the monkey parietal cortex, including: i) an insensitivity to the modality in which stimuli are presented (i.e., visually or aurally; Buchsbaum et al., 2005; Corbo & Orban, 2017; Okada & Hickok, 2009), ii) a markedly different response pattern depending on whether it is engaged in sensory or motor processing (Hickok et al., 2009), and iii) a preference for a specific motor effector (i.e., vocal over manual; Pa & Hickok, 2008). These findings have established a role for Spt in the integration of sensorimotor information for the vocal tract (Hickok et al., 2009; 2011), but recent evidence has suggested its role may be more specific to laryngeal control (Hickok, 2016). For instance, tracking a moving sound source by modulating pitch covertly activates Spt more than tracking that sound source by moving the tongue (Isenberg et al., 2012). The importance of Spt for speech production is corroborated by the fact that the distribution of lesions in conduction aphasia, a disorder marked by speech repetition deficits, is concentrated in this area (Buchsbaum et al., 2011).

Area Tpt conspicuously shares many features with Spt, indicating some possible correspondence. Both areas sit in a transitionary patch of cortex that bounds temporal and parietal cortex. The multisensory properties Spt exhibits are matched by the microstructural properties of Tpt (i.e., auditory-like features but lacking hallmarks of auditory cortex). In addition, tracer studies in macaque monkeys show that Tpt is uniquely positioned among auditory areas, with long range connections that can support multisensory integration. For example, this area shows the weakest connectivity to primary auditory cortex among auditory areas, and the strongest connectivity to portions of posterodorsal prefrontal cortex implicated in multisensory processing that lie near the frontal eye fields (i.e., Brodmann area 8, typically dorsal; Hackett, 2015; Chavis & Pandya, 1976; Smiley et al., 2007). Further, some tracer studies have demonstrated connectivity between the sulcal portion of Tpt and areas 44 and 45B of the inferior frontal gyrus (Petrides & Pandya, 1988; 2002; Frey & Petrides, 2014). Coincidentally, Spt also exhibits connectivity to the human homologue of area 44 (Isenberg et al., 2012), which is thought to code motor programs for speech production (Hickok, 2012).

3.1.6 Other evidence for functional dissociations in the PT

Several functional neuroimaging studies testing multiple different functions in the same group of participants have illustrated functional dissociations within the PT as well, posing a challenge to accounts that ascribe single function to PT. For instance, studies have distinguished pitch processing in anterolateral PT from spatial hearing in more posteromedial PT (Warren & Griffiths, 2003). At the same time, spatial hearing has been shown to involve a different portion of the PT than sensorimotor processing for speech in

more posterior area Spt (Isenberg et al., 2012). Another task-fMRI study has shown three functional dissociations within the PT by delineating cortical areas using macroscopic landmarks on structural MRI scans, then treating these areas as regions of interest (Tremblay et al., 2013). This research has implicated the anterolateral section of the PT in purely auditory processing, the anteromedial section in processing speech feedback, and the posterior section in sensorimotor integration (Tremblay et al., 2013). It is not clear, however, whether structural MRI scans can provide adequate resolution for defining cortical areas accurately, and many of the cortical areas in and around the PT have been described not to follow macroscopic landmarks (e.g., Clarke & Morosan, 2012; Fullerton & Pandya, 2007; Morosan et al., 2005a,b; Zachlod et al., 2020).

3.1.7 Connectivity data for mapping functional organization of the PT

Studies attempting to understanding the functions supported by the PT face several challenges. This anatomical area is difficult to define and highly variable across individuals. Its microstructural properties suggest that it contains at least three separate functional areas that are unlikely to respect its anatomical boundaries. At the same time, the precise number of areas within the PT and their spatial configuration remains uncertain because architectonic studies have shown some heterogeneity in their parcellations and have covered slightly different patches of cortex. Though these parcellations of the PT remain incredibly useful, it can be difficult to generalize them from a handful of post-mortem brains to the functional anatomy of participants in task-fMRI with its coarser resolution. Nevertheless, task-fMRI has successfully demonstrated functional dissociations within the

PT. The issue, however, is that individual task-fMRI studies attempting to parcellate the PT cannot possibly test all of the tasks that have been reported to activate this area. As such, it can be difficult to understand how functional dissociations reported within the PT in individual studies map onto each other, particularly because the PT and the cortical fields that underpin it are so highly variable across individuals. Here, we advocate for a more systematic approach of investigating the functional organization of the PT—using its participation in different functional networks to carve it up into areas. Spontaneous activity at rest exhibits a rich structure that reflects the functional architecture of the brain (Seitzman et al., 2020; Smith et al., 2014). Functional connectivity during this resting state can be used to map many different networks in the same participants, networks that traditionally could only be accessed through multiple task-fMRI studies (Smith et al., 2014). Analyzing functional connectivity patterns within a particular region can be used to determine which parts of the region participate in different networks (Eickhoff et al., 2015; 2017; 2018).

Many previous studies have taken the same approach of parcellating a particular brain region based on the functional or structural networks in which it participates (e.g., Eickhoff et al., 2015). However, no study has explicitly attempted to parcellate the PT. While some whole brain parcellations have covered the PT, modeling connectivity across the entire brain reduces the sensitivity of the analysis to more subtle connectivity patterns within regions (e.g., Eickhoff et al., 2015; 2018). Further, despite the evidence from architectonic and task-fMRI studies that the PT contains multiple areas, whole brain parcellations based on functional connectivity have not shown much functional diversity within in (e.g., Joliot et al., 2015). This is curious because connectivity-based parcellations

applied to other regions of the brain have been reported to show good correspondence with many features of architectonic parcellations (e.g., Anwander et al., 2006; Beckmann et al., 2009; Chao et al., 2009; Genon et al., 2018; Gordon et al., 2014; Glasser et al., 2016; Kahnt et al., 2012; Zhang et al., 2014). That connectivity data from neuroimaging *can* be sensitive to cortical areas has been supported by comparisons between invasive and noninvasive methods (structural: Gao et al., 2018; functional: Hori et al., 2020).

3.1.8 Current research

This research aims to elucidate functional organization in the PT by performing a careful connectivity-based parcellation of the area using cluster ensembles that leverage the strengths of different clustering algorithms for discovering different structures in the data. In order to describe a functional organization of the PT that connects to prior task-fMRI work, we defined and focused on the larger cortical zone that has become associated with the PT across functional neuroimaging studies. Because cortical fields underpinning the PT can extend beyond the area's boundaries, focusing on this larger posterior perisylvian (PPS) zone also improved the odds that the full extent of areas within the PT would be captured by the parcellation.

A large amount of effort in the current work was directed at ensuring the production of a high quality parcellation of the PPS zone. At the level of acquisition and preprocessing, we elected to collect multiecho resting state data in a relatively large group of participants (N=136) and relied on a robust framework for denoising that exploits the linear echo-time dependence property of BOLD-signal changes (Kundu et al., 2012; 2013; 2017). Critically, this approach requires minimal spatial smoothing. We then used a

subsampling procedure to evaluate 12 different clustering approaches (i.e., combinations of 4 clustering algorithms and 3 distance metrics) on their ability to produce balanced parcellations that were both reliable and complex (i.e., had more parcels). That is, a subsampling procedure was used to evaluate the internal validity of clustering approaches by measuring how consistently parcels could be assigned across many subsamples. This procedure was repeated across a range of parcellation complexities. The most balanced clustering approaches with the most distinct parcellations were consolidated into cluster ensembles to produce more robust parcellations for each participant. These participant-level ensembles were used to investigate individual variability in optimal parcellation complexity. Group-level ensembles were then created either by consolidating over all participants, or only those participants with the same optimal parcellation complexity, allowing for the evaluation of which parcellation complexity could be fit to the entire group while still retaining the unique features of functional organization (as expressed by optimal model complexity) within-participants. The ensemble approach for fusing participant-level parcellations was evaluated against direct clustering of group-level functional connectivity data by determining which of these methods produced parcellations that could be more accurately used to classify external data. The same classification analysis was extended to several existing atlases to benchmark our parcellations of the PPS zone. In order to understand what functions are associated with the networks that carved up the PPS zone, classifiers that were trained to discriminate between parcels based on whole brain connectivity from the benchmark analysis were tested on meta-analyses of different behaviors. The functions of the parcels themselves were decoded by correlating the

reliability with which a cluster could be assigned to the PPS zone, with the likelihood that different behaviors produced response there.

3.2 Methods

3.2.1 Participants

One hundred and thirty-seven adults (78 female) were recruited for this study approved by the Local Research Ethical Committee at the University of Cambridge. The volunteers were right-handed, native English speakers with no known history of neurological disease. Data from a total of thirty-five of those subjects was published in prior work (Kundu et al., 2013).

3.2.2 MRI data acquisition

Images were acquired on a Siemens Trio 3T MRI Scanner using a 32-channel receive-only head coil (Siemens Medical Solutions) located at the University of Cambridge. Functional images were acquired with a multiecho EPI sequence [repetition time (TR), 2.47 s; flip angle, 78°; matrix size, 64 × 64; in-plane resolution, 3.75 mm; field of view (FOV), 240 mm; 32 oblique slices, alternating slice acquisition slice thickness 3.75 mm with 10% gap; iPAT factor, 3; bandwidth (BW) = 1,698 Hz/pixel; echo time (TE) = 12, 28, 44, and 60 ms]. Subjects were instructed to fixate on a centered crosshair over two functional runs (~12 minutes in total). Anatomical images were acquired using a T1-weighted magnetization prepared rapid gradient echo (MPRAGE) sequence [176 × 240 FOV; 1-mm in-plane resolution; inversion time (TI), 1,100 ms].

3.2.3 Preprocessing resting state data

Multiecho independent component analysis (ME-ICA) was used to preprocess resting state data (Kundu et al., 2012). This method was chosen because it requires no spatial smoothing and can track anatomy more closely than standard ICA methods for denoising resting state data (Kundu et al., 2017). Estimation of functional connectivity is complicated by the inability to discriminate true neuronal activity in the BOLD signal from noise (e.g. physiological noise, motion, etc.). One way to deal with this issue is to model that noise, either by using filtering techniques to estimate it, or by collecting some other data externally and in parallel with image acquisition (i.e., physio signals). Such techniques are susceptible to artifacts, for example by underestimating noise (Power et al., 2012) or removing true signal as (arguably) in the case of global signal regression (Saad et al., 2012). An entirely different approach to dealing with this issue is to exploit a well-known property of BOLD signal. By collecting imaging data at multiple echo times, it is possible to distinguish true BOLD signal from noise by testing for echo-time dependence. This testing can be facilitated by ICA, which can be used to model timecourses within the resting state signal. The resulting components can be tested for echo-time dependence and used to produce more specific estimates of functional connectivity at individual and group levels (Kundu et al., 2012).

Preprocessing was performed with AFNI (Cox, 1996). Six-parameter rigid body motion correction and other parameters of interest for coregistration were estimated from functional images collected with the middle TE. Volumes acquired during steady-state for functional data were removed, and the following volume was used as the EPI image for

further processing. Data were then de-obliqued and coregistered to anatomical scans using twelve-parameter affine transformation with the local Pearson correlated signed cost function. The gray matter of the EPI image was extracted with FAST (Smith et al., 2004) and used as a weight mask for this transformation. All transformations were then concatenated into a single matrix which was used to spatially align the TE datasets after slice-time correction. Following this, anatomical images were deskulled and nonlinearly warped to the 1 mm MNI template using FNIRT (Andersson et al., 2008). The same warp was applied to each of the TE datasets.

The dimensionality of the concatenated mutliecho EPI data was reduced using principal component analysis, and the lower-dimensional dataset was decomposed into independent components in the temporal domain using FastICA (Bingham & Hyvarinen, 2000). The resulting independent component timecourses were fit to the TE-dependence model of BOLD signal changes, as well as the TE-independence model for non-BOLD signal changes using weighted least-squares. Goodness of fit was then computed for each voxel with an F-test that compared the residuals of that fit to the residuals of the null model, which was defined as the sum of the squares of signal changes. This resulted in F-maps that were then averaged and weighted by voxel component weights, defined as partial correlation coefficients normalized with the Fisher transform. For every component, pseudo-F-statistics κ and ρ, corresponding to BOLD-like modulations and initial intensity modulations were then computed from the weighted averages. Functionally related BOLD components are characterized by high κ and low ρ values, which were clearly distinguishable using the elbow method. This method was automated to extract BOLD components by finding the point within the sorted pseudo-F statistics that was most

distant from the highest and lowest endpoints. A more detailed account of the exact preprocessing pipeline employed here is presented by Kundu and colleagues (2013).

3.2.4 Computing group-level and subject-level functional connectivity

Group-level functional connectivity was estimated from resting state data using multiecho independent component regression (ME-ICR), by applying principal component regression to the BOLD-weighted independent components generated with ME-ICA (Kundu et al., 2013). This approach allowed for the estimation of BOLD degrees of freedom, thereby enabling normalization of connectivity estimates, and further accounted for changes in those degrees of freedom due to variable sensitivity within BOLD signal to subject motion. The ME-ICR process entailed generating partial correlation coefficients for each ICA component by variance normalizing the BOLD timeseries data and regressing on the ICA matrix used to decompose the BOLD components. Functional connectivity was estimated at the subject-level by performing Pearson correlations between independent BOLD component coefficients for seed and target voxels and subsequently normalizing the correlation using the Fisher transform. Degrees of freedom for this transform were taken as the number of significant BOLD signal related components generated by ME-ICA. Group-level functional connectivity maps were then generated by performing a one-sample t-test over subject-level functional connectivity maps using 3dttest in AFNI.

3.2.5 Defining the left posterior perisylvian (PPS) zone

The objective of this study was to elucidate functional organization within the broader zone of cortex that functional neuroimaging studies have associated with the PT. Thus, we elected to define a large region of interest (ROI) for parcellation based on the coordinates reported in functional neuroimaging studies that have focused on the PT. This larger ROI is referred to as the posterior perisylvian (PPS) zone to distinguish it from purely anatomical PT.

The meta-analysis for defining the ROI was performed using Neurosynth, a large database that provides a framework for associating information about word frequencies within the abstracts of published neuroimaging studies to the activation foci that they report (Yarkoni et al., 2011). Specifically, we first identified all papers in the neurosynth database that used the phrase "planum temporale" at a minimum rate of 1 out of 1000 words, which prior research has found to be sufficient for removing studies that use a term incidentally (Yarkoni et al., 2011). This threshold identified a total of 85 studies, out of which the least frequent use of the phrase "planum temporale" occurred at a rate of 5 out of 1000 words. To ensure that this inclusion criterion did not significantly affect the area of cortex identified as the PPS zone, we also performed an analysis of studies using the phrase "planum temporale" at a rate of 10 out of 1000 words, which identified 49 studies in the database.

The meta-analysis that was performed for the two frequency thresholds identified voxels that differentially activated for studies that used the phrase "planum temporale" at that frequency compared to those that didn't (i.e., all other studies in the database not

matching the criterion). This analysis proceeded by separating all foci reported in the group of papers that used the phrase “planum temporale” frequently from the foci in the remaining group of papers that did not. Contingency tables were then constructed for these two groups of papers at every voxel of the brain, including the number of times activity was reported, and the number of times the phrase “planum temporale” was used. A two-way chi-square test for independence was then performed over those contingency tables, and the resulting whole-brain map was corrected for multiple comparisons using the false-discovery rate (FDR) threshold of $p < 0.01$. Note that this approach does not measure consistency of activity as typically done in meta-analytic studies, and results in meta-analyses with higher specificity. The importance of couching brain-behavior associations in specificity is discussed at length in the Chapter 2.

Meta-analyses for both study inclusion thresholds are presented below in Figure 2, along with a visualization of a probabilistic map for anatomical PT taken from the Harvard-Oxford atlas (Desikan et al., 2006). Notably, the more stringent study inclusion threshold did not substantially impact the extent of the resulting ROI (blue and yellow outlines in panel A). However, the more liberal study inclusion threshold was selected for ROI definition because it: i) extended further posteriorly, consistent both with the location of area Spt, and with areas of high inter-subject variability in anatomical PT, ii) more clearly entered a portion of the posterior superior temporal sulcus, consistent with the extent of cortical fields within the PT, and iii) had relatively stronger associations throughout PT, and weaker associations within planum polare relative to the more stringent study inclusion threshold (as expressed by the colors of vertices in panel A). Note, the meta-analysis that defined PPS included activity in premotor cortex that was removed from the

ROI that was submitted to the parcellation analysis. This exclusion will be clear in figures going forward. The area is reported here because it will be shown to be an important feature in distinguishing some of the networks that parcellate the PPS zone.

We have focused on parcellating the left PPS zone in the current work because functional organization within the PT appears to be more complex in the left hemisphere. For example, the aforementioned area Spt has been mapped in the left hemisphere. However, in analyses not reported here for the sake of brevity, we have also parcellated a right hemisphere PPS zone using the same methods. Even though this parcellation was performed independently using exactly the same procedure, it was strikingly similar to the left PPS parcellation.

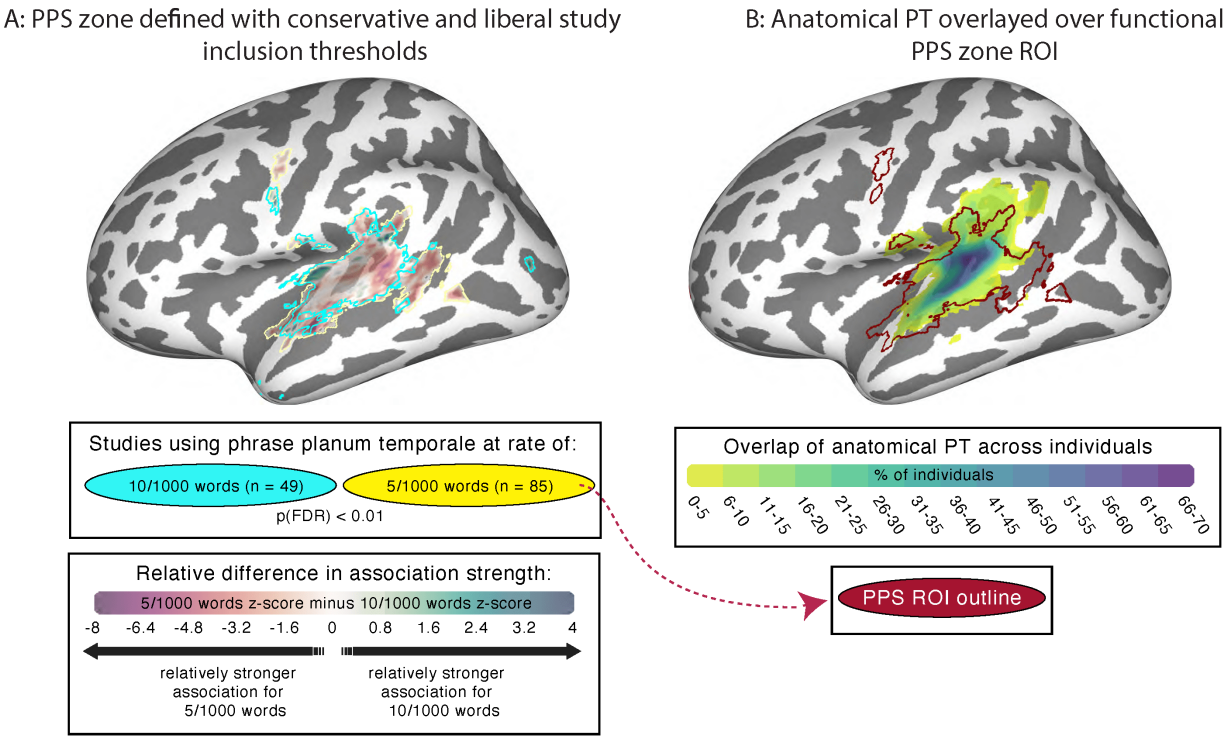

Figure 2: Defining the posterior perisylvian zone (PPS). Two meta-analyses are presented for the phrase "planum temporale" (panel A). The significant area of activity associated with each meta-analysis is outlined in blue and yellow. The blue color corresponds to a study inclusion criteria of 10/1000 words (within an abstract), and the yellow color corresponds to a study inclusion criteria of 5/1000 words. Red and teal colors on the brain represent the relative difference in likelihood of seeing activity for in the two sets of studies. The 5/1000 words meta-analysis is selected as our ROI and its outline is superimposed (in red) over the boundaries of anatomically defined PT across individuals in panel B. Blue colors on the brain in this panel represent greater overlap of anatomical PT boundaries across individuals.

3.2.6 Overview of clustering approaches

In order to generate a reliable parcellation of the PPS zone, 12 different clustering approaches were compared within two different procedures for generating group-level parcellations—either by clustering group-level functional connectivity data directly, or by fusing subject-level parcellations together (fusion process described in section 3.2.9). The 12 clustering approaches were constructed by pairing one of four different clustering

algorithms (i.e., k-means, k-medoids, self-tuning spectral clustering, and Louvain) with one of three different distance metrics (i.e., Euclidean distance, Pearson distance, and eta^2). Each of the 12 clustering approaches were used to generate 29 parcellations of incrementally increasing complexity (i.e., containing from 2 to 30 clusters).

Clustering algorithms were selected primarily based on their popularity for generating parcellation schemes, but also for their ability to identify different kinds of clusters. For instance, k-means and spectral clustering have been described as common approaches to brain parcellation (Eickhoff et al., 2015) and used extensively in prior work (*k-means*: e.g., Anwander et al., 2006; Cha et al., 2017; Crippa et al., 2011; Garcea & Mahon, 2014; Nanetti et al., 2009; Jakab et al., 2012; Shen et al., 2003; Yang et al., 2017; *spectral clustering:* e.g., Craddock et al., 2012; Dillon & Wang, 2020; Ji et al., 2016; Parisot et al., 2016; Shen et al., 2013; Thirion et al., 2006; Wang et al., 2012). To represent exemplar-based clustering approaches, k-medoids, which has some precedent for parcellating brain structures was included as well (e.g., Lefranc et al., 2016; Salehi et al., 2017). Finally, Louvain clustering, a community detection algorithm that has attracted recent attention in graph-based connectivity analysis was also included (Akiki & Abdallah, 2019; Ji et al., 2019; Kurmukov et al., 2017; Messe et al., 2020; Meunier et al., 2009; Nicolini et al., 2017; Taylor et al., 2017; Vanni et al., 2017; Williams et al., 2019). These clustering algorithms have various complementary strengths. While k-means clustering is simple and highly efficient, it is sensitive to noise and struggles to discover non-spherical clusters, clusters of different sizes, and clusters with different densities (e.g., Shukla & Naganna, 2014). Spectral clustering is more computationally expensive, but can find arbitrarily shaped, or touching clusters (e.g., Meila, 2016). K-medoids shares many similarities with k-means, but it is less

efficient for larger datasets; however, it is less sensitive to outliers (i.e., because a mean is easily influenced by extreme values), and selects an exemplar from the data to represent each cluster (e.g., Park & Jun, 2009). The Louvain clustering method shares similarities with spectral clustering but operates by maximizing a modularity measure that selects the number of clusters appropriate for the data based on a resolution parameter that acts as a cluster density threshold (e.g., Traag et al., 2019). Prior studies clustering gene expression data indicate that sometimes the selected distance measure can have just as much impact on the resulting partition as the clustering algorithm itself (e.g., Jaskowiak et al., 2014). Here, we tested two distance measures that tended to produce more discrepant results in this prior work: Euclidean distance and Pearson distance (i.e., $d = 1 - r$). We also tested a third distance measure, eta^2, which is largely the same as Pearson distance but modified to account for scaling and offset when comparing functional connectivity maps (Cohen et al., 2008). As a result, we did not expect this distance measure to produce substantially different results.

3.2.7 Implementation of clustering approaches

All clustering approaches were implemented in MATLAB. K-means clustering using the k-means++ algorithm (Arthur & Vassilvitskii, 2007) was performed using the native statistics and machine learning toolbox in MATLAB. Clustering began with the random selection of a voxel in the PPS zone, which was treated as the first centroid. The distance between the data vector associated with that voxel and the vectors associated with each other voxel in the PPS zone was then calculated. All subsequent centroids were chosen at random using a weighted probability distribution proportional to the distance between a given voxel's

vector of data and the vector of the closest previously chosen centroid. This procedure was repeated until all centroids were chosen. Clustering was repeated 500 times, each time initializing the first centroid in a new position. The clustering solution that minimized the within-cluster sums of point-to-centroid distances was retained as the final clustering solution. For k-means, squared Euclidean distance was used rather than Euclidean distance, in which case each centroid was defined by the mean of the points contained within that particular cluster. Pearson distance for k-means was computed as one minus the sample correlation between points. Centroids were determined by the component-wise mean of the points within that cluster, after centering and normalizing those points to zero mean and unit standard deviation. A distance measure for eta^2 was computed using in-house scripts, taking one minus the estimate for eta^2. This distance was operationalized within the context of k-means in an identical fashion to pearson distance.

Self-tuning spectral clustering was implemented as the spectral clustering algorithm, using the original scripts from the paper proposing this technique (Zelnik-Manor & Perona, 2005). Two features of this approach make it "self-tuning": i) the optimal number of clusters within the data is automatically determined based on an analysis of the eigenvalues of the affinity matrix, and ii) the affinity matrix is locally scaled, allowing it to be driven by local statistics within neighborhoods of points. We do not select the optimal number of clusters using this self-tuning procedure and instead rely on a measure of reliability across clusterings of subsamples. However, locally scaling the affinity matrix eliminates the trial-and-error process associated with selecting a single scaling parameter in conventional spectral clustering. At the same time, *K*, a parameter for determining the size of the neighborhood over which local scaling is implemented still requires some

minimal tuning. Our approach to this involved a parameter search for K=5…1000. The selected value for K was based on a qualitative inspection of preliminary clustering results for the PPS zone. We observed that very low K values (i.e., small neighborhood sizes) tended to produce a large number of singleton clusters, and one very large cluster. However, after reaching a certain neighborhood size, parcellations would stabilize and appear largely indistinguishable no matter how much larger the neighborhood size became. Thus, we chose the lowest K after which the parcellations appeared to stabilize under a varying number of predetermined clusters. This ended up being K = 100 for all distance metrics. Otherwise, the procedure for self-tuning spectral clustering was identical to more common spectral clustering techniques (Zelnik-Manor & Perona, 2005).

The Pattern Recognition and Machine Learning Toolbox (Chen, 2020) was used to implement k-medoids, which was realized using the Partitioning Around Medoids algorithm (e.g., Kaufman and Rousseeuw, 1990). This procedure closely mirrored that of the k-means++ algorithm. First, a randomly chosen set of medoids was selected based on the number of requested clusters. Each data point was then associated to its closest medoid. Then, for each point within a medoid, the average dissimilarity between that point and all other points in the medoid was computed, and the point with the lowest dissimilarity replaced the data point that was originally selected as a medoid. This process of reassigning medoids based on dissimilarity was repeated until dissimilarity could not be decreased.

Finally, Louvain clustering (Blondel et al., 2008), a type of hierarchical clustering, was implemented using the "generalized Louvain" MATLAB code (Mucha et al., 2010; Lucas et al., 2019). This clustering entailed creating a modularity matrix that defines the quality

of a partition based on the placement of any pair of samples into the same cluster or community, from a weighted adjacency matrix that is constructed using a similarity measure. In conventional Louvain clustering, a partition is generated in two steps, first by identifying small clusters or communities by optimizing modularity locally, then by combining nodes belonging to the same community, so as to generate a new network. This process is repeated until modularity cannot be maximized. The generalized Louvain clustering process is only different in so far as it operates over a precomputed modularity, rather than an adjacency matrix. Consistent with prior brain parcellation work relying on modularity maximization algorithms, we used a fixed resolution parameter of 1 (Tian & Zalesky, 2018). Therefore, the algorithm automatically detected the number of clusters putatively supported by the data. Because the resolution parameter effectively replaces the choice of clusters made a priori, there was no reasonable way to estimate parcellations of varying complexity (i.e., number of clusters) using this method, which distinguished it from the others above.

3.2.8 Dimensionality reduction

For parcellating group-level functional connectivity, whole-brain functional connectivity maps computed for each voxel within the PPS were downsampled to 2mm, masked to remove all voxels outside the brain, then concatenated into a single matrix. Because brain voxels are not independent units, and neighboring voxels exhibit similar functional connectivity patterns, these whole-brain functional connectivity maps contained a large amount of redundant information. To reduce the feature-space of this connectivity matrix, Principal Component Analysis (PCA) was applied prior to clustering. The connectivity

matrix was mean centered, and singular value decomposition (SVD) was used to estimate its principal components. For retaining components, a permutation-based procedure was employed to determine at which point components become statistically indistinguishable from noise (Linting et al., 2011; McIntosh & Lobaugh, 2004). To establish a null distribution, the functional connectivity matrix was sampled without replacement a total of 10,000 times. Critically, in each iteration of the permutation procedure, the samples within each feature were permuted independently, thereby breaking the correlational structure between features. SVD was then performed on the resampled data and used to compute the cumulative variance explained by the addition of each successive component into the PC model (i.e., the sum of eigenvalues). All PCs after which the cumulative variance explained in the original data was lower than the cumulative variance explained in more than 0.1% of the permuted datasets were discarded from further analysis. This resulted in the retention of 81 significant components ($p < 0.001$) explaining a total of 96% of the variance in connectivity to the PPS zone while reducing the feature-space by 228,372 dimensions. To test whether the PCA procedure impacted clustering results we performed k-means clustering performed over the component scores to clustering performed over the "raw" functional connectivity maps (analysis not reported here for the sake of brevity). There were no visible qualitative differences in the results, but the dimensionality reduction step significantly reduced the computational costs associated with generating cluster ensembles.

Parcellating subject-level functional connectivity did not require a dimensionality reduction step. This is because each BOLD component modeled with ME-ICA carries with it coefficients in the spatial domain that fundamentally represent the loading of each brain

voxel onto the set of BOLD-weighted independent components, or timeseries, modeled within a subject. Correlation between these spatial BOLD component coefficients for a pair of voxels determines their functional connectivity. Thus, we submitted these coefficients directly to each clustering approach. It is worth noting that although this temporal ICA procedure departs from more conventional spatial ICA for the definition of resting state networks (RSNs), it has the positive effect of allowing RSNs to overlap to a greater extent because independence is not being enforced in the spatial domain. At the same time, these temporal ICA-based RSNs are still functionally distinct because they are characterized by independent timecourses. Prior research has avoided using temporal-ICA in this manner because it requires collecting high quality data (i.e., because there are many more voxels than timepoints) and is more computationally expensive to handle (i.e., the mixing matrix becomes unwieldy by modeling inter-voxel covariance; McKeown et al., 1998; Calhoun et al., 2001; Smith et al., 2012). Nevertheless, prior work has shown high similarity between RSNs generated with temporal and spatial ICA (Smith et al., 2012) and the current research relies on multi-echo data, which substantially improves data quality (Kundu et al., 2012; 2013; 2017).

3.2.9 Consensus clustering for creating and fusing cluster ensembles

The framework of consensus clustering, an ensemble method that has received substantial attention for class discovery in gene-expression data (e.g., Herschkowitz et al., 2007; Kiselev et al., 2017; Lancichinetti et al., 2012; Monti et al., 2003; Sturm et al., 2012; Swift et al., 2004; Wang et al., 2014), was used to both evaluate the reliability of different clustering

approaches and to fuse together cluster ensembles. The principle of this method as it is most commonly implemented is to use a resampling technique to compute consensus (interchangeably used with agreement and consistency) across multiple clusterings (Monti et al., 2003). Although bootstrapping has been used in consensus clustering, it can distort the compactness of the data and therefore subsampling is often the preferred resampling technique (Kim & Lee, 2007; Monti et al., 2003). Consensus is defined as the proportion of times that a pair of samples from the data are placed into the same cluster when they are subsampled together (e.g., Monti et al., 2003). This measure is computed for each pair of samples in the data to form a symmetric similarity matrix that contains a consensus solution which can be retrieved using similarity-based clustering algorithms (e.g., Fred & Jain, 2005; Monti et al., 2003; Strehl & Ghosh, 2002; Swift et al., 2004). A consensus matrix can also be computed over subsamples clustered using multiple different clustering approaches, thereby quantifying agreement across different methods (e.g., Boongoen & Iam-On, 2018; Simpson et al., 2010). Thus, the consensus matrix serves as the 'consensus function' that maps a cluster ensemble onto a single solution. Combining clustering solutions can yield results that are more stable, less sensitive to noise, and which take advantage of the strengths of different clustering approaches for finding different kinds of clusters (see clustering approaches in methods). As one example, differences in the way clustering approaches handle outliers can be mitigated (Simpson et al., 2010). Although consensus serves to consolidate clustering solutions, it is also a rich source of information about internal reliability. The consistency with which a clustering algorithm can yield the same clusters repeatedly is a good indicator of its internal validity (e.g., Monti et al., 2003; John et al., 2020; Senbabaoglu et al., 2014; Swift et al., 2004). Most studies have used

consensus clustering in the context of a single clustering approach, generating a more robust cluster solution while using some measure of stability, or reliability, computed over consensus matrices in order to select an optimal number of clusters that can be fit to the data (e.g., Monti et al., 2003; John et al., 2020; Senbabaoglu et al., 2014; Simpson et al., 2010; Swift et al., 2004).

Here, we performed consensus clustering to evaluate clustering approaches, select solutions with the optimal number of clusters, and fuse cluster ensembles (for a detailed schematic of the analysis pipeline, see Figure 3). Finding the optimal number of clusters within solutions was part of our strategy for evaluating the 12 different clustering approaches detailed in the preceding sections. This evaluation determined which clustering approaches were used to form cluster ensembles. Some analyses involved forming cluster ensembles over parcellations of group-level data, while others involved larger ensembles that contained many participant-level parcellations. For all clustering approaches that were employed, random subsamples of 60% of the voxels in the PPS zone were selected thousands of times, and a parcellation was generated each time. Consensus was computed over these thousands of parcellations to generate a single consensus matrix. This procedure was repeated 29 times for each predetermined level of parcellation complexity that was tested (i.e., number of clusters in the parcellation; 2-30). An optimal parcellation complexity was then selected based on the internal validity measures computed for each of these consensus matrices. Optimal parcellation complexities and their associated validity measures were identified for all clustering approaches and used to guide the selection of *which* clustering approaches would be submitted into cluster ensembles. Independent cluster ensembles were formed at each level of parcellation complexity. Cluster ensembles

were fused by averaging together the original consensus matrices used to estimate the internal validity of each parcellation within the ensemble. Eventually, a single ensemble parcellation was chosen from the set of ensembles that represented different levels of parcellation complexity. Overall, two such sets of ensembles were created—one containing parcellations of group-level functional connectivity, and one containing parcellations of RSNs from individual participants. Hierarchical clustering with average linkage was used to extract the final parcellations embedded within the consensus matrices generated for cluster ensembles (e.g., Fred & Jain, 2005; Goder & Filkov, 2008; Monti et al., 2003; Strehl & Ghosh, 2002; Swift et al., 2004). The number of clusters submitted to hierarchical clustering was equal to the parcellation complexity associated with the ensemble. The only exception to this was for Louvain clustering, which automatically selected the appropriate number of clusters for each subsample. As a result, all consensus matrices generated using Louvain clustering were themselves clustered with this approach.

3.2.10 Measuring reliability

Internal validity of consensus matrices was estimated using the proportion of ambiguously clustered pairs (PAC) measure (Senbabaoglu et al., 2014). PAC is based on capturing the flatness of the mid-segment of a cumulative distribution function (CDF) curve generated for a consensus matrix by taking the fraction of sample pairs with consensus indices falling within some sub-interval (see steps 1 and 2 in panel E of Figure 3). The intuition behind this measure is that a clustering that is particularly reliable will be associated with a consensus matrix that is comprised largely of ones and zeros, reflecting the consistency with which sample pairs are either placed in the same cluster, or in different clusters,

across subsamples. This kind of behavior within a consensus matrix is characterized by a CDF curve that has two large peaks around consensus values of one and zero, and a flat middle segment. Here, we use the common interval of [0.1,0.9] to compute PAC for each consensus matrix. We chose PAC because work has shown that this approach is more robust for determining the number of clusters in well-characterized datasets compared to other methods like qualitatively inspecting the "blockedness" of consensus heatmaps (i.e., heatmaps of consensus matrices), or looking for an elbow in the change of area under the CDF curve of consensus matrices when iteratively increasing the number of clusters (Senbabaoglu et al., 2014). Note that hereafter when we refer to the "reliability" of a parcellation, we refer to its PAC.

One issue with PAC that has only recently received attention is that it can show bias towards cluster models with a large number of clusters (John et al., 2020). In a typical PAC curve calculated over the parcellation complexities we investigated, PAC would dramatically increase with complexity, then at some point begin to decrease, often monotonically. This behavior occurs because for very dense partitions with many clusters, consensus matrices are almost uniformly comprised of zeros, reflecting no consistent assignment of clusters across subsamples. The CDF curves for such matrices show a large peak at zero, sometimes a small peak around 1, and a relatively flat middle segment. The large amount of agreement about samples not belonging together artificially reduces PAC. We developed a method to identify consensus matrices that showed overall poor agreement despite increasing PAC by inspecting the extent to which consensus matrices were both bimodal and multimodal. The dip statistic was used to measure the multimodality of a consensus matrix by computing the difference between an empirical

distribution function and the unimodal distribution function that minimizes this difference (Hartigan & Hartigan, 1985). A p-value for the dip statistic, reflecting whether a distribution is likely to be multimodal as opposed unimodal, was provided by comparing the obtained dip value with dip values calculated for 10,000 bootstrapped samples of the same size from a uniform distribution. The bimodal coefficient (BC) was used to measure bimodality by weighting the skewness of a consensus distribution against its excess kurtosis, scoring highly those distributions that had high absolute values of skewness and low values of kurtosis (Pfister et al., 2013). The dip statistic was explicitly used to screen implausibly low PAC values while the BC was used to confirm that a peak in PAC corresponded to a peak in bimodality, and therefore was not driven by the PAC bias towards more complex parcellations.

3.2.11 Selecting optimal parcellation complexities

An automated procedure was developed for identifying locally optimal parcellation complexities. For a visual description of the procedure, see Figure 3 (panel E). Initially, we intended to use only the p-values associated with each dip curve to exclude unimodal consensus distributions that exhibited low PAC, selecting a single parcellation complexity from the remaining solutions. However, some consensus distributions marked by implausibly low PAC have slight, and often multiple, small peaks around consensus values other than zero. These consensus distributions could generate 'significantly' multimodal distributions but still reflected poor agreement across parcel assignments. Inspection of the dip curves suggested that identification of a plateau in the dip curve was effective for excluding these consensus distributions from consideration. Further, the plateau in the

curve usually aligned with the beginning of the often-monotonic decrease in PAC observed across higher parcellation complexities. The dip plateau was recognized automatically by taking the absolute value of successive differences along the dip curve, which centered it around zero within some tolerance. To determine a good tolerance within which this plateau could be marked, the noise variance of the curve was estimated using the EVAR algorithm (Garcia, 2010), and then multiplied by a tunable tolerance threshold value. If a particular tolerance threshold did not capture any points on the curve, it was doubled iteratively until a plateau could be discovered. Typically, the plateau in dip occurred either at the same time or before dip values started to become insignificant. However, this was not always the case. As such, any parcellations with insignificant dip values (i.e., likely unimodal consensus matrices) were excluded from consideration for local optima. Within the range of solutions not eliminated by the dip curve, a set of potential optimal parcellations was compiled by identifying those parcellations that constituted a valley on the on the PAC curve (i.e., were flanked by higher PAC values on both sides), and a peak on the BC curve (i.e., were flanked by lower BC values on both sides). The locally optimal parcellation was defined as the one that had: i) the highest complexity while remaining within 1.5 standard deviations of the lowest PAC value on the PAC curve, and ii) the highest BC value on the BC curve. In cases where the PAC and BC curves that were trimmed to exclude implausibly low PAC values showed no valleys or peaks, all of the solutions along this curve were considered.

The automated steps of identifying dip plateaus and then optimal solutions were compared against qualitative methods for performing these steps (i.e., eyeing a plateau in dip, and eyeing a valley in PAC that was “reasonably” close to the lowest PAC value). The

selection of a good tolerance threshold for identifying the dip plateau was directly based on the comparison between the automated method as performed over a range of tolerance thresholds, and the qualitatively identified plateaus in 25 participants. A single tolerance threshold was selected based on how closely it approximated the qualitative dip plateaus in participants across all clustering approaches (correlation coefficients varied between 0.6 and 0.83; $p < 0.05$; intra-class correlation coefficients and kappa values were also computed and showed high consistency). Using this tuned tolerance value, optimal solutions were then automatically identified and compared to qualitatively identified optimal solutions. Correlations between these two methods for identifying optimal solutions showed high consistency (correlation coefficients varied between 0.54 and 0.82; $p < 0.05$; intra-class correlation coefficients and kappa values were also computed and showed high consistency). After being applied to all participants, the selected tolerance threshold was inspected and was found to generally align with qualitative determinations in the dip plateau. The optimal models selected automatically also showed good consistency with qualitative inspections.

3.2.12 Comparison between clustering approaches based on optimal models

Clustering approaches were evaluated independently for parcellations generated from group-level and participant-level functional data. This was done to ensure that the selected approaches were tailored to the structure of each type of data. Due to the considerable computational expense of performing the subsampling procedure we describe 348 times (12 clustering approaches by 29 parcellation complexities) for each of 137 subjects, only

2000 subsamples of the data were taken when generating participant-level parcellations compared to 5000 subsamples for generating group-level parcellations. In addition, the number participants over which the 12 clustering approaches were evaluated was pared down to 25, still resulting in over 17 million performed parcellations. The two clustering approaches that were selected were then carried out over the remaining 112 participants, putting the total number of parcellations performed at just over 30 million. In comparison, only roughly 1.3 million parcellations were generated for group-level functional connectivity data. Because parcellating participant-level data revealed one clustering algorithm performed especially poorly (k-medoids), we excluded this algorithm from evaluation in group-level data. For a visual description of how cluster evaluation differed between group and participant-level analyses see panels C and D of Figure 3.

Selection of clustering approaches followed slightly different procedures for group-level and participant-level data. Recall that each clustering approach was used to generate parcellations of varying complexity. For group-level data, an optimal parcellation complexity was selected for each clustering approach by evaluating the internal validity of parcellations as a function of increasing parcellation complexity (see panels E and F in Figure 3 for visualization). The selection of optimal parcellations was performed on the basis that parcellations should be relatively complex, but still show high reliability. The reliability of the chosen models was then compared across clustering approaches to identify those approaches that showed higher internal validity. Separately, pairwise similarities were measured between the parcellations generated by each pair of clustering approaches, both across all parcellation complexities, and at each level of parcellation complexity (see panels F in Figure 3 for visualization). Two clustering approaches were

then selected on the basis that they both produced relatively complex but highly reliable parcellation models that were also highly dissimilar. Dissimilarity was maximized to encourage the formation of a cluster ensemble that contained parcellations tuned to different patterns in the data. Only two approaches were selected to form ensembles due to the computational considerations described previously. When clustering participant-level data, the same process of selecting an optimal parcellation complexity was carried out within-participants (see panels G in Figure 3 for visualization). The resulting distribution of optimal parcellation model complexities across participants, as well as the distribution of reliabilities associated with those models, was used to select two clustering approaches. The selection process for participant-level data was similar, weighting the reliability of optimal models produced by various clustering approaches against their ability to generate different parcellation schemes, the latter of which was inferred from the distributions of optimal model complexities (see panels G in Figure 3 for visualization). Thus, part of the clustering approach for participant-level data involved analyzing the individual variability in functional anatomy expressed as the largest number of parcels that can be *reliably* investigated within an individual. Selecting optimal models was couched in the idea of discovering the most complex parcellation scheme that could be reasonably investigated further.

3.2.13 Adjusting the complexity of participant-level parcellations based on optimal model reliabilities in the group

The reliability of locally optimal parcellation models across participants was used to make small refinements to the selection of locally optimal models within subjects. Any

participant-level parcellation that had a reliability which was 2 standard deviations above the mean for locally optimal parcellations, received a bump up in parcellation complexity. That is, the number of parcels within that parcellation was increased until reliability was within 2 standard deviations of the mean. On the other hand, if the reliability of a participant-level parcellation was 2 standard deviations below the mean, complexity was decreased until it was within that range. There are two motivations for this parcellation complexity adjustment procedure. First, it allowed for the possibility of null parcellation models (i.e., when no parcellation complexity generated a reliability value within 2 standard deviations of the mean). Second, the procedure for selecting locally optimal parcellation models involved searching for peaks in reliability when reliability was plotted as a function of parcellation complexity. However, it is possible for a parcellation to be very close to a local peak in reliability, without actually constituting a peak itself. This parcellation would not be selected as a locally optimal model, even though it may be a perfectly reasonable model to investigate. The inspection of optimal models across participants that was used to select clustering approaches, as described in the preceding section, was carried out over the distributions that were group-adjusted in this way.

3.2.14 Selecting optimal parcellation complexities at the level of cluster ensembles

Cluster ensembles were generated to combine clustering approaches at each level of parcellation complexity. In the group-level analysis where group-level functional data was clustered, each ensemble contained only two parcellations, one for each clustering approach. In the case of the participant-level analysis, each ensemble contained two

parcellations for *each* participant in the dataset. This allowed us to fuse parcellations from individuals into a group-level parcellation that could be compared directly to the parcellation from the group-level analysis. To fuse each cluster ensemble, the consensus matrices associated with the parcellations within the ensemble were averaged together. Finally, an optimal ensemble complexity was selected separately for group and participant-level analyses. Selecting an optimal ensemble complexity for the group-level analyses was straightforward, involving the application of the previously described method of selecting optimal models, but applied over the consensus matrices that fused parcellations from different clustering approaches. Thus, this selection process followed the same principle of choosing a solution that was relatively complex, but still highly reliable. However, selecting an optimal ensemble complexity for the participant-level analyses followed a more involved procedure. This is because our measure of reliability could not resolve variability in consensus introduced by consolidating parcellations across both participants and clustering approaches. To reduce this variability, we deployed the same procedure used to identify optimal ensemble complexity in the group-level analysis, but on each participant in the dataset. In this way, each participant had a set of cluster ensembles that consolidated clustering approaches at each level of complexity, allowing us to identify optimal ensemble complexities for each participant. This was immensely useful, providing a range of complexities that could be reasonably investigated in the superordinate set of ensembles that consolidated parcellations across both clustering approaches and participants. The issue of selecting an optimal complexity within this range could be conceptualized as selecting a single complexity value that would be a good fit to the participants as a group. To determine this complexity value, we created a third set of ensembles for the participant-

level analysis that collapsed clustering approaches only across those participants for whom each level of complexity was determined to be optimal. The parcellations contained within these ensembles were compared to the parcellations contained in ensembles that collapsed across all participants. This comparison was made at each level of parcellation complexity to determine which complexity resulted in a group-level parcellation that reasonably reflected the unique parcellations of individuals. A detailed visualization of the strategies used to select optimal ensemble complexities can be found in Figure 3 (panels I and H).

3.2.15 Classification of external coactivation data using parcellation schemes

A classification analysis was used to test whether a cluster ensemble of participant-level parcellations could better predict external data than parcellations generated directly from group-level functional connectivity data. Voxels in the PPS zone were classified according to these different parcellation schemes, using their pattern of coactivation across the brain. Coactivation refers to the likelihood that two voxels will be found to activate together across published functional neuroimaging studies. Critically, coactivation data is complimentary to functional connectivity (as well as structural connectivity) because it captures networks that process information irrespective of task conditions or specific neural processes. However, prior work has established exceptionally strong similarity between functional connectivity and coactivation networks (e.g., Di et al., 2013). Thus, we expected classification accuracy to be high.

This classification analysis was also extended to several popular atlases, masked to exclude voxels outside the PPS zone, in order to benchmark the parcellations that we have

generated. Multiple modalities were represented by the atlases. First, a purely functional atlas was selected (atlas of intrinsic connectivity of homotopic areas, or AICHA; Joliot et al., 2015). Second, a purely anatomical atlas (Harvard-Oxford atlas; Desikan et al., 2006). Third, a multimodal atlas that combined functional and structural data (Human Connectome Project's multimodal atlas; Glasser et al., 2016).

Coactivation maps were generated for each voxel in the PPS zone, using roughly the same procedure involved in defining the PPS zone ROI. That is, for a given seed voxel, all studies in Neurosynth that reported activity within 4mm of that seed were separated from studies that did not. For the two groups of studies, contingency tables were then constructed at every voxel of the brain, populated by frequencies with which activity was reported in that particular voxel, and the number of times the studies reporting activity in that voxel also happened to report activity in the seed region (i.e., 4mm sphere drawn over seed PPS voxel). A two-way chi-square test for independence was performed over these contingency tables to identify voxels that coactivated with the seed. The parameter of 4mm was selected based on visual inspection of coactivation maps. Coactivation for smaller spheres often generated sparse or empty maps. To facilitate comparison of coactivation and functional connectivity maps for voxels in the PPS zone, functional data were downsampled to 2mm MNI space.

The LIBSVM library (Chang & Lin, 2012) was used to perform linear c-support vector classification of coactivation data according to each of the aforementioned atlases, and the parcellations of the PPS zone that we have generated. Each feature in the coactivation matrix was scaled between -1 and 1, and support vector class weights were set for each class inversely proportional to class size. A one-versus-one approach was

implemented for multi-class classification, such that a different model was trained to separate each pair of classes. Stratified 10-fold cross-validation was used to estimate both accuracy and the Matthews correlation coefficient, a variant of the correlation coefficient designed to deal with class imbalances and sensitive to both true negative and true positive rates (Chicco & Jurman, 2020; Matthews, 1975). The Matthews correlation coefficient varies between -1 and 1, with 1 representing perfect agreement between predictions and the target, -1 representing perfect disagreement, and 0 representing an average random prediction. Performance measures were compared using macro-averages. Cross-validation was repeated 20 times to ensure performance was unrelated to the random partitioning of the dataset into train and test sets.

3.2.16 Decoding parcels and mapping network differences

Two approaches were taken to decode the final parcellation scheme that we focus on. In order to decode the functions of networks attached to each parcel, the classifiers trained to learn the mapping between coactivation maps and parcels were tested on the meta-analysis of each phrase contained in the Neurosynth database (Yarkoni et al., 2011). In other words, the same procedure from section 3.2.15 was employed, but the test data set (i.e., coactivation maps for voxels) was replaced with a batch of meta-analyses. In order to capture the consistency with which each meta-analysis was classified, the procedure was repeated 1000 times, each time training the classifier on newly formed folds. In order to constrain the number of meta-analyses that were tested (Neurosynth contains 1,335 terms for which a meta-analysis can be generated), only those meta-analyses that generated a mean z-score of 0.8 or above within the left PPS zone were considered (leaving 165 terms).

Each meta-analysis was generated using the same procedure described for defining the PPS zone meta-analytically and was FDR-corrected at $p < 0.001$.

Decoding the functions of the individual parcels proceeded by taking all meta-analyses in Neurosynth, masking out regions outside the PPS zone, then correlating the resulting maps with item-to-cluster consensus maps. Item-to-cluster consensus maps were generated separately for each parcel in the parcellation scheme and represent the mean consensus between any voxel in the PPS zone and all other voxels within a particular parcel. Thus, a parcel was associated with a cognitive process if the likelihood of seeing activity in the PPS zone was related to the reliability with which voxels could be assigned to that parcel.

3.2.17 Surface space projections

In order to facilitate visualization of parcellations and connectivity maps, volumetric data in MNI152 space was transformed into the fsaverage surface coordinate system using a highly accurate precomputed transformation matrix based on a registration fusion approach using advanced normalization tools (Wu et al., 2018). Because the data were being mapped onto a lower dimensional space, a single surface vertex on the boundary of two binary parcels from a parcellation could be associated with voxels belonging to multiple clusters in volume space. Where this occurred, the value of a surface vertex was determined based on the cluster that most frequently mapped onto it in volume space.

Figure 3: Outline of group-level and subject-level parcellation analysis

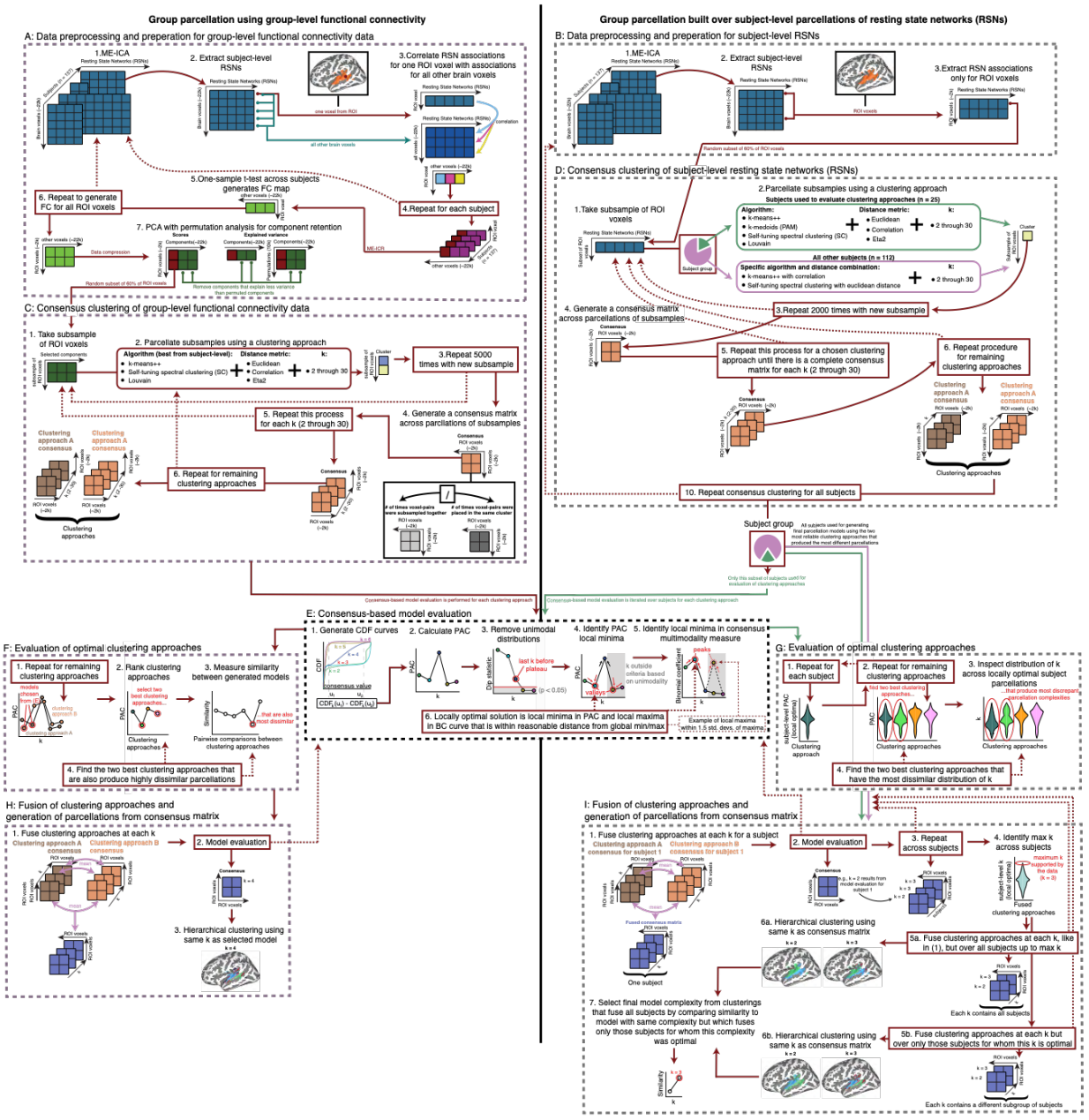

Figure 3: Outline of group-level and subject-level parcellation analysis. Parcellation over group-level functional connectivity data is presented on the left, and parcellation of subject-level functional connectivity data is presented on the right. Panels marked as A and B represent the preprocessing steps prior to clustering (but after ME-ICA). Panels C and D represent the consensus clustering procedure. Panel E represents how measures of reliability over consensus matrices are used to select the model with the optimal number of clusters (i.e., parcellation complexity). Panels F and G represent how different clustering

approaches were then selected to form cluster ensembles based on their ability to produce relatively complex but still highly reliable models. Panels H and I represent how the most reliable clustering approaches that generated the most dissimilar parcellations were fused, evaluated, and clustered to produce the final parcellations.

3.3 Results

In the preceding sections we described two methods for generating a group-level parcellation, either by consolidating participant-level parcellations, or by clustering over group-level data. Because generating participant-level parcellations involved repeating the same core procedure across many participants, we begin by describing the group-level parcellations.

3.3.1 Parcellating group-level functional connectivity using different clustering approaches

The process of generating a single parcellation from group-level functional connectivity data began with an evaluation of which clustering approaches would be used to create cluster ensembles. This evaluation involved selecting optimal parcellation models for each clustering approach from many levels of parcellation complexity, then comparing those models' reliability across clustering approaches. The process and result of model selection for clustering approaches is presented in Figure 4. Comparing the distributions of values within each consensus matrix to PAC curves clearly shows—across all clustering approaches—the PAC bias towards larger parcellation complexities (c.f., PAC curves in third column of panels and consensus distributions in first column of panels in Figure 4). That is, PAC scores can often improve at higher complexities, despite dwindling agreement

within the consensus matrices to which these scores correspond. This lack of agreement is reflected in consensus distributions that are more unimodal, peaking only around zero and reflecting growing agreement that no pairs of voxels should be placed in the same parcel. The same lack of agreement is reflected in cumulative distribution functions, which show curves that have slightly less flat middle segments, weak inflections around consensus values on one, and strong inflections around consensus values of zero (Figure 4, second column across panels).

PAC curves overlaid on dip and BC curves illustrate that BC more closely mirrored trends in PAC (Figure 4, fourth column across panels). In particular, dip curves had the tendency to plateau at relatively early parcellation complexities. Within the range of plateauing dip values, both PAC and BC scores fluctuated in a way that appeared related. PAC and BC curves were correlated to illustrate that identifying the plateau in dip separated portions of the PAC curve that correlated positively with BC, and those that correlated inversely with BC (Figure 4; panel J). That is, prior to the plateau in dip, as the proportion of ambiguously clustered pairs within a consensus matrix decreased, the bimodality of the consensus distribution increased. This indicated alignment between the two measures about the reliability of parcellations. However, after the dip plateau, PAC continued to decrease, while the BC decreased as well. Thus, PAC reflects the bimodality of the consensus distribution only up to a certain point of parcellation complexity. We advocate for an approach that combines BC and/or PAC with the dip statistic because BC can show sharp inflections at higher complexities, despite plateauing dip values (e.g., Figure 4, panel E). Dip plateaus roughly aligned with the point at which consensus distributions became unimodal, at least for the k-means algorithm (Figure 4, third column

for panels A-C). Spectral clustering tended to produce consensus matrices that were significantly multimodal at higher complexities relative to k-means (Figure 4, third column for panels D-F). However, the plateau in dip curves for spectral clustering still aligned with the point at which PAC began to show bias towards larger complexities (c.f., column 3 for panels D-F in Figure 4). Thus, identifying the dip plateau was more effective for eliminating implausibly low PAC values than using the p-value associated with the dip statistic.

Our method for selecting a set of candidate locally optimal parcellations involved identifying valleys in PAC that corresponded with peaks in BC and preceded the plateau in the dip curve. However, sometimes these potential solutions all showed relatively low reliability (e.g., Figure 4, panels A, B, column 4). In such cases, the set of potential solutions was shifted to the batch of parcellation complexities that preceded the first (unreliable) valley/peak. This batch of potential solutions always contained the most reliable parcellation, ensuring that a highly reliable model could be selected. Inspecting CDF curves qualitatively for relative flatness in midsegment (which is what the PAC captures) would have converged on the selected solutions (c.f., columns 2 and 4, Figure 4).

Overall, most clustering approaches showed the same locally optimal parcellation complexity of 5 parcels. The exception to this was k-means using Euclidean distance, for which 6 parcels was locally optimal, and spectral clustering using Pearson distance, for which 4 parcels was locally optimal. We found that although k-means clustering with Euclidean distance produced highly reliable parcellations at high complexities, this approach often generated poorly organized parcellation schemes at those higher complexities (Figure 4, panel C, column 5). Louvain clustering generated parcellations with 2 and 13 parcels depending on which similarity measure was used, however, only the

correlation coefficient produced a reasonable consensus matrix, showing evidence of 2 parcels (Figure 4, panels G-I). At the same time, other clustering approaches all produced a 2-parcel solution that looked identical but was more reliably assigned across subsamples. In the case of Louvain clustering using eta^2, PAC for the only model was low, but the CDF curve and consensus distributions indicated PAC was being overestimated. It is likely that Louvain clustering with Eta^2 and Euclidean distance was less reliable because the number of clusters assigned to each subsample of the data were very high, making it less likely that any two voxels were assigned to the same cluster.

The organization of parcellations that were selected as optimal and potentially optimal for each clustering approach showed consistency. The 2-cluster parcellation scheme appeared to be identical across clustering approaches, broadly separating everything dorsal and ventral to the lateral superior temporal gyrus (STG; see Figure 4, column 5 across panels). However, this parcellation scheme also hinted at greater organizational complexity. For instance, a dorsal area in the PPS zone intruding into the parietal operculum (PO) was associated with the ventral PPS cluster despite being dorsal to lateral STG. Likewise, the most posteromedial segment of the PPS zone that encroached on the supramarginal gyrus (SMG) was also associated with the ventral cluster despite being a dorsal area. Across all clustering approaches, the area of the PPS that was most consistently isolated as its own parcel was auditory cortex. Clustering approaches showed several other similarities with respect to parcel assignment. This included a tendency for a single parcel to combine portions of PO, posterior PT, usually temporooccipital MTG (toMTG) and rarely planum polare (PP; e.g., k=5 in panels A,E,F; k=4 in panels C,D; k=7 in panel A; k=10 in panel A,I). The association of the PP with this subnetwork was exclusively found by k-

means clustering using Euclidean distance and spectral clustering using Pearson distance. Louvain clustering was able to separate toMTG from this subnetwork at higher complexities, whereas other clustering algorithms could not (c.f., k=10 in panel A and I). Another trend across clustering approaches was the tendency for posteromedial PPS near SMG to be combined into a single parcel with posterior superior temporal sulcus (pSTS) and lateral STG (e.g., k=5,7 in panel A, k=4 in panel C,D, k=5 in E,F). Qualitatively, less consistency across clustering approaches was observed at higher parcellation complexities.

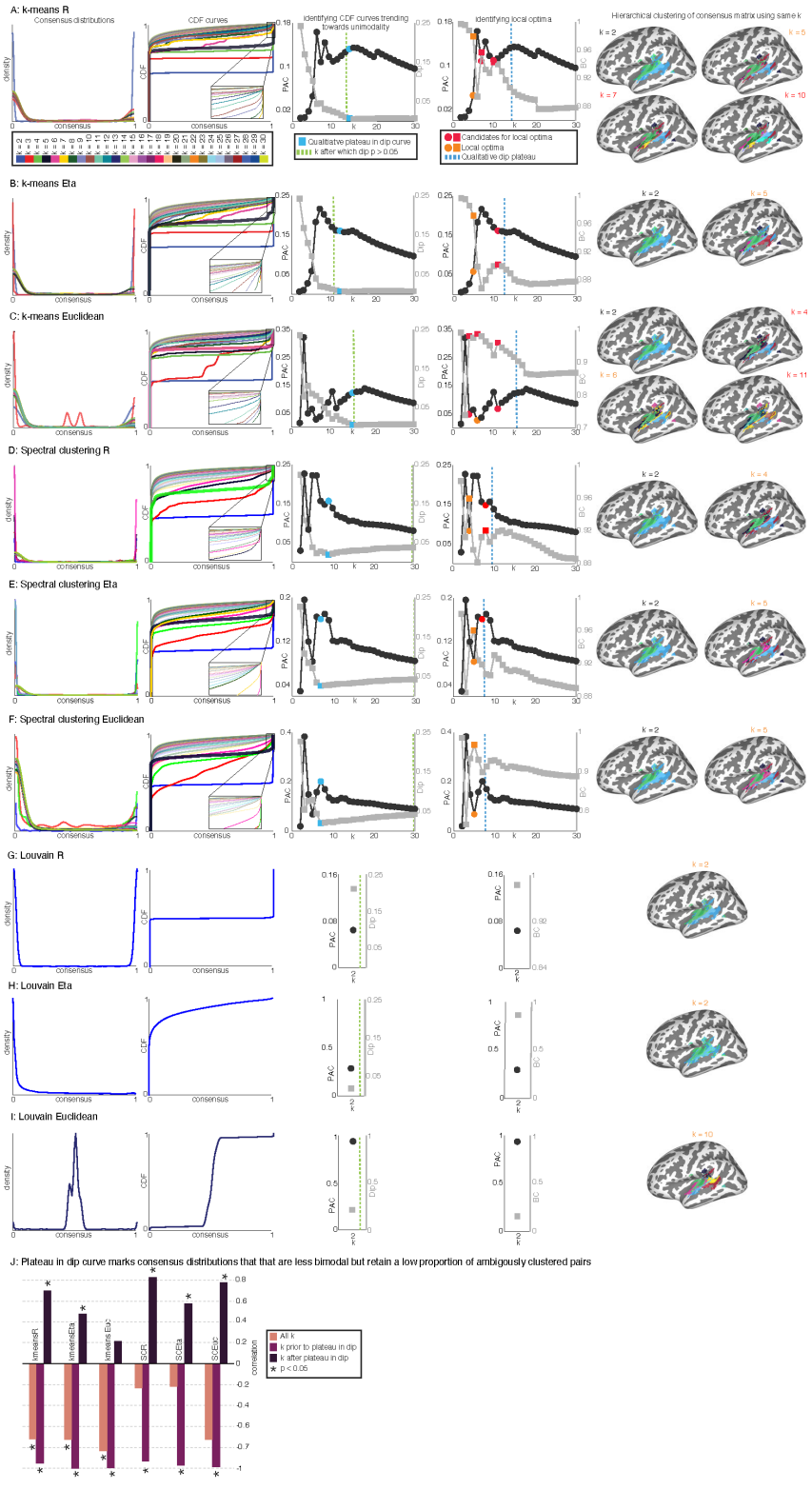

Figure 4: Selecting optimal models for parcellating group-level functional connectivity for different clustering approaches. Each panel A-I plots different information about the consensus matrices generated by each clustering approach. Each consensus matrix

represents a parcellation complexity between 2 and 30 (referred to as k). Information about consensus matrices is arranged across the first 4 columns of each panel, and the last column visualizes the parcellations contained within the set of consensus matrices that were identified as either optimal, or potentially optimal. The first four columns show the following from left to right: consensus distributions, cumulative distribution functions, curves for the proportion of ambiguously clustered pairs (PAC) and dip values, and curves for PAC and bimodal coefficients (BC). In the third column that shows PAC values (connected black dots) and dip values (connected grey squares), the point at which consensus matrices become unimodal is marked by a green dotted line. The point at which the dip curve appears to plateau based on qualitative inspection is also marked by blue squares and dots. The dip plateau was used to constrain the range of solutions considered to be candidates for the optimal solution based on PAC and BC. Parcellation complexities after the dip plateau are marked on the CDF curves in the second column by a 40% reduced opacity. The fourth column shows BC (grey squares) and PAC (black dots) curves in relation to this plateau (blue dotted line). Coinciding peaks in BC and valleys in PAC preceding the dip plateau were considered candidate optimal solutions and are shown in red. When these candidates were too unreliable based on standard deviation from the lowest PAC score on the curve (e.g., panels A,B), all solutions up to the first peak/valley were treated as candidates instead. The optimal solution is shown in orange and was selected based on being the highest complexity candidate within 2 standard deviations of the most reliable solution. The chosen optimal parcellation complexity is also marked on the CDF curves in the second column using a thicker line width. All candidate models are visualized on the surface in the last column (orange labels correspond to the selected solution, red labels correspond to the remaining candidates). Parcellations with 2 parcels are also shown because they represented the most reliable models overall (black labels). Panel J shows correlations between PAC and BC curves for each clustering approach. Correlations are carried out separately over all parcellation complexities (salmon bars), those parcellation complexities that precede the dip plateau (lighter purple bars), and parcellation complexities that follow the dip plateau (darker purple bars)

3.3.2 Fusing approaches for clustering group-level functional connectivity

Cluster ensembles were created at each parcellation complexity by combining parcellations from two clustering approaches, then a single optimal ensemble complexity was selected. To ensure cluster ensembles contained reliable but heterogenous parcellations, we also performed an analysis of how similar the parcellations contained within the consensus

matrices of each clustering approach were to each other. To this end, normalized variance of information (NVI) was used, although we found that other measures (i.e., Dice, normalized mutual information, Rand index adjusted for chance) exhibited identical patterns. NVI was selected over the Rand index because the procedure for adjusting this index for chance does not account for autocorrelation and spatial smoothing in the underlying data. NVI is highly similar to normalized mutual information but has the property of being a true metric.

NVI was computed between each pair of clustering approaches, at each possible parcellation complexity. NVI values were also averaged across parcellation complexity to rank dissimilarity between clustering approaches (Figure 5, panel A). Note, Louvain clustering is not included in the analysis because it produced a parcellation at a single complexity, and overall generated parcellations with either poor reliability, poor complexity, or both. This similarity analysis confirmed that all clustering approaches produced the same 2-parcel solution. It also indicated that the most dissimilar parcellations were generated for spectral clustering with Euclidean distance and k-means clustering, either with Pearson distance, or the eta^2 coefficient. K-means clustering with eta^2 produced slightly more similar parcellations in the range of complexities we expected to characterize the PPS zone (i.e., somewhere between ~5 to ~10 clusters; PPS is large, containing anterior STG, auditory cortex, toMTG, PO, and the PT the latter of which contains at least 3 cortical fields). In addition, the previous analysis established spectral clustering with Euclidean distance and k-means clustering with Pearson distance to produce the most reliable optimal models. For these reasons, ensembles were formed from parcellations produced by k-means clustering with Pearson distance and spectral clustering with Euclidean distance.

Overall, k-means clustering with eta^2 and Pearson distance produced the most similar parcellations irrespective of complexity. However, similarity for parcellations generated using these two measures wasn't always guaranteed. For instance, spectral clustering using these two measures produced parcellations that were less similar to each other than the parcellations produced by different clustering algorithms that used the same distance measure.

Using the same NVI metric, we also quantified the extent to which parcellations contained in the consensus matrix resembled parcellations that would have been produced otherwise, without consensus clustering (Figure 5, panel B). All clustering approaches exhibited exceedingly high parcellation similarities over all complexities, suggesting that consensus clustering typically had a modest impact on the clustering solution. K-means clustering using eta^2 and Pearson distance measures tended to be benefit most from consensus clustering. Consensus clustering had the smallest impact on spectral clustering using eta^2 and k-means clustering using Euclidean distance. For most clustering approaches, consensus clustering had a much higher impact at higher complexities. Spectral clustering tended to be impacted less by consensus clustering at higher complexities, particularly when using eta^2 or Euclidean distance. We note that qualitative comparisons between the parcellation schemes in this analysis revealed no recognizable differences at most parcellation complexities. As such, NVI in the range of 0.2 to 0.3 was considered to reflect nearly perfect agreement.

Fusing together the consensus matrices for k-means and spectral clustering generated more heterogeneity in consensus values (Figure 5, panel C). This was anticipated given that the fusion process involved averaging the two clustering approaches with the

most differently structured consensus matrices. Optimal ensemble complexity was determined using the same procedure for model selection applied to the individual clustering approaches. Thus, we looked for a complex ensemble that still showed *relatively* high agreement despite containing parcellations from dissimilar clustering approaches. A locally optimal 5-parcel ensemble was identified. Coincidentally, both of the clustering approaches that formed the ensemble also showed a locally optimal 5-parcel model. Qualitative inspection of the parcellation generated by the optimal ensemble showed that the fusion process mainly favored the k-means parcellation (Figure 5, panel D). Spectral clustering appeared to contribute to the ensemble parcellation by slightly reducing the extent to which the large auditory cortex parcel intruded into dorsomedial portions of the PT. NVI quantitatively confirmed the preference given to k-means clustering in the locally optimal ensemble model, but also showed that ensemble models were just as often more similar to the spectral clustering parcellations, depending on which specific complexity was being assessed (Figure 5, panel D).

Figure 5: Ensemble parcellation of group-level functional connectivity

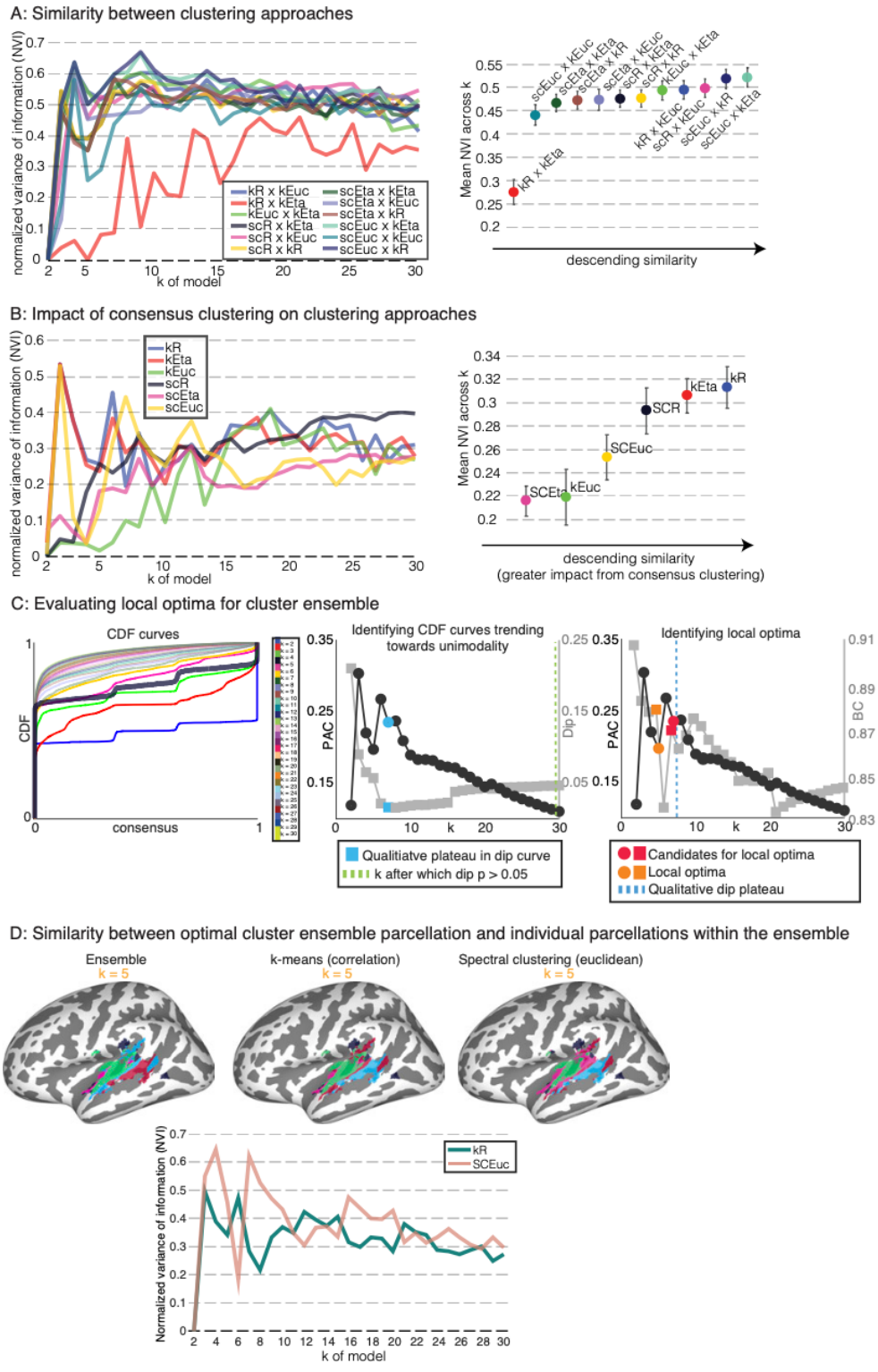

Figure 5: Ensemble parcellation of group-level functional connectivity. Panel A: Similarity between the parcellations generated by each pair of clustering approaches (excluding Louvain clustering for low reliability) is examined to select two reliable clustering approaches that generate dissimilar parcellation schemes. On the left, similarity is computed separately at each parcellation complexity (i.e., k), and the right shows the same similarity values collapsing across parcellation complexities (with SEM bars). Note, low normalized variance of information means greater similarity. Panel B: Similarity is also measured between parcellations generated through consensus clustering (i.e., by subsampling from the data and forming a cluster ensemble over subsamples) and parcellations generated without consensus clustering to understand the impact of the subsampling procedure on the final solution. On the left, similarity is shown at each level of complexity, and on the right similarity is collapsed across complexities (with SEM bars). Panel C: Chosen clustering approaches based on reliability measured in Figure 4 and similarity measured in panel A of this figure were used to create cluster ensembles at each level of parcellation complexity. A single ensemble complexity was then selected using the same model selection procedure that searched for optimal models for each clustering approach. The leftmost chart shows CDF curves for the consensus matrix formed from each ensemble. The optimal solution is highlighted by a thicker line. The middle chart shows PAC and dip curves for each ensemble, indicating where the dip statistic (grey) plateaued relative to PAC (black). The rightmost chart shows PAC (black) and BC (grey) curves, indicating peaks and valleys (red) of reliability and bimodality of consensus distributions. These were considered as potential candidates for the optimal solution (orange). Note, here all peaks and valleys are shown in red, but an optimal solution was only selected from the range of PAC values that preceded the dip plateau. Panel D: Optimal ensemble parcellation visualized on the surface, along with the parcellations from each of the clustering approaches that formed the ensemble (top row). Similarity between the ensemble parcellation and each of the clustering approach parcellations is shown as a function of parcellation complexity in the bottom row. Abbreviations: kR = k-means with Pearson distance; kEta = k-means using eta^2; kEuc = k-means with Euclidean distance; scEuc = spectral clustering with Euclidean distance; scR = spectral clustering using correlation; scEuc = spectral clustering using Euclidean distance-based similarity; scEta = spectral clustering using eta^2

3.3.3 Parcellating participant-level functional data

The process of selecting clustering approaches to generate cluster ensembles was independently carried out for participant-level data. For each clustering approach, an optimal model was selected within a participant (i.e., an optimal parcellation complexity). The reliability and complexity of models selected across participants was then compared

between clustering approaches. The distributions of locally optimal parcellation reliabilities and complexities in a subset of 25 participants is presented in Figure 6 (panels A and B). The initially generated distribution of optimal model reliabilities was used to refine optimal models at the participant-level by slightly increasing or decreasing optimal model complexity so that the reliability of the new models was within the scope of optimal models selected across the group (see methods for more details). The effect of group-level adjustment to the overall distributions in reliability and complexity of optimal models was negligible, mainly serving to highlight those participants for whom even a two-parcel solution was too unreliable relative to the group (i.e., see the addition of participants with an optimal model of 1-parcel in group-adjusted distributions). Paired t-tests between the original and group-adjusted distributions of optimal models revealed no significant differences (see rectangular matrix in panel C, Figure 6).

Individual variability in parcellation complexity was highest for the k-medoids algorithm, particularly when combined with Euclidean distance. Coincidentally, this algorithm also performed poorly according to the distribution in PAC. Although the eta^2 and correlation coefficients are highly related, the correlation coefficient (or Pearson distance) tended to produce a slightly wider range of optimal parcellation complexities across participants, but at the cost of slightly higher mean and median in PAC. The only exception to this pattern was Louvain clustering, which generated much higher parcellation complexities for the correlation coefficient than for eta^2. Clustering with Euclidean distance tended to produce a slightly wider range of optimal parcellation complexities compared to the other distance measures, but only sometimes showed lower mean and median in PAC (i.e., for k-means and Louvain clustering, but not spectral

clustering). This trend towards larger optimal parcellation complexities was also observed when clustering group-level functional data. Some clustering approaches appeared to divide participants into groups that had relatively more and less reliable parcellations. Notably, often this would not correspond to a bimodal distribution in parcellation complexity. This appeared to be the case for Louvain clustering with either the eta^2 coefficient, or Euclidean distance, and for spectral clustering with either correlation or eta^2 coefficients.

Differences between clustering approaches with respect to both reliability and complexity were tested using paired two-sample t-tests (Figure 6, panel C). This confirmed that k-medoids generated significantly higher complexities than all other clustering approaches ($p < 0.05$). It also showed that Louvain clustering often produced parcellations with significantly higher complexities when compared to k-means or spectral clustering. Nevertheless, both Louvain clustering and k-medoids also produced significantly less reliable parcellations ($p < 0.05$). This was true across all comparisons to clustering approaches that relied on k-medoids, and for all comparisons to clustering approaches that relied on Louvain, with the only exception being that k-medoids used with the correlation coefficient was also significantly less reliable than Louvain clustering with eta^2 and Euclidean distance. Some differences between the reliability of spectral and k-means clustering were also observed ($p < 0.05$). However, there was no significant difference between the reliability of parcellation models generated with k-means clustering and those generated with spectral clustering using either Pearson or Euclidean distance. There were also no significant differences in reliability between different distance measures executed with the same clustering algorithm (i.e., for spectral clustering and k-means).

In order to fuse clustering approaches that generated equally reliable models, we constrained our final choices of clustering approaches to spectral clustering using Pearson or Euclidean distance, and k-means clustering using all three distances. Because the purpose of the ensemble was to leverage the strengths of different clustering approaches, it was ultimately decided that a spectral clustering and a k-means clustering approach would be consolidated (i.e., because these clustering algorithms are demonstrably sensitive to different kinds of clusters). Spectral clustering using Euclidean distance was given preference because it was less likely to split participants into more discrete groups of high and low reliability and showed a larger range of optimal parcellation complexities. Because k-means clustering using Euclidean distance produced a relatively large range of complexities as well, k-means clustering using Pearson distance and eta^2 were given preference as the complement to spectral clustering with Euclidean distance, encouraging some heterogeneity in the resulting cluster ensembles. The distribution of complexities and reliabilities were highly similar between k-means clustering with Pearson distance and eta^2. Ultimately, k-means clustering with Pearson distance was selected as the complement to spectral clustering with Euclidean distance to maintain some parity with the cluster ensembles formed from group-level functional data and to decrease the amount of time necessary to compute cluster assignments (eta^2 clustering was performed using in-house scripts that were not as heavily optimized).

The two selected clustering approaches were extended to those participants that were not included in the initial evaluation of clustering approaches. Optimal models were selected for each of these participants using these clustering approaches to observe how the distribution of optimal model complexities and reliabilities was impacted. The range of

optimal model complexities (Figure 6, panel D) and reliabilities (Figure 6, panel E) did not change for clustering approaches. Across all participants, there was no significant difference between optimal parcellation complexities identified by these two clustering approaches ($p < 0.001$). However, k-means clustering with Pearson distance did produce models with significantly higher reliability ($p < 0.001$). The actual difference in mean reliability between the two approaches was modest and consistent with what was observed in the smaller subset of participants, with k-means clustering producing less ambiguous parcel assignments for on-average 2% of voxels in the PPS zone. Optimal ensemble models were also selected within-participant. Reliability of optimal ensembles was significantly lower than the reliability of optimal models selected by each clustering approach, as anticipated ($p < 0.001$). On-average, optimal ensembles produced more ambiguous parcel assignments relative to each of the individual clustering approaches in 12% of PPS voxels. Given that the clustering approaches individually showed high reliability, the decrease in consensus for parcel assignments must reflect the fusion of different patterns in the data discovered by the clustering approaches. At the same time, the fusion of these discrepant patterns does not typically lead to the discovery of additional clusters, or the collapse of existing clusters. This is evinced by the fact that creating ensembles did not produce optimal models that differed with respect to optimal parcellation complexities for individuals (Figure 6, panel D).

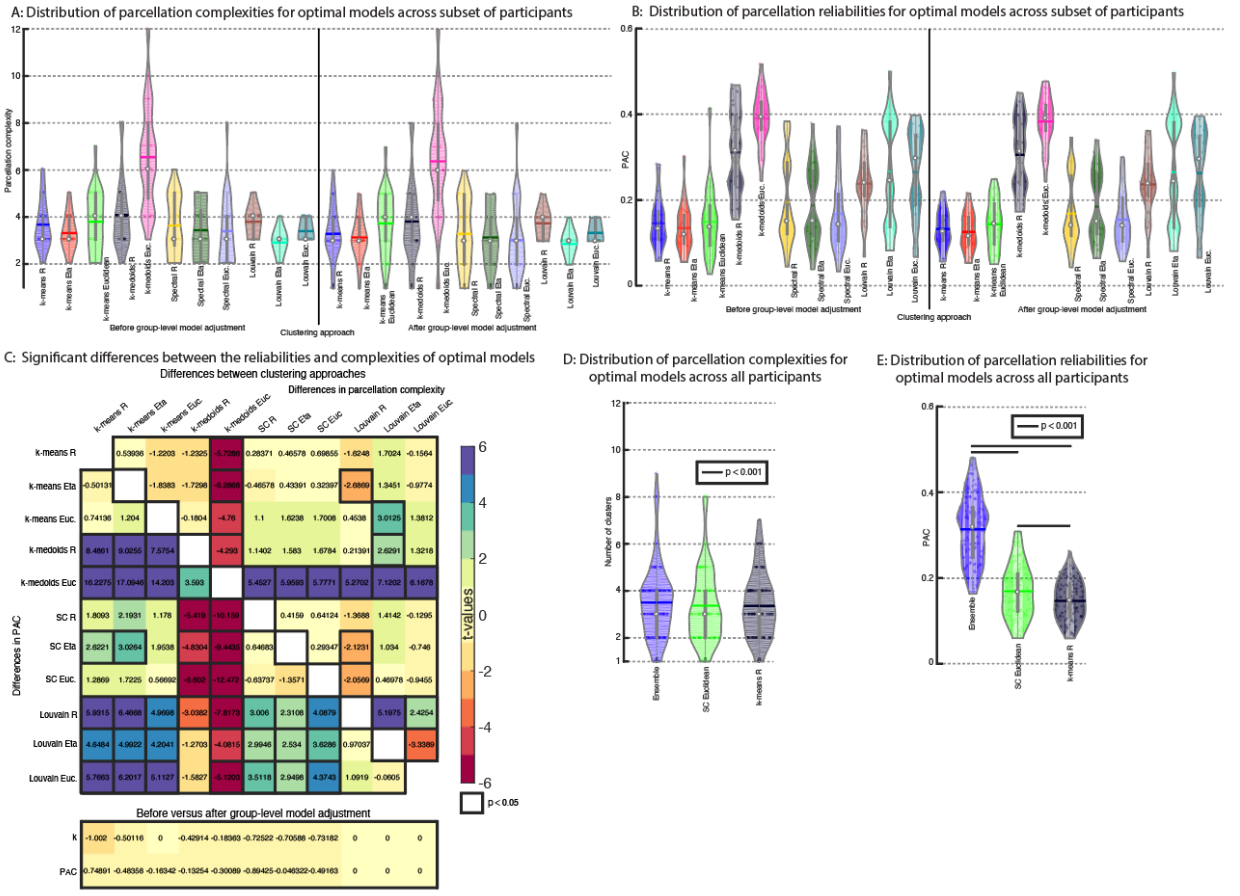

Figure 6: Comparing PAC and parcellation complexity across clustering approaches for subject-level parcellations. Panel A: Distributions of optimal parcellation model complexities across a subset of participants are presented for each clustering approach as violin plots. In each violin, the median is represented by a white dot and the mean by a thick horizontal line. A vertical line marks the division between group-adjusted models and non-group-adjusted models (i.e., group-level reliability was used to nudge optimal participant-level parcellation complexities either up or down to be in line with the average). Panel B: Distributions of optimal parcellation model reliabilities across a subset of participants are presented for each clustering approach as violin plots. Panel C: The square heatmap represents the results of t-tests carried out between each pair of clustering approaches. T-values are shown within each square and outlined boxes represent significant t-values ($p < 0.05$). Separate t-tests were carried out to test for differences between optimal parcellation model complexities (upper triangle) and differences between optimal parcellation reliabilities (bottom triangle) across participants. The lower rectangular heatmap represents t-tests performed for each clustering approach, testing whether adjusting optimal models within-participants based on group reliability resulted in optimal models with significantly different complexities or reliabilities. Panel D: Distributions of optimal parcellation model complexities are shown across all participants for the two clustering approaches selected to form cluster ensembles based on the results presented in the other panels. Cluster ensembles were created for each participant at each

level of parcellation complexity, and the same optimal model selection procedure used for individual clustering approaches was carried out to identify optimal ensemble models within-participants. The distribution of complexities for these optimal ensemble models is also presented in this panel. T-tests were carried out to test whether optimal model and ensemble complexities differed across participants (none did; $p < 0.001$). Panel E: Distributions of optimal parcellation model reliabilities are shown across all participants for the two clustering approaches selected to form cluster ensembles. The distribution of optimal ensemble model reliabilities is also shown. T-tests were carried out to test whether optimal model and ensemble reliabilities differed across participants and significant differences are marked by solid horizontal lines ($p < 0.001$).

3.3.4 Forming an ensemble model that collapses across participants

In the preceding section, we identified two clustering approaches to form cluster ensembles within participants. Our next aim was to consolidate participant-level ensembles by creating new cluster ensembles that contained all of the parcellations generated by the chosen clustering approaches for all participants. These group ensembles were created separately at each level of parcellation complexity. However, collapsing across both participants and clustering approaches introduced a degree of variability in consensus that our reliability measure could not resolve, making it difficult to use this measure to select an optimal group ensemble. Thus, our strategy for selecting a single group ensemble involved leveraging the information that we had about which parcellation complexities were optimal for within-participant ensembles that combined the two selected clustering approaches. In other words, we looked at optimal complexities within-participants to understand which complexity could be best fit to the group. First, we constrained our search for the optimal group ensemble complexity to the range of optimal ensemble complexities identified within participants. Then, we tested how well each of

these parcellation complexities could be fit to the group by comparing group ensembles that collapsed across all participants, to a different set of group ensembles that only collapsed across those participants for whom that complexity was optimal. Put otherwise, parcellation complexities were deemed a good fit to the group if they resulted in an organizational scheme of the PPS (i.e., parcellation) that managed to retain the characteristics of the organization found in participants to whom this complexity was uniquely suited. Measuring how well the subgroup ensemble parcellation compared to the (whole) group ensemble parcellation involved computing the same NVI metric from previous similarity analyses (Figure 7A). The decision of which final group ensemble model to focus on was based on relative differences in the NVI metric—the overall aim was to select the most complex model that still showed high similarity to the subgroup ensemble model. Surprisingly, although NVI generally tended to increase with parcellation complexity, an 8-parcel solution showed exceptionally high similarity between group and subgroup ensemble parcellations. Indeed, the 2-parcel solution was the only one that showed higher similarity. This suggested that even though a parcellation complexity of 8 parcels was far above the optimal complexity of the average participant (which was 4), an 8-parcel solution fit to most participants generated a relatively consistent pattern of organization that retained the characteristics of organization present in participants for whom this solution was optimal. Although relatively few participants showed an optimal 8-parcel solution, multiple other solutions were optimal for small groups of participants (i.e., ranging from 1-14) but failed to show high group-subgroup similarity.

To get a better sense of whether the similarity that was observed for the selected 8-parcel group ensemble model was reasonably high, we carried out the same procedure of

comparing parcellations resulting from ensembles that collapse across all participants to those resulting from ensembles that just collapse over those participants for whom a certain complexity is optimal, but in the two individual clustering approaches that were previously consolidated (Figure 7A). Group and subgroup ensemble parcellations generated with k-means clustering were found to be virtually always more similar than those generated with spectral clustering (the only exception was 2-parcels). For spectral clustering, the 8-parcel solution was overall the most similar, however, the same complexity in the ensembles that consolidated clustering approaches showed comparable similarity. The 8-parcel solution selected for the group ensembles that consolidated clustering approaches presented one of the highest group-subgroup similarities. The only other solutions that exhibited higher similarity were 2 and 3-parcel solutions produced with spectral and k-means clustering.

The 8-parcel solution we selected revealed a complex arrangement of areas (Figure 7A). The superior temporal plane contained 4 areas—the most anterior aligned with auditory cortex, 2 more areas fell squarely within the boundaries of anatomical PT, and one last area roughly aligned with the location where we might expect Spt. The auditory area (AA) covered both Heschl's gyrus (HG) and Heschl's sulcus (HS), as well as some of anterior PT. The first area situated largely within the PT aligned with mid-to-posterior PT (mpPT) and extended dorsomedially. An area in posterior PT (pPT) formed a parcel subnetwork with non-adjacent areas in the PO and toMTG (pPT subnetwork). The area in the vicinity of Spt was posterodorsal and intruded into SMG (posterodorsal PPS; pdPPS). Portions of lateral and ventral STG also separated into anterior and posterior areas. Lateral STG contained an anterior area just ventral to auditory cortex (anterolateral STG; alSTG) and a

posterior area ventral to the PT (posterolateral STG; plSTG). Ventral STG contained an anterior area just beneath anterolateral STG (anteroventral STG; avSTG) and a posterior area beneath posterolateral STG (posteroventral STG; pvSTG). Anteroventral and posteroventral STG areas extended into the ventral bank of the STS, with pvSTG also including a large portion of pSTS and intruding into posterior middle temporal gyrus (pMTG).

Although our selection of the 8-parcel solution from the group ensembles was based on computing parcellation similarities with NVI, we also qualitatively inspected similarities between group and subgroup ensemble parcellations by visualizing them on the surface (Figure 7B). In the ensembles that consolidated clustering approaches, this confirmed generally good group-subgroup ensemble similarity for the 8-parcel solution. Nevertheless, collapsing across participants introduced differences in organizational patterns at most levels of complexity (i.e., excluding 2-parcel solutions). In relation to the 8-parcel solution fit to all participants that we described above, the following differences were found by only collapsing across participants with an optimal 8-parcel solution: pdPPS formed a parcel subnetwork with plSTG; pPT and PO areas were slightly larger, and their subnetwork excluded toMTG; and toMTG formed a parcel subnetwork with a posterior portion of plSTG. Ultimately, collapsing across participants produced parcels with better definition, undoubtably due to the increased sample sizes. Visualizing parcellations generated from the individual clustering approaches that we consolidated showed the influence that each clustering approach had on the final 8-parcel solution that we focus on. Both clustering approaches generated a similar organizational scheme of areas, but k-means clustering failed to separate anterior and posterior portions of lateral STG, while spectral clustering

failed to separate pdPPS from the pPT subnetwork. At the same time, both clustering approaches found additional parcels unique to each approach that were relatively small and sometimes spatially unresolved. Thus, combining clustering approaches produced a parcellation that only kept the unique parcels discovered by each of the clustering approaches that were more robust and consistent.

As a further test of whether the 8-parcel solution is a good fit to individuals with lower optimal parcellation complexities, we identified the networks that distinguished these parcels in the whole group of participants, then tested whether the same networks distinguished these parcels in subgroups of participants with low optimal parcellation complexities. If participants with lower optimal parcellation complexities showed poor evidence for an 8-parcel organization of the PPS zone, we expected that fitting the group-defined parcels onto their data would fail to identify the same network differences seen at the level of the entire group, where the 8-parcel organization was more likely to be a good fit (i.e., because the group always contained more participants with higher optimal parcellation complexities). The networks associated with each parcel were mapped by producing a separate functional connectivity map for each parcel, then performing a winner-take-all classification analysis that assigned voxels in the brain to the PPS parcel they were most strongly connected to (Figure 7C). In the entire group of participants, this analysis also provided information about which networks presumably drove the clustering results, and we describe those networks in detail first (Figure 7C, top left panel). We note that AA is distinguished by a sensorimotor network that primarily spans auditory cortex, motor cortex, posterior insula, medial visual cortex, and the cingulate gyrus; mpPT is distinguished primarily by connectivity to two areas in dorsal premotor cortex that align

with prior descriptions of the frontal and prefrontal eye fields (c.f., Glasser et al., 2016); plSTG is distinguished by connectivity to an area in dorsal premotor cortex that is sandwiched between the eye fields that aligns with a portion of previously described area 55b, but does not intrude as far into posterior middle frontal gyrus (pMFG; Glasser et al., 2016); the pPT subnetwork is distinguished by connectivity to a wide network of parietal areas (anterior SMG, PO, superior parietal lobule), temporooroccipital areas (toMTG and temporooccipital inferior temporal gyrus; oITG), anterior middle frontal gyrus (aMFG), anterior insula, frontal operculum, ventral premotor cortex (vPMC), dorsal somatosensory cortex on the medial surface, and the cingulate motor area; alSTG is distinguished primarily by connectivity to right hemisphere areas of the MFG but also the anterior segment of the inferior frontal gyrus (IFG) pars triangularis; avSTG is distinguished by connectivity to superior frontal gyrus (SFG), anterior temporal lobe (ATL), and a more anterior portion of angular gyrus (AG); pvSTG is distinguished by connectivity along mid-to-posterior STS, pMTG, portions of posterior inferior temporal gyrus (pITG), more dorsal AG, pMFG, in the left hemisphere, and most of IFG; curiously pdPPS is more strongly connected to a small ventral portion of IFG pars opercularis, pSMG, and some segments of aMFG.

An identical winner-take-all analysis was performed using the same 8 cluster parcellation of the PPS zone but using only subgroups of participants to calculate functional connectivity maps. Participants showing an optimal 2, 3, and 4-parcel solution identified remarkably similar networks that distinguished the 8 parcels (Figure 7C). The features that distinguished these parcels across all participants were present in each group, even when those features were relatively specific. For example, plSTG was always distinguished by a very small dorsal premotor cluster that was flanked by premotor areas associated with

mpPT. At the same time, pdPPS was always marked by strong connectivity to a small portion of pars opercularis that was abutted anteriorly by vPMC associated with the pPT subnetwork, and posteriorly by IFG associated with pvSTG. Further, alSTG was always marked by stronger connectivity to the right hemisphere, particularly along the MFG. The other networks showed even stronger similarities, suggesting that plSTG, alSTG, and mpPT in particular may be more difficult to identify in some participants.

Finally, we showed that the similarity we measured between the group and subgroup ensemble parcellations that contained 8 clusters was high by demonstrating that a bigger source of variance in parcellation similarity was the choice to either create ensembles from group functional connectivity data, or from participant-level parcellations (Figure 7D). Parcellations generated by the analysis of group-level and participant-level data were compared between ensembles that consolidated clustering approaches, and those that did not. Similarity was substantially lower for the 8-cluster group-subgroup parcellation comparison than over the first 14 complexities comparing parcellations generated by the analysis of group-level and participant-level data.

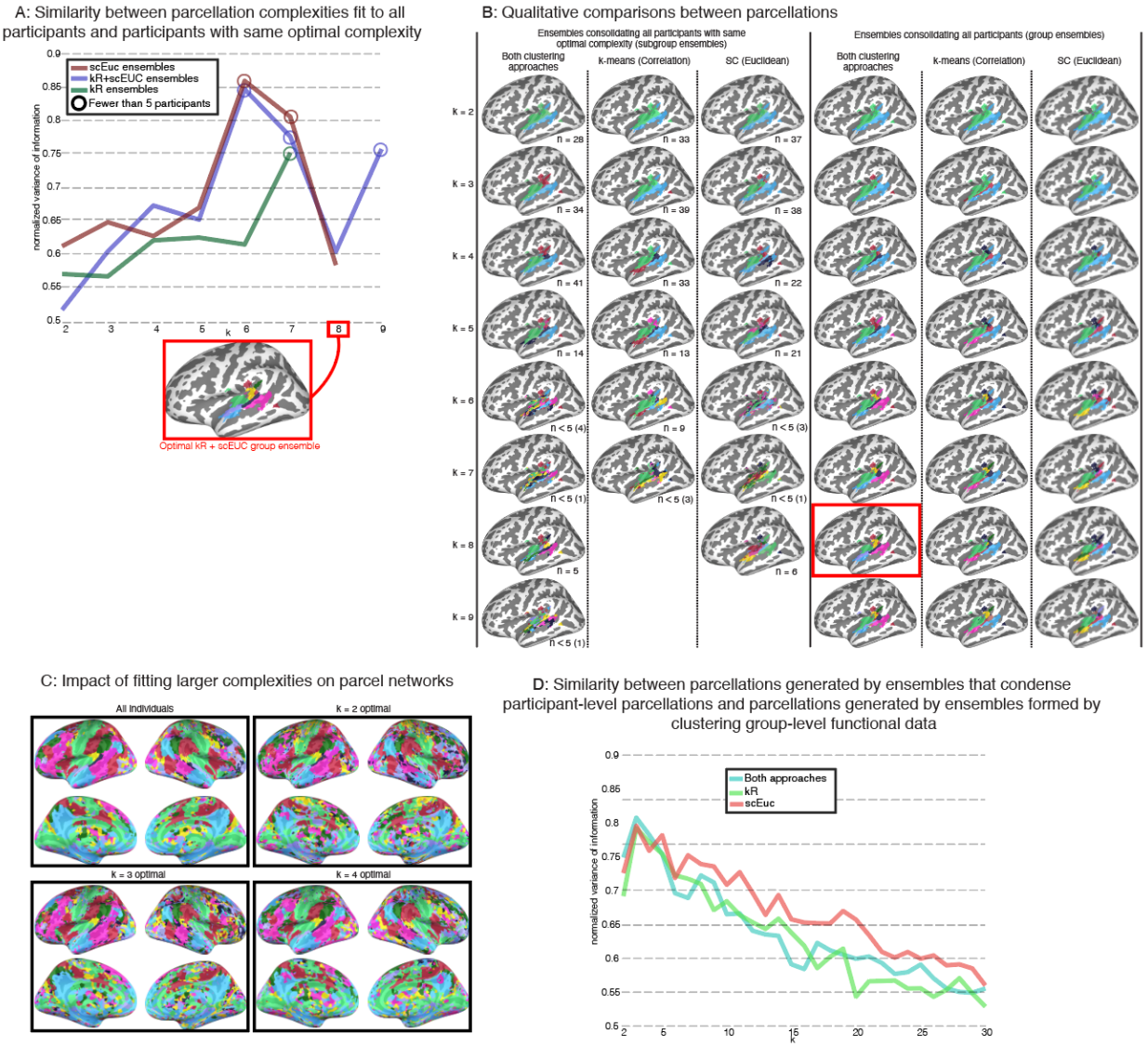

Figure 7: Comparison between consensus clustering at individual and group levels. Panel A shows the similarity between the parcellation extracted from a group cluster ensemble generated over all participants and a parcellation extracted from a group cluster ensemble generated over participants with the same optimal parcellation complexity. Ensembles consolidating clustering approaches (brown) show a high similarity for a relatively complex solution, which we consider the optimal group ensemble complexity (red box). Similarities for the individual clustering approaches that were consolidated are shown to give a reference point for the optimal group ensemble complexity similarity value. Panel B visualizes the parcellations for which similarities in panel A were computed. Panel C shows the vertices most strongly connected to each parcel in the 8-parcel optimal ensemble parcellation. In the top left panel, connectivity to parcels is defined over all participants. In the other panels, connectivity is defined on subgroups of participants for whom a much lower complexity was optimal than fit to the entire group. In these subgroups, the

networks that define each parcel are conserved. Panel D shows similarity values for comparisons between parcellations from panel A and equivalent parcellations but generated over group-level functional connectivity data (i.e., from Figures 4 and 5). In the legend, "both approaches" refers to comparison between parcellations formed from ensembles that condensed both clustering approaches.

3.3.5 Classification of coactivation data

We next sought to determine more specifically whether ensembles that combined participant-level parcellations were superior to ensembles that consolidated parcellations formed by clustering group-level functional data. The parcellations generated by both approaches were externally validated by testing whether coactivation for voxels in the PPS zone could be accurately classified according to their parcel schemes (Figure 8). A classification analysis using c-SVC was performed for parcellation complexities of 2, 5, and 8. Although we did determine a single optimal ensemble complexity for both methods of generating group parcellations, these complexities did not match (i.e., 8-parcels was best when combining participant-level parcellations, but 5-clusters was best when parcellating group-level data). Thus, to ensure that differences in accuracy of classification were not strictly the result of different parcellation complexities, both 8 and 5-cluster parcellations generated by both methods for producing group parcellations were tested. Two-cluster parcellations were also tested purely for the reason that they were always the most reliable solution. The classification performance for these parcellations was benchmarked against classifiers trained on a popular structural atlas (Harvard-Oxford), functional atlas (AICHA), and multimodal atlas (HCP MMP 1.0). Performance was tested using stratified 10-fold CV and repeated multiple times to ensure performance wasn't impacted by the specific

selection of voxels belonging to a parcel. The significance of classifier performance was evaluated using a permutation analysis (see methods for more details).

Overall, ensembles that condensed participant-level parcellations had substantially better accuracy, and higher Mathews correlation coefficients (MCC). Increases in parcellation complexity tended to inversely affect accuracy by a few percentage points and lowered mcc by roughly 10%. Nevertheless, voxels in the PPS zone could be classified into the most complex 8-cluster parcellation built up from individual participants with just over 90% accuracy and a 0.74 MCC. Trends were generally more dramatic for MCC than accuracy. The 5-cluster and 8-cluster parcellations built up from participant parcellations were predicted with roughly 10% higher accuracy and were roughly 50% more similar to the predicted parcellations compared to parcellations generated from group-level data (i.e., MCC). The 2-cluster parcellation showed only a slightly higher accuracy and MCC for parcellations built up from participants. Performance for the most complex 8-cluster parcellation of the PPS built up from participant-level parcellations was higher compared to performance on any of the benchmarked atlases--accuracy was roughly 5% higher and MCC was higher by 0.17. However, the atlases outperformed an 8-cluster parcellation of the PPS formed by clustering group-level functional data. Notably, performance on the HCP MMP atlas was likely underestimated because this surface space atlas was registered to volume space in order to perform the parcellation. Classifier performance on all parcellations and atlases was significant ($p < 0.001$).

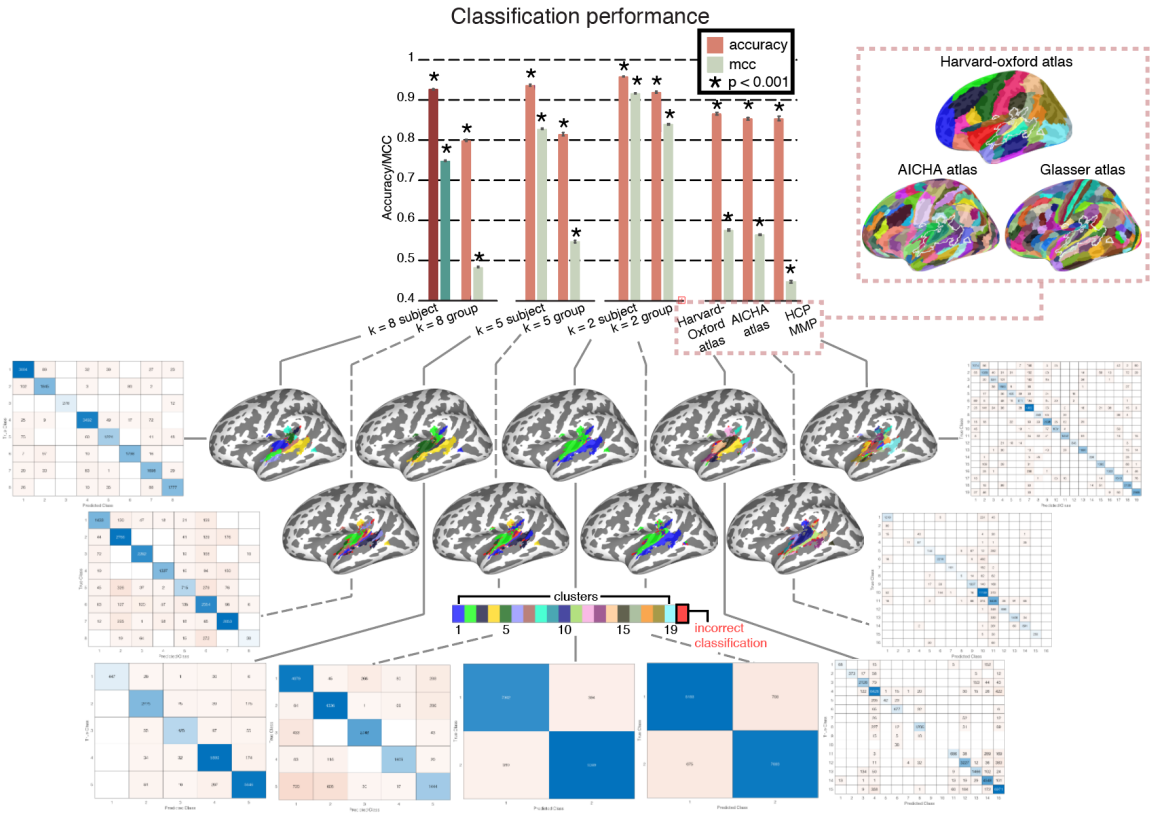

Figure 8: Classification of coactivation data using different parcellation schemes. Accuracy for classifying voxels in the PPS zone into different parcellation schemes according to external coactivation data is presented. The bar plot shows classifier performance: accuracy in blue and Matthews correlation coefficient in red. The leftmost two bars are oversaturated to highlight the solution we focus on. Parcellation complexities are referred to as 'k'. The short-hand for parcellations built up from clustering of group-level functional data is 'group' and the short-hand for parcellations formed by condensing participant-level parcellations is 'subject'. Each of the tested parcellations is depicted below the corresponding bar and connected by a line. Vertices highlighted in red represent misclassifications. Below each parcellation is the confusion matrix showing classifier performance. The leftmost 3 of the tested parcellations were areas extracted from several atlases. Atlases are shown in the red dotted box on the right.

3.3.6 Investigating connectivity networks in the best performing 8-cluster parcellation of the PPS zone

Now that we have established the parcellation that tends to be best predicted by external data, we more carefully consider the connectivity patterns associated with each of its parcels. Because the parcellation we focus on is the one that condensed participant-level parcellations, we have already shown the brain areas that are most strongly associated with each of its parcels using a winner-take-all classification analysis of functional connectivity computed from each parcel (Figure 7C). However, this previous analysis was aimed at more generally describing patterns of connectivity associated with each parcel across multiple subgroups of participants with different sample sizes, and in this context, functional connectivity maps were uncorrected. In Figure 9, we repeated the winner-take-all analysis while correcting each parcel's functional connectivity map for multiple comparisons (Figure 9A). This did not result in networks that were significantly different from what we observed in Figure 7, and we refer to the text describing this figure for their more thorough description. To probe more subtle differences in parcel-level functional connectivity, we repeated the winner-take-all analysis on different groupings of the 8 clusters in the parcellation. For example, we were interested to see which clusters on the superior temporal plane showed greater connectivity to temporal areas dominated by avSTG, pvSTG, plSTG, and alSTG. Further, we wanted to better characterize the functional connectivity of the pPT subnetwork, and to understand how strongly more ventral areas were connected to somatosensory and motor cortex, the connectivity to which was dominated by AA.

Analysis of parcels on the superior temporal plane included the AA, mpPT, pdPPS, and pPT (without PO and toMTG), and indicated that connectivity to mid STS was strongest to pdPPS, while connectivity to pSTS appeared to be strongest for mpPT (Figure 9B). The pPT area of the PO-toMTG-pPT subnetwork overall showed strong similarity to the same areas that distinguished this subnetwork from the rest of the PPS zone in Figure 9A. However, pPT also appeared to show stronger connectivity to a portion of pMFG that was overall more strongly connected to pvSTG. The mpPT area showed stronger connectivity along the entire lateral STG, as well as dPMC just anterior to pMFG, which included both of the eye fields that distinguish mpPT from the rest of the PPS zone, but also the area sitting between the eye fields that is overall more strongly associated with plSTG. In parallel, mpPT showed stronger connectivity to temporal visual areas. Comparison of areas on the lateral STG showed alSTG to have a connectivity profile generally more similar to AA, and plSTG to show a connectivity profile more similar to the other areas on the superior temporal plane (i.e., connectivity to PO, STS, plSTG, PT, IFG, dPMC, vPMC; Figure 9C). Curiously, one exception to this was connectivity to visual temporal areas, which appeared to be more associated with areas posterior to AA on the superior temporal plane (i.e., mpPT) but was more strongly associated with alSTG. The two ventral clusters in the PPS zone were also compared, showing a similar pattern, with avSTG exhibiting relatively stronger connectivity in a network similar to the one associated with AA (but slightly sparser connectivity to motor and somatosensory cortex, and stronger connectivity to ATL), and pvSTG showing stronger connectivity to areas that have been associated with other parcels on the superior temporal plane (Figure 9D). Thus, comparisons across areas distributed from more dorsal portions of the PPS zone to more ventral portions of the PPS

zone appear to broadly reflect an anterior-posterior difference, with anterior areas (i.e., AA, alSTG, avSTG) expressing stronger connectivity to a similar network that includes auditory, motor, and somatosensory areas, and posterior areas expressing stronger connectivity to a similar network of association areas, as well as two premotor regions (vPMC and dPMC). In addition, we examined connectivity to areas of the pPT subnetwork, which showed that connectivity data may be able to decompose this subnetwork, even though the clustering analyses were not able to (Figure 9E). The pPT in particular showed relatively stronger connectivity in a network that resembled that of mpPT, but also included ventral motor cortex, and IFG, particularly the portion of pars opercularis overall more strongly associated with pdPPS. The PO showed stronger connectivity to both vPMC and auditory cortex, although the comparison between areas on the superior temporal plane indicated that pPT is also connected to vPMC.

Because areas around premotor and motor cortex appeared to distinguish many parcels, we took a closer look at the connectivity between the parcels and a handful of ROIs that were selected to map onto some of the areas that we observed to distinguish parcels in the winner-take-all analysis (Figure 9).This included the eye fields associated with mpPT (FEF and PEF), and the area of vPMC (6r) that was associated with pPT. To try to capture the area of the dPMC connected more strongly with plSTG, we included area 55b, located at the transition between posterior middle frontal gyrus and premotor cortex, and dorsal laryngeal motor cortex (dLMC), which often contains a gyral component near area 55b (Bylek, 2021). In addition, we included an analysis of connectivity to auditory cortex as a contrast to the motor-driven ROIs. Briefly, areas FEF, PEF, 6r, and 55b were extracted from the HCP's MMP 1.0 atlas. The dLMC was defined by performing an activation likelihood

estimation (ALE) meta-analysis of studies (N=9) reporting activity in dorsal laryngeal motor cortex. Most of these studies also found foci in ventral laryngeal motor cortex (vLMC), and the resulting areas were separated into two ROIs. A high threshold ($p < 0.0001$) was applied to the meta-analysis in order to minimize overlap between area 55b and dLMC. First, this analysis showed that most parcels of the PPS zone were functionally connected to areas along premotor and motor cortex, as well as auditory cortex. The analysis also showed that most parcels had much higher connectivity to auditory cortex than premotor or motor cortex. The ROI analyses demonstrate that mpPT shows stronger connectivity to the eye fields, and that connectivity to the motor cortex ROIs is relatively high for AA, mpPT, avSTG, and plSTG. Interestingly, we note that parcel connectivity to auditory cortex could explain up to 96% variability in parcel connectivity to motor cortex (dLMC: $r=0.95$, $p < 0.01$, vLMC: $r = 0.98$, $p < 0.01$). Unfortunately, the ROIs for 55b and dLMC did not appear to capture the dPMC area most strongly connected to plSTG. Area 55b had relatively high functional connectivity to plSTG, but had higher functional connectivity to pvSTG, reflecting the greater overlap between 55b and the network of areas that activate most strongly for pvSTG (i.e., Figure 9A).

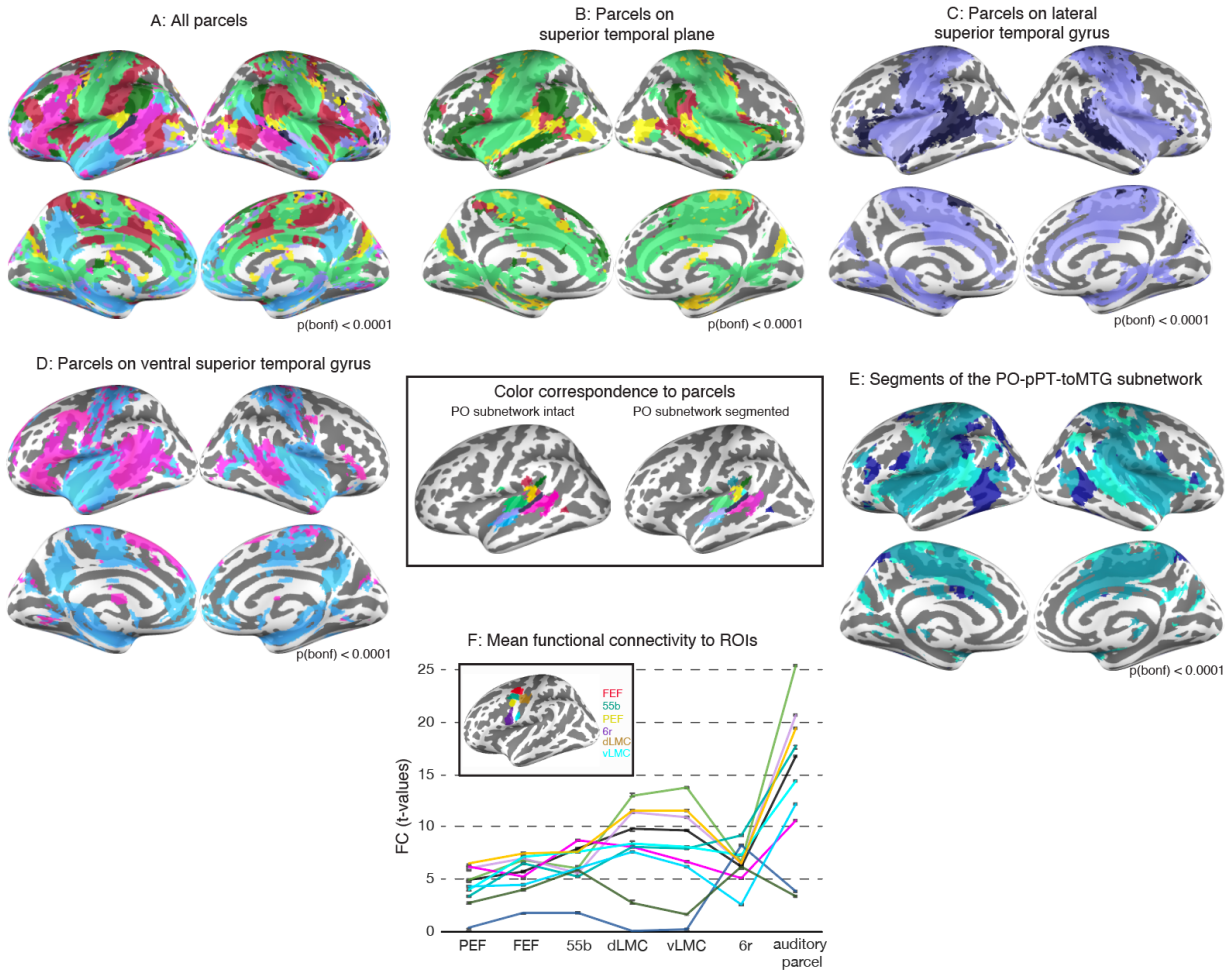

Figure 9: PPS zone networks. Functional connectivity maps were generated for each parcel of the PPS zone and a winner-take-all analysis was performed classifying each voxel with the parcel to which it is most strongly connected. The result is visualized by assigning the color of the "winning" parcel to each vertex on the surface. The parcel associated with each color can be found in the black box in the center of the figure. Panel A shows the result of this analysis when all parcels in the PPS zone are considered. The remaining panels show comparisons between groups of increasingly more ventral parcels. Panel B shows the same analysis but when only parcels on the superior temporal plane are considered (note, only the pPT portion of the red pPT-PO-toMTG subnetwork is used as it is the only area of this subnetwork on the superior temporal plane). Panel C shows the same analysis for parcels on lateral superior temporal gyrus. Panel D shows the analysis for parcels on ventral superior temporal gyrus. Panel E shows the same analysis for areas that makeup the pPT-PO-toMTG subnetwork. Across these panels, anterior clusters tended to show stronger connectivity to the same network. An ROI analysis was performed to better capture connectivity in areas of premotor cortex across parcels. This was motivated by evidence from the other panels that areas along premotor cortex are important for distinguishing many parcels. ROIs are shown superimposed on the figure in a black box. An additional auditory cortex ROI (i.e., the auditory parcel in the PPS zone) is included to contrast with

the motor regions. Mean functional connectivity between each parcel and the voxels within an ROI is computed (error bars are SEM).

3.3.7 Decoding parcels

The functions of networks attached to each parcel in 8-cluster parcellation scheme were decoded by testing the classifiers trained on coactivation data, on meta-analyses in neurosynth (Figure 10, panel A). Meta-analyses that generated a mean z-score of less than 0.8 for voxels in the PPS were exclude from analysis, trimming 1335 terms in neurosynth to just 165 terms. Classifiers were tested on meta-analyses a total of 1000 times. Each meta-analysis was assigned to the cluster it was most frequently classified into across these 1000 repetitions. Any cluster assigned with a frequency less than 0.7 was deemed too inconsistent to decode (only ~10 terms were excluded in this manner). The results of this analysis indicated that the network of AA was associated with low-level auditory and somatosensory processing. The network of alSTG was associated with higher-level auditory processing, including processing of speech sounds and music. Just posterior to this, plSTG had a network implicated in speech production. Moving dorsally, the pPT network (note, the original classification analysis did not decompose this subnetwork) was associated with sensorimotor processing and secondary somatosensory cortex. The two clusters on either side of pPT did not have networks that were associated with any meta-analyses, and included the mpPT, and pdPPS. Finally, the two parcels in ventral STG had networks which were associated with higher-level language processing (pvSTG), and the integration of higher-level auditory information. For instance, the network attached to the pvSTG was involved in language, sentence processing, word processing, lexical-semantics, phonology,

and reading. It was also associated with movements, motor, and premotor cortices. The avSTG was part of a network that was implicated in auditory processes, including phonetics, voice processing, processing pure tones and other sounds, speech perception, and musicianship. However, the network of this parcel was also associated with multisensory integration, comprehension, syntax, linguistics, vision, and orthography. The interpretation that this network has a strong connection to auditory processing is also based on its connectivity patterns from the preceding section.

Parcels within the 8-cluster parcellation scheme were decoded by correlating meta-analyses in Neurosynth with item-to-cluster-consensus maps indicating the average PAC value between each voxel in the PPS zone and each cluster in the parcellation (Figure 10, panel B). This presented some functional information for parcels which did not have easily decodable networks. For instance, the mpPT was more strongly associated with the planum temporale than other clusters. Its item-to-cluster consensus map also correlated more strongly than other parcels with the meta-analysis of the ventral premotor cortex, which may reflect its coactivation with vPMC, however, it did not show particularly strong functional connectivity to vPMC. The mpPT also showed modest associations with meta-analyses that were most strongly associated with the AA, supporting its role in auditory processing. Indeed, relatively high associations for the same set of meta-analyses were also found for alSTG, mirroring the pattern of FC to AA that we observed in the preceding section. Finally, the mpPT was exclusively associated with rehearsal, implicating it in speech production. The pdPPS parcel also had a network that was difficult to decode. Although this parcel was most strongly associated with 'comprehension' and 'sentence', other areas showed a strong association with these meta-analyses as well. However, pdPPS

exclusively associated with 'sentence comprehension' and word form. Further, it had strong associations with pMTG and syntax, and it was strongly anticorrelated with meta-analyses for planum temporale and auditory cortex. Finally, pdPPS was moderately related to other higher-language meta-analyses most strongly associated with pvSTG, including phonological processing and lexical-semantics. For many of the parcels, this analysis reinforced the same functional associations that the network-level decoding provided. For instance, it clearly implicated the auditory cluster in lower-level auditory processing. Notably, however, audiovisual processing was also associated with this parcel. The pPT subnetwork was exclusively associated with somatosensory processing, mirroring the decoding results for its network. Notably, this area was also exclusively associated with sensorimotor processing, suggesting the sensorimotor associations with the AA network were being driven by areas in motor cortex. The avSTG parcel was associated more strongly than other parcels with speech perception, constituting a part of the functional associations attached to its network. The alSTG parcel was associated with higher-level auditory processing more than the other parcels, but also exhibited associations to lower-level auditory processes, including processing speech sounds. Finally, the pvSTG parcel was associated with various aspects of higher-level language processing, including comprehension, lexical-semantics, phonology, phonetics, and word processing. It was also associated with the STS, which this parcel contained, as well as social processing.

Figure 10: Decoding parcels and their networks

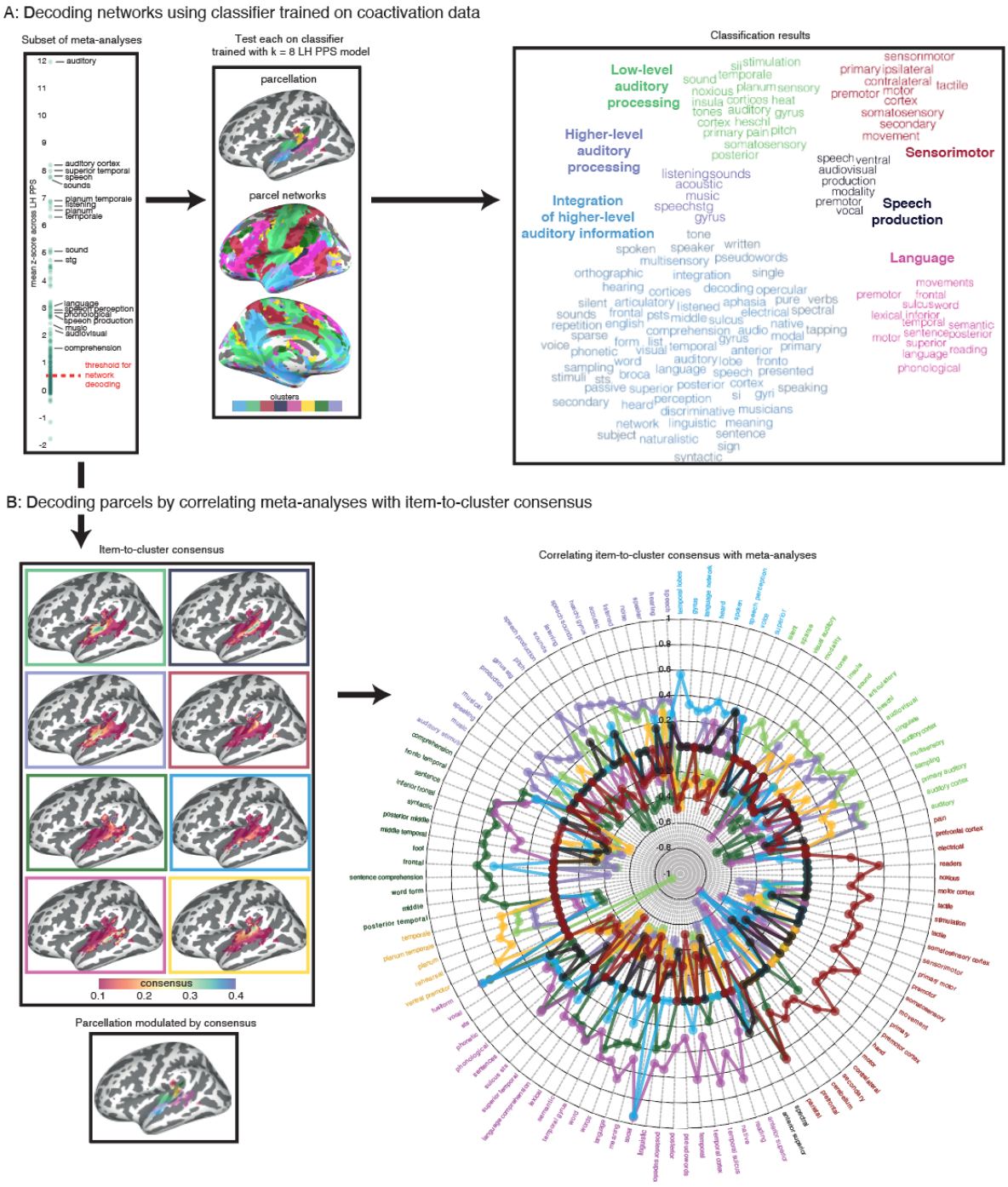

Figure 10: Decoding parcels and their networks. Panel A shows the process for decoding networks. In the leftmost box, mean likelihood of seeing activity within the PPS is shown for each meta-analysis in Neurosynth. A threshold of 0.8 was applied to remove a subset of meta-analyses unrelated to the PPS. The middle box represents the process of training a classifier to learn the parcellation 8-cluster parcellation scheme from coactivation data. The network depicted in this box is for illustrative purposes and not the actual coactivation

network that was learned by the classifier. The rightmost box presents the result of testing the classifier trained on coactivation data, on each of the meta-analyses from the leftmost box. The color of each word in the word clouds corresponds to the cluster that that particular meta-analysis was grouped with. Bolded words on the periphery of each cluster are interpretations of cluster function based on the entire word cloud for each cluster. Panel B shows the process for decoding parcels. The leftmost box shows item-to-cluster consensus maps for each cluster in the parcellation. They represent the reliability that a vertex can be assigned to a particular cluster. The rightmost chart shows the decoding analysis, performed by correlating each item-to-cluster consensus map with each meta-analysis in Neurosynth. Each meta-analysis on this chart is color coded based on the cluster to which it was most strongly associated with.

3.4 Discussion

3.4.1 General summary of findings and their implications

Because areas around premotor and motor cortex appeared to distinguish many parcels, we took a closer look at the connectivity between the parcels and a handful of ROIs that were selected to map onto some of the areas that we observed to distinguish parcels in the winner-take-all analysis (Figure 9).This included the eye fields associated with mpPT (FEF and PEF), and the area of vPMC (6r) that was associated with pPT. To try to capture the area of the dPMC connected more strongly with plSTG, we included area 55b, located at the transition between posterior middle frontal gyrus and premotor cortex, and dorsal laryngeal motor cortex (dLMC), which often contains a gyral component near area 55b (Bylek, 2021). In addition, we included an analysis of connectivity to auditory cortex as a contrast to the motor-driven ROIs. Briefly, areas FEF, PEF, 6r, and 55b were extracted from the HCP's MMP 1.0 atlas. The dLMC was defined by performing an activation likelihood estimation (ALE) meta-analysis of studies (N=9) reporting activity in dorsal laryngeal motor cortex. Most of these studies also found foci in ventral laryngeal motor cortex

(vLMC), and the resulting areas were separated into two ROIs. A high threshold ($p < 0.0001$) was applied to the meta-analysis in order to minimize overlap between area 55b and dLMC. First, this analysis showed that most parcels of the PPS zone were functionally connected to areas along premotor and motor cortex, as well as auditory cortex. The analysis also showed that most parcels had much higher connectivity to auditory cortex than premotor or motor cortex. The ROI analyses demonstrate that mpPT shows stronger connectivity to the eye fields, and that connectivity to the motor cortex ROIs is relatively high for AA, mpPT, avSTG, and plSTG. Interestingly, we note that parcel connectivity to auditory cortex could explain up to 96% variability in parcel connectivity to motor cortex (dLMC: $r=0.95$, $p < 0.01$, vLMC: $r = 0.98$, $p < 0.01$). Unfortunately, the ROIs for 55b and dLMC did not appear to capture the dPMC area most strongly connected to plSTG. Area 55b had relatively high functional connectivity to plSTG, but had higher functional connectivity to pvSTG, reflecting the greater overlap between 55b and the network of areas that activate most strongly for pvSTG (i.e., Figure 9A).

3.4.2 Comparing clustering approaches

Using consensus to measure the reliability of clustering approaches, we found spectral clustering and k-means to perform particularly well. Prior results have found that spectral techniques tend to have higher reproducibility when compared to k-means (Arslan et al., 2018; Thirion et al., 2014). That is, parcellations are relatively more similar when spectral clustering is applied to different scans of the same individual, or to bootstrapped data. We report slightly higher reliability in k-means clustering relative to spectral clustering, partly as a result of stabilizing each application of k-means by minimizing sum of squared errors

over 300 repetitions with different random initializations. For example, we found that k-means clustering produced significantly more reliable parcellation models than spectral clustering across our participants, even though these models did not significantly differ with respect to complexity. The decision to stabilize k-means in this manner is motivated by prior work showing high variability in k-means based parcellations and suggesting that roughly 250 repetitions are necessary to draw accurate anatomical boundaries using structural connectivity (Nanetti et al., 2009). The impact of repetition is evinced by the results we present for k-medoids, which shares many similarities with k-means, but was not stabilized by multiple repetitions in the current work. This algorithm showed remarkably poor reliability compared to k-means. One possibility is that k-means clustering was better-suited to the characteristics of the dataset used in the current research (e.g., multiecho resting state data). We recommend the consensus clustering framework for studies considering relying on k-means clustering, as nesting repetitions of k-means within a procedure that extracts the most stable partition will help to further stabilize the results. By extracting a clustering solution from the consensus across many repetitions of k-means clustering, the influence of the tendency for k-means to get stuck in local optima is further minimized. Indeed, we found consensus clustering with k-means using Pearson distance and eta^2 to be the most drastically impacted clustering approach by the consensus clustering framework. However, we note that none of the clustering approaches were dramatically impacted by the consensus clustering procedure individually. That is, the parcellations extracted from the consensus matrix, and the parcellation that would have been extracted without consensus clustering were qualitatively indistinguishable.

The fourth clustering algorithm that we tested was Louvain clustering, for which we also report relatively poor reliability. A caveat to the comparison between Louvain clustering and other algorithms is that in Louvain clustering, the number of clusters was automatically estimated from each subsample of the data. Thus, while reliability was computed for the other 3 algorithms independently for each predetermined number of clusters, in Louvain clustering, the reliability measure collapsed across solutions with varying numbers of clusters. This variability likely explains the poorer reliability that we observed. At the same time, larger parcellation complexities could not be extracted from the ensembles of Louvain parcellations. Thus, more substantial adjustments to the consensus clustering framework are necessary to create ensemble parcellations from Louvain clustering, and to evaluate its reliability relative to other clustering approaches.

Our internal reliability analysis of clustering approaches highlighted the fact the choice of distance measure may have a bigger impact on a parcellation than the choice of clustering algorithm. For example, we found that clustering group-level functional data using k-means and spectral clustering with the same distance measure produced parcellations with higher similarity (on-average) than using either of those clustering algorithms with different combinations of distance measures. However, for each clustering algorithm, different distance measures seemed to produce slightly more reliable parcellations. For k-means clustering, Pearson distance and eta^2 performed best while for spectral clustering, it was Euclidean distance. In general, we report high similarity between parcellations produced when clustering using Pearson distance and the eta^2 coefficient, which shares many similarities with Pearson distance, but has been developed for estimating similarity between whole brain connectivity maps (Cohen et al., 2008). Finally,

we observed that under very specific conditions (e.g., when clustering principal components of group-level functional connectivity using k-means clustering with Euclidean distance), clustering approaches could produce results that were highly reliable, but also systemless. Consequently, we caution that while reliability is valuable for understanding which parcellations are more likely to be reproduced by a clustering approach, it is nevertheless important to consider the plausibility of the underlying parcellation.

3.4.3 Clustering group-level data versus consolidating participant-level parcellations

The performance of different clustering approaches relative to each other was roughly the same when clustering either group-level data, or individual subjects. However, the parcellations themselves were highly dissimilar. We report much better results when clustering individual subjects and fusing the resulting parcellations. Qualitatively, this generated much smoother, and more contiguous clusters overall. Further, parcellations built up over subjects could be used to classify external data much more accurately. Thus, we encourage future parcellation work to avoid clustering over group-level data when possible.

We also describe a set of methods that make it tractable to consolidate participant-level parcellations, and even fuse together different clustering approaches in the process. In our validation of parcellations using external data, we showed that these methods performed favorably, including against existing parcellation schemes. The set of methods we promote can be used to select optimal parcellation complexities in an automated fashion for individual subjects. It can also adjust optimal participant-level parcellations

based on group-results, allowing for the identification of outlier participants where there is less evidence that a parcellation can be performed. Finally, in considering the unique functional anatomy of participants as expressed by their optimal parcellation complexities, we describe a way of determining the most complex parcellation that can be fit to the entire group, based on the ability of the group parcellation to reflect the unique features of participant-level parcellations.

The automated method we describe for selecting optimal parcellation complexities within the consensus clustering framework aligns with intuitions about good clustering solutions based on the distribution of consensus values within a consensus matrix. Given that studies using consensus clustering often make a qualitative decision about the optimal number of clusters supported by a dataset based on these intuitions, we hope that the automated procedure we describe can be used to more clearly articulate the basis for choosing a good model. Further, this automated procedure can be deployed in problems where many different datasets need to be clustered (e.g., subjects), and its performance can be fine-tuned to a particular dataset based on simple parameters. For example, the tolerance threshold that determines the point at which certain measures of reliability in consensus clustering begin to poorly reflect important properties of the consensus distribution can be adapted to different datasets and clustering approaches.

3.4.4 Comparing the parcellation of the PPS zone to architectonic work

The spatial organization of the parcellation we promote shows consistency with architectonic descriptions of cortical fields underlying the PT, as well as neighboring areas

of the STG. One cluster that we found posterior to Heschl's gyrus (HG) on the superior temporal plane (referred to as middle-to-posterior PT or mpPT) was consistent with the location of areas PaAe (Fullerton & Pandya, 2007; Galaburda & Sanides, 1980), PA (Rivier & Clarke, 1997; Wallace et al., 2002), and especially Te2.2 (Morosan, 2005a,b; Clarke & Morosan, 2012; Zachlod et al., 2020). Consistency was higher with area Te2.2 than others primarily because the work that has mapped these areas has separated cortical fields on the superior temporal plane from cortical fields within lateral STG. The functional connectivity data reported here suggests also that areas on the superior temporal plane do not extend into lateral STG. Here, we found two clusters on the lateral STG, one anterior (anterolateral STG; alSTG) and one posterior (posterolateral STG; plSTG). Together, the site of these clusters is consistent with the location of area Te3 (Morosan, 2005a,b; Clarke & Morosan, 2012; Zachlod et al., 2020) and STA (Rivier & Clarke, 1997; Wallace et al., 2002). Coincidentally, area Te3 has been mapped in studies using observer-independent architectonic parcellation methods, bearing some similarity to the data-driven approach for generating parcellations that we promote here.

This observer independent work has additionally identified two areas ventral to Te3 that are in the superior temporal sulcus (STS). One of these areas overlaps with the upper bank of the STG. Although our ROI did not extend far enough to capture the most ventral of these two STS areas, we do find evidence for the more dorsal area, called STS1. In our data, the corresponding parcel is referred to as anterolateral STG (avSTG) because it appears to extend into a sizeable area of ventral STG. In our work, we also found an area in posteroventral STG (pvSTG), but this portion of cortex has not been mapped by the observer-independent architectonic work described above. The point of separation

between posterior and anterior areas on the STG was found to align in the current work, suggesting that there may be a broader transition zone between multiple cortical areas in mid-to-posterior STG (i.e., pvSTG versus avSTG, and alSTG versus plSTG).

We also found multiple clusters that shared spatial correspondence with descriptions of the location of area Tpt. A posterodorsal cluster within the PPS (pdPPS) was consistent with the description of area Tpt as intruding on SMG and located at the transition of temporal and parietal cortex (Fullerton & Pandya, 2007; Galaburda & Sanides, 1980). Notably, this cluster overlapped with the PT, but it is referred to here as the pdPPS to clarify that it intruded into adjacent anatomical areas. At the same, Tpt has been shown to extend into pSTS, and the pvSTG cluster that we found was consistent with descriptions of "sulcal Tpt". Further, area Tpt extends into STG and therefore may partly correspond with the aforementioned plSTG cluster that we mapped. Area Tpt has been indicated to extend into portions of the parietal operculum (PO) as well. Our parcellation described a subnetwork of areas that included the PO, a portion of posterior PT (pPT), and temporooccipital middle temporal gyrus (toMTG). When we investigated the functional connectivity networks associated with each of the areas in this subnetwork, we found substantial differences in connectivity, but particularly for toMTG. Given that the pPT area of this subnetwork also aligns with the location of Tpt (i.e., transition between temporal and parietal cortex), it's possible that the pPT and PO areas that we have mapped reflect a single cortical field that was spatially disconnected in our parcellation. Indeed, for many parcellations that we did not focus on, we found these two areas to be connected as a single cluster. Thus, it's possible that by collapsing over participants or clustering approaches, a

specific portion of the superior pPT that connects these areas was lost in the service of representing more consistently identified parcels.

3.4.5 Comparing the parcellation of the PPS zone to other functional neuroimaging atlases

The parcellation of the PPS zone that we promote was benchmarked against several commonly used atlases to ensure that it has good generalizability. Testing how well these different parcellations could be used to classify external coactivation data showed that our parcellation performed favorably, and even slightly better than the others. Relative to these other atlases, our parcellation had several distinct features. For one, none of the tested atlases showed a distinction between anterior and posterior segments of lateral STG. In the HCP MMP atlas, lateral STG was splintered into a dorsal and a ventral cluster. In the AICHA atlas, a cluster remarkably similar to anatomical PT was identified, but no evidence of further segmentation within this area was reported, even though this atlas contained twice as many regions in the PPS zone compared to our parcellation. All of the atlases, as well as our parcellation, identified posteroventral and anteroventral STG. The AICHA atlas showed a similar clustering of the PO and posterior PT to what we observed (i.e., a kind of subnetwork). On the other hand, the HCP MMP atlas distinguished posterior PT from PO. The mpPT cluster we report was roughly segmented into two clusters in the HCP MMP atlas, a dorsal and a ventral one. Both of the functional atlases exhibited stronger parcellation complexity anterior to HG. All of the atlases, as well as our parcellation, picked out a pdPPS cluster, although it appeared to be located slightly more ventral in the HCP MMP atlas. Finally, both HCP MMP and AICHA atlases distinguished pvSTG into anterior

and posterior clusters, which we do not report in our parcellation. We do note, however, that similar clusters emerged at higher parcellation complexities in our data, but these parcellations were deemed slightly too unreliable in our model-selection process. We caution that it is difficult to compare atlases in an entirely unbiased way but emphasize that the parcellation we promote clearly achieved very good generalizability overall. For example, the HCP MMP atlas may perform better on data modeled in surface space and may generalize better to some modalities than others because it is a multimodal atlas. In this work, this atlas had to be projected into volume space (but using tools developed by HCP itself), undoubtedly introducing some inaccuracies. The data we used to test the atlases may also be a factor in atlas performance. We show that our parcellation better predicts coactivation data across task-fMRI studies. It's possible that the AICHA atlas would perform better on functional connectivity data, although we believe this is unlikely because prior work has established strong similarity between functional connectivity and coactivation data. In any case, the results we report suggest that the parcellation promoted here would serve future task-fMRI work much better. This result aligns with the overarching goal of this work—to understand how prior task-fMRI results implicating the PT reflect its functional organization.

3.4.6 Networks in the PPS zone

Connectivity to several large networks distinguished areas in the PPS zone. Although our parcellation was not able to discriminate auditory cortex from other adjacent early auditory areas, the large auditory cluster (auditory area; or AA) that we found exhibited stronger connectivity to a motor and somatosensory network that closely followed

previous work mapping connectivity in auditory cortex (Yuan et al., 2018). When we compared connectivity between parcels as a function of moving from the dorsal to the ventral portions of the PPS zone (i.e., comparing AA and superior temporal plane areas; alSTG and plSTG; avSTG and pvSTG), we found that anterior clusters in the PPS zone showed relatively stronger connectivity to the same set of areas that were most strongly connected to AA, particularly auditory cortex, the insula, and motor cortex. Thus, the PPS zone appears to be partly organized around connectivity to the areas on the auditory *network*. This finding explains the basis on which our parcellation appears to have defined an anterior-posterior transition point along the STG. Although prior architectonic and connectivity parcellation work has distinguished such a transition point in ventral STG and on the superior temporal plane, this study is the first to show that such a transition point also exists on lateral STG and follows the same organizational principle that distinguishes anterior and posterior areas in other portions of superior temporal cortex.

Posterior areas of the PPS zone also showed a similar connectivity pattern when compared to their more anterior counterparts—particularly to portions of dorsal premotor cortex, inferior frontal gyrus (IFG), ventral premotor cortex (vPMC), and mid-to-posterior STS. Each of the posterior clusters (i.e., mpPT, pPT-PO-toMTG, pdPPS, plSTG, pvSTG) was more strongly associated than *any* of the other clusters with different portions of this network. For example, mpPT and plSTG were most strongly connected to different portions of dPMC, the pPT-PO-toMTG subnetwork was most strongly connected to vPMC (though the toMTG area alone showed much lower connectivity there than the rest of the subnetwork), pdPPS was most strongly connected to an anterior portion of posterior

inferior frontal gyrus, and pvSTG was most strongly connected to most of IFG, as well as the STS.

We found densely concentrated differences in connectivity to parcels in different portion of dorsal premotor cortex (dPMC). For example, mpPT was most strongly connected to the eye fields (frontal and premotor), while plSTG was most strongly connected to an area of dorsal premotor cortex (dPMC) sandwiched in between the eye fields. Curiously, this dPMC area aligned with the portion of dPMC that was significantly associated with studies on the PT, suggesting that some functional neuroimaging studies treating functions in the PT may be tapping into a plSTG-dPMC network. This portion of the dPMC was adjacent to the dorsal laryngeal motor cortex (dLMC) mapped in previous functional neuroimaging studies. Notably, activity for dLMC has been often found in separate sulcal and gyral portions (e.g., Belyk et al., 2021), and the location of this dPMC area is relatively close to gyral dLMC. At the same time, pvSTG expressed preferential connectivity to an area anteriorly abutting the dPMC associated with plSTG. This area transitions between posterior middle frontal gyrus (pMFG) and dPMC. It was aligned with the location of area 55b as delineated in prior work, which has been defined myeloarchitectonically (but also using myelin estimates in vivo) based on the fact that it is a virtual island of low myelination in an otherwise densely myelinated area of cortex (Glasser et al., 2016).

3.4.7 Functions in the PPS zone

Functions of the PPS zone were decoded by correlating the reliability of cluster assignments to the PPS zone with the likelihood of seeing activity in the PPS zone for

different meta-analyses generated with Neurosynth. The networks of these clusters were decoded as well using the trained classifier models that we used to benchmark our parcellation against other atlases. That is, by testing the classifiers trained to distinguish between parcels based on coactivation networks (i.e., activity patterns that cooccur across all functional neuroimaging studies) on different meta-analyses. Auditory processing within the PPS appears to be organized by levels of complexity, starting with lower-level auditory analysis in AA (which includes anterior portions of the PT), more complex auditory analysis of speech sounds and music in the alSTG, and higher-level speech and voice processing in the avSTG. Remarkably, network-level decoding converged on roughly the same functional assignments for each cluster. The only exception to this was avSTG, which took part in a functionally broader network that we speculate to be involved in integrating auditory information with a wide range of other processes.

It is worth pointing out that our parcellation grouped together portions of anterior PT and auditory cortex. Given that we found scarce evidence of early auditory processing within other portions of the *PT* (i.e., not anterior PPS zone areas), this suggests that auditory response commonly attributed to the PT in task-fMRI work mapped onto our auditory parcel. The close correspondence between functional connectivity of auditory cortex and anterior PT suggests that anterior PT is involved in general auditory processing. This may be consistent with the computational hub hypothesis that postulates the PT is involved in processing complex sounds, but crucially only within anterior PT (Griffiths & Warren, 2002).

Two parcels within the PPS zone were explicitly linked to speech. One of these parcels—more clearly within the PT—showed a connectivity pattern consistent with area

Tpt as mapped in prior work in primates. Area Tpt has been shown to be connected to portions of posterodorsal prefrontal cortex near the eye fields (Hackett, 2015; Chavis & Pandya, 1976; Smiley et al., 2007) and here we found area mpPT to show strong connectivity to the frontal and premotor eye fields. Based on neuron recordings, work with primates has shown that area Tpt is multisensory, lacks tonotopic organization, and contains different distributions of neurons that respond to head rotation and sound location (Leinonen et al., 1980). These response properties may explain the connectivity pattern of area Tpt, which may serve to facilitate head movement and sound localization. Work in humans has implicated some portions of the PT in spatial hearing (Isenberg et al., 2012). However, despite the connectivity correspondence between Tpt and mpPT, we do not find mpPT to correspond to spatial hearing. Instead, it associates with speech rehearsal. Prior task-fMRI work has mapped an area with similar multisensory response properties to Tpt within a more posterior portion of PT that appears to be important for speech production. This area Spt is similar to Tpt in so far as it exhibits a strong auditory response and performs a computation that involves integrating sensory and motor information, but for a different motor effector—the larynx (see introduction for further support; also Hickok, 2017 for review). Curiously, the area mpPT that we have mapped here does not appear to correspond to area Spt, which has a more posterior location and a dramatically different connectivity and response profile (Isenberg et al., 2012). For example, here we find that mpPT does not associate with speech, music, speech perception, speech production, or sounds, but it does associate with hearing, tones, and production. We also found that its network was difficult to decode and did not appear across task-fMRI studies. We speculate that one possibility is that mpPT may be involved in head rotation, which

may explain why task-fMRI studies have not managed to find this network consistently enough for us to be able to decode it. The association of this area with rehearsal, but not overt speech production, suggests that it may be involved speech imagery (Price, 2012) or processing self-generated speech (Tremblay et al., 2013). Perhaps this area is also involved in spatial hearing, although we do not see explicit evidence of this here.

Within the PT, task-fMRI studies have dissociated self-generated speech from sensorimotor integration processes for speech (Tremblay et al., 2013) and spatial hearing from sensorimotor integration processes for speech (Isenberg et al., 20120). Here, we report that mpPT, putatively involved in self-generated speech and possibly spatial hearing, dissociated from an area involved in sensorimotor integration and an area involved in speech production. Thus, overall, our results appear to corroborate prior task-fMRI work but add more complexity to the functional organization of the PPS zone. Speech production studies more broadly were associated with the plSTG, as well as the plSTG network. Thus, the dPMC area near dLMC that distinguished this cluster appears to be an area commonly reported in speech production studies. Given that the same dPMC area was significantly activated in studies of the PT (i.e., the analysis we used to define the PPS zone), it appears to be the case that some studies that have implicated the PT in speech production are localizing the plSTG network. We suspect that the dPMC area associated with the plSTG represents the dLMC. We note that while we attempted to construct dLMC and vLMC regions of interest within which to probe functional connectivity in the results, these efforts are a coarse approximation of the actual location of these areas. Notably, plSTG showed some association with processing speech and speech sounds, as well as phonological information. In combination with its connectivity near dLMC, this makes it a

possible candidate for area Spt, which acts as an interface between auditory-phonological and articulatory networks (Hickok, 2017). However, it is also possible that sensorimotor integration within the pPT area reflects Spt. Because pPT is such a small portion of the pPT-PO-toMTG network that we mapped, we were not able to decode this area separately. As such, future work is necessary for understanding the functional differences between pPT and plSTG. One possibility that has been previously suggested is that posterior temporal cortex may be organized in a similar way to visuomotor integration areas in the intraparietal sulcus (Hickok, 2017). That is, it may contain multiple areas that perform a similar computation specialized for different motor effectors. For example, speech production requires the coordination and control of not just the larynx, but also supralaryngeal articulators, and respiration. Thus, it may be the case that pPT performs a similar kind of computation, but not for controlling the larynx. Critically, the connectivity of pPT with ventral premotor cortex is most consistent with descriptions of structural connectivity for Spt (Isenberg et al., 2012). Curiously, comparisons in connectivity between macaques and humans has indicated that, in humans, area Tpt is much more strongly connected to the same portion of vPMC that we found to be associated with the pPT and PO (Neubert et al., 2014). Thus, ultimately, it is somewhat difficult to resolve which areas in our parcellation correspond to area Tpt, and it may be the case that Tpt is comprised of multiple functional areas.

Two clusters in the PPS zone also appeared to be associated with higher-level speech processing. The pdPPS was much more strongly associated above other clusters in syntax, sentence comprehension, and word form, while the pvSTG cluster was much more strongly associated with reading and social cognition. Overwhelmingly, the strongest

associations for pvSTG were related to language and speech processing, but particularly phonological and lexical processing. This aligns with the fact that pvSTG contains the pSTS, an area important for phonological processing (Hickok & Poeppel, 2007). Critically, most of the associations between pvSTG and pdPPS clusters overlapped, suggesting they were both involved in higher-level speech and language processes. Despite these functional associations with higher-level language processing, the pdPPS was connected more strongly than other areas to a very focal portion of anterior inferior frontal gyrus, which may play a role in speech production (Hickok & Poeppel, 2007). The network for this area did not map onto any task networks from the functional neuroimaging literature. We note that a strikingly similar area to pdPPS has been parcellated in prior work and related comprehension, similar to our findings here (Glasser et al., 2016). It is possible that this area exhibits functional connectivity that is substantially different from the task networks in which it has been observed to participate.

Chapter 4

Meta-analytic and structural connectivity based parcellations of left temporal and inferior parietal cortex for localizing speech perception

4.1 Introduction

4.1.1 The direction of the ventral stream pathway

Auditory comprehension involves transforming spectrotemporal information in the continuous acoustic signal into speech sound representations that can be used for auditory word recognition. While the initial stages of acoustic analysis unequivocally unfold bilaterally around Heschl's gyrus (e.g., Barton et al., 2012), there are differing views about whether the ventral processing stream supporting speech perception is directed anteriorly towards mid-to-anterior superior temporal gyrus (Rauschecker & Scott, 2009; DeWitt & Rauschecker, 2012), or posteriorly towards mid-to-posterior superior temporal sulcus (Belin & Zatorre, 2000; Hickok & Poeppel, 2007; Hickok & Poeppel, 2016). This disagreement reflects a major point of divergence in dual stream models of auditory and speech processing. One model for auditory processing rooted in nonhuman primate anatomy suggests that primary auditory cortex divides anterior ventral stream regions for

auditory object identification from posterior regions that form the dorsal stream for spatial hearing (Rauschecker & Tian, 2000). Under this model, speech perception involves increasingly anterior areas to auditory cortex, while speech production involves posterior sensorimotor areas that facilitate spatial hearing (Rauschecker & Scott, 2009). An alternative model for speech processing based in functional neuroimaging and lesion work places the point of divergence between the ventral and dorsal streams much further posteriorly around the posterior superior temporal sulcus (pSTS; Hickok & Poeppel, 2007; 2016). A posterior location for speech perception undermines work with primates showing selectivity for species-specific calls in regions of the anterior belt (Rauschecker & Tian, 2000). At the same time, an anterior location for speech perception cannot account for the association between left posterior temporal lobe damage and auditory comprehension deficits, as well as the strong response in pSTS that has been documented by functional neuroimaging studies for tasks which require processing phonemic information (Hickok & Poeppel, 2007).

Although functional neuroimaging has generally established that listening to speech sounds generates robust response in superior temporal cortex, the results of efforts to localize speech perception more precisely have varied. For instance, studies that have compared passive listening to sentences with passive listening to sounds that match the acoustic complexity of speech have typically implicated anterior STS (aSTS; Evans et al., 2016; Pelle et al., 2010; McGettigan et al., 2012; Narain et al., 2003; Scott et al., 2003; Spitsyna et al., 2006). At the same time, other studies have shown that pSTS is more sensitive to intelligible speech and less sensitive to acoustic variation compared to aSTS (Okada et al., 2010). Further, response in more posterior portions of STS has often been

associated with processing larger amounts of phonemic information (e.g., Okada et al., 2006; Liebenthal et al., 2005; Specht et al., 2009).

4.1.2 Meta-analysis of speech perception

Meta-analyses have sought to resolve some these discrepancies but have sometimes themselves produced unclear results. For instance, one study compared activity across 19 functional neuroimaging studies that contrasted sublexical speech perception tasks with speech control stimuli (Turkeltaub & Coslett, 2010). Across these studies, they found significant response within a large area of the superior temporal gyrus (STG) and STS that extended into regions both anterior and posterior to Heschl's gyrus (Turkeltaub & Coslett, 2010). When the meta-analysis was restricted to the 6 experiments (note, not studies) that contrasted speech stimuli with more closely matched speech-like control conditions, an area clearly within mid-to-posterior STS emerged as the site of phonetic processing. In addition, a contrast meta-analysis between experiments that used explicit decision tasks (n=13), which drive attention to phonological information, and experiments that used passive monitoring tasks (n=10) showed involvement of a slightly more posterior portion of the STS, indicating that speech perception processes are directly posteriorly. Another meta-analysis pooled 49 speech perception studies that were more heterogenous, focusing on phonetic, word-level and sentence-level speech stimuli contrasted against non-speech acoustic controls (DeWitt & Rauschecker, 2012). The results were consistent with prior work, implicating a large region spanning anterior to posterior portions of the STG and STS in speech perception (DeWitt & Rauschecker, 2012). However, the likelihood of seeing activity in these studies was somewhat higher in the STG than the STS. Further, the same

meta-analytic study took a more careful analysis approach, separately examining experiments that relied on speech stimuli of different lengths. In this analysis, although stimuli of all lengths showed some likelihood of activating mid-to-posterior STS, there was evidence for an anteriorly directed speech processing pathway. Phoneme length stimuli (n=14) consistently produced activity in mid-STG, but word-length stimuli (n=16) consistently produced activity in a slightly more anterior area, and sentence-level stimuli (n=19) in an even more anterior area.

4.1.3 Considering sample sizes and domain specificity in meta-analysis

Clearly meta-analyses have converged on the importance of mid-STG/STS for phonetic processing. However, the extent to which more anterior and posterior regions are involved in phonetic and phonological processes remains unclear. One weakness of prior meta-analytic work is the lack of robust sample sizes. While tens of functional neuroimaging studies can add up to hundreds of subjects, the effects reported for each study are purely associated with a single voxel, which limits the spatial resolution of the analysis (e.g., Gorgolewski et al., 2015; Samartsidis et al., 2017). As a result, it's unclear how effective conventional meta-analysis can be for accurately delineating functional boundaries (Salimi-Khorshidi et al., 2009; Van Essen, 2009). For example, in the most common framework for meta-analysis, the foci attached to each study is modeled by a Gaussian probability distribution with width based on inter-subject and inter-study variability (Eickhoff et al., 2012). Empirical work on this approach suggests that at least 17 experiments are necessary to prevent a single experiment from biasing the results, and at least 20

experiments are necessary to achieve sufficient power for moderate effects (Eickhoff et al., 2016). These *minimal* sample size recommendations complicate the more specific meta-analyses that studies have carried out, and which have implicated both posterior and anterior areas in speech perception. Critically, these sample sizes also assume exceedingly high homogeneity in the experiments being analyzed, and greater sample sizes are required when there is more variance in tasks, or even in the sample sizes of the individual experiments being analyzed (Eickhoff et al., 2016).

Prior meta-analytic work has also focused on a relatively narrow definition of speech perception and has not mapped the extent to which areas that respond to speech participate in other functions. A wide range of speech tasks presumably entail speech perception and activate lower-level speech areas involved in phonetic and phonological processing. Studying these tasks and their associated patterns of brain activity in relation to each other can be useful for understanding how speech areas are hierarchically organized. At the same time, most meta-analyses of speech tasks are highly domain-specific and do not consider whether the implicated areas might reflect non-speech functions. For instance, mapping speech perception using sublexical tasks has been argued to involve executive control and working memory (Hickok & Poeppel, 2007; 2016). Similarly, anterior regions activated for speech perception tasks that involve sentences or phrases may reflect syntactic, combinatorial, or prosodic processes (Hickok & Poeppel, 2007; Hickok & Poeppel, 2016). Parallel issues arise for linguistic processes like syntax (Campbell & Tyler, 2018). A more systematic comparison between patterns of brain activity and different tasks is particularly important for understanding which functions are performed within the anterior and posterior portions of STS and STG, which have not only been associated with

speech perception, but also a wide range of other processes that include semantics, orthography, biological motion perception, face processing, theory of mind, audiovisual integration, etc. (e.g., Hein & Knight, 2008; Liebenthal et al., 2014; Venezia et al., 2017; Wilson et al., 2018). Thus, mapping the boundaries of speech perception areas situated in such a functionally dense neighborhood requires considering patterns of activity not just associated with speech, but also a range of other processes.

4.1.3 Current Work

In this work, we aimed to extend prior meta-analytic efforts attempting to localize speech perception by considering the issues outlined above, chiefly concerns about small sample sizes and domain specificity. We take on this task in order to better understand whether speech perception is directed anteriorly towards anterior superior temporal gyrus (STG) or posteriorly towards mid-to-posterior superior temporal sulcus (STS) as specified by different dual stream models. In order to improve the sample size of our meta-analysis, we relied on Neurosynth, a database that has used text-mining techniques to amass tens of thousands of functional neuroimaging studies. In Neurosynth, information about word frequencies within the abstracts of published neuroimaging studies is extracted and associated with the activation foci that studies report (Yarkoni et al., 2011). Consequently, frequently used phrases in the neuroimaging literature act as an index for behavior. Critically, this approach has been validated against much smaller databases that catalogue published functional neuroimaging studies more carefully (Yarkoni et al., 2011; Poldrack & Yarkoni, 2016). Aside from increasing sample sizes relative to prior meta-analytic work, we also sought to delineate areas involved in speech perception in a way that was less domain

specific, reflecting the role of areas involved in speech across the entire functional neuroimaging literature. In a conventional meta-analysis, a functional network is mapped from a small group of studies that are typically hand-picked to reflect a specific cognitive process. Here, we took a radically different and data-driven meta-analytic approach by extracting unique patterns of brain activity across many different behavioral interventions, allowing us to isolate patterns of activity associated with speech perception, while at the same time characterizing the broader behavioral context in which these activation patterns have been observed. Specifically, we compared patterns of activity in temporal and temporoparietal cortex associated with the 3,107 features in Neurosynth (i.e., frequently used phrases in the functional neuroimaging literature that serve as a proxy for cognitive processes). By gauging the similarity between activity patterns associated with each pair of features, we extracted a set of abstract, lower-dimensional functional networks that represent the structure of brain-behavior relations embedded in this database. Because we defined unique functional networks, the individual brain areas that comprised these networks could be shared. Speech related functional networks were compared to each other directly in order to understand which areas contributed uniquely to different aspects of speech processing. However, to systematically appreciate how speech networks related not just to each other, but to other domains, we performed a functional parcellation of the temporal and inferior parietal cortex. This allowed us to define functional areas that exhibited similar association patterns across the functional networks that carved up the behavioral space. Thus, we could identify speech perception areas in a more data-driven way that accounted for the fact these areas may be involved other domains. In order to validate the boundaries of functional areas defined in this way, a similar data-driven

approach was applied to structural connectivity data to test if it could independently recover the same functional boundaries. If the functional and structural areas that were mapped converged, it would suggest that the subtle differences in patterns of activity in temporal and parietal cortex that are reported across the functional neuroimaging literature are grounded in tangible differences between structural networks in the brain.

4.2 Methods

4.2.1 Region of interest

The primary objective of this study was to understand which portions of superior temporal cortex are associated with speech perception. As such, we constrained our analysis of functional networks to left temporal cortex. Restricting the area over which the multivariate meta-analysis of the current study was performed was additionally necessary to make the connectivity-based parcellation that acted as a validation of functional areas more tractable. However, we did expand our region of interest (ROI) to include neighboring portions of cortex that have been implicated in speech perception. The inferior parietal cortex was included in the ROI on the basis that regions in the vicinity of supramarginal gyrus have been implicated in the maintenance of phonemic information and the activation of phonological representations (Binder, 2017; Hickok & Poeppel, 2016). The region of interest (ROI) for this study was constructed from several anatomical atlases. The MNI structural atlas (Mazziotta et al., 2001) was used to define the main area of temporal cortex. Regions surrounding temporal cortex were extracted from the Harvard-Oxford atlas (Desikan et al., 2006), and included the parietal operculum, posterior

supramarginal gyrus, anterior supramarginal gyrus, angular gyrus, and central operculum. The central operculum was included due to its proximity to dorsal superior temporal gyrus. The resulting ROI was comprised of over 23,000 voxels (see Figure 13).

4.2.2 Meta-analysis

The neurosynth framework and database (Yarkoni et al., 2011) was used to generate meta-analyses for frequently used phrases in functional neuroimaging studies. The Neurosynth database is compiled by scraping abstract text and activation foci from published neuroimaging work, representing data from over 14,000 studies and over 3,000 phrases. In order to construct lower-dimensional functional networks, we performed a separate meta-analysis for each of these phrases, then analyzed the pairwise similarities between them.

Each phrase was meta-analyzed by comparing foci within the group of studies frequently using the phrase, to the foci of the remaining studies in the database (i.e., those that didn't frequently use the phrase). Frequently using a phrase in this context referred to using it at a rate of more than 1/1000 words within an abstract. This threshold sufficiently controls for studies using phrases incidentally (Yarkoni et al., 2011). Meta-analysis was performed using a two-way chi-square test for independence over contingency tables generated at each voxel in the brain (see Figure 13 for visualization). That is, at every voxel, association between the presence or absence of the phrase, and the presence or absence of activation foci was tested for the two groups of studies. Chi-square values were then converted to z-scores. All z-scores outside of our temporal and inferior parietal cortex ROI were excluded from analysis. By comparing the set of studies that use a phrase frequently to all other studies in the database, these "reverse inference" meta-analytic maps provide

more information about the specificity with which a brain region responds (see also Chapter 2 for additional details). That is, a conventional meta-analysis models the consistency of activity reported across a set of studies reflecting some cognitive process. However, an area that is consistently activated within one set of studies may be consistently activated across many other sets of studies representing other cognitive processes. By more carefully comparing a selected group of studies against the other studies in the database, the Neurosynth framework maps brain-behavior associations at each voxel that better reflect the specificity with which a particular cognitive process elicits activity there. One disadvantage to the Neurosynth approach is that the set of phrases it contains are highly redundant (e.g., 'default', 'default mode', 'dmn'), and can be relatively abstract. For example, studies frequently using the phrase 'speech perception' and 'speech production' also frequently use the phrase 'speech'. Here, we organized Neurosynth phrases into groups by analyzing the similarity between the patterns of activity associated with each pair of phrases.

4.2.3 Measuring similarity between meta-analyses

Similarity between meta-analyses was evaluated using distance correlation, a measure of statistical dependence that considers both linear and nonlinear associations (Székely et al., 2007). Using distance correlation, we generated a symmetric similarity matrix capturing the statistical dependencies between each pair of frequently used phrases in Neurosynth (3,000+).

Distance correlation was used as the measure of similarity between meta-analyses to increase sensitivity to complex relationships. For instance, two meta-analyses might

produce networks with some small overlapping brain regions, and some larger unique regions. The likelihood of seeing activity within voxels near the shared regions might increase proportionately for the two networks. However, the voxels near the regions unique to each network might show an inverse or more complex relationship, whereby the likelihood of seeing activity for one set of studies increases, while the likelihood of seeing activity for the other set either decreases or remains very low. If both linear and nonlinear relationships between these two networks are considered, they would be deemed more similar.

An analysis of the impact of considering non-linear relationships is presented in Figure 11. Using distance correlation to investigate the relationship between studies frequently using the phrase speech perception and other phrases in Neurosynth resulted in a much stronger relationship with studies on orthography (Figure 11, panel A). Looking at these two meta-analyses clearly revealed regions where activity was high for one set of studies and low for the other (i.e., inferior temporal gyrus for orthography and portions of dorsal superior temporal gyrus for speech perception), and where activity was high for both sets of studies (i.e., around posterior superior temporal sulcus and posterior superior temporal gyrus). A similar analysis was carried for other frequently used phrases in order to understand which were most impacted by non-linear dependencies (Figure 11, panel B). A number of these corresponded to domain-general systems (e.g., concentration, flexibility, incentive, nogo, prediction error), methods (e.g., cortical thickness, voxel-based morphometry), and intriguingly, complex psychiatric and neurologic symptoms (e.g., dementia, impulsivity, substance abuse).

Briefly, computing distance correlation involves generating a matrix of pairwise Euclidean distances between samples for each of the two vectors being evaluated. Each of the resulting two distance matrices are then double centered so that row and column means are subtracted, and the grand mean is added. Distance covariance is then calculated between the two double centered distance matrices $\mathcal{X}$ and $\mathcal{Y}$ with elements *(i,j)* by:

$$\mathcal{V} = \sqrt{\frac{1}{n^2}\sum_{i,j=1}^{n} \mathcal{X}_{ij}\,\mathcal{Y}_{ij}}$$

Much like in the case of Pearson correlation, distance correlation is then computed by dividing distance covariance $\mathcal{V}$ by the square root of the product of the distance variances for $\mathcal{X}$ and $\mathcal{Y}$ (i.e., distance variance for $\mathcal{X}$ and $\mathcal{Y}$ is calculated as distance covariance of the matrix with itself). Distance correlation is always positive and exactly zero when there is no dependence between the two variables being compared.

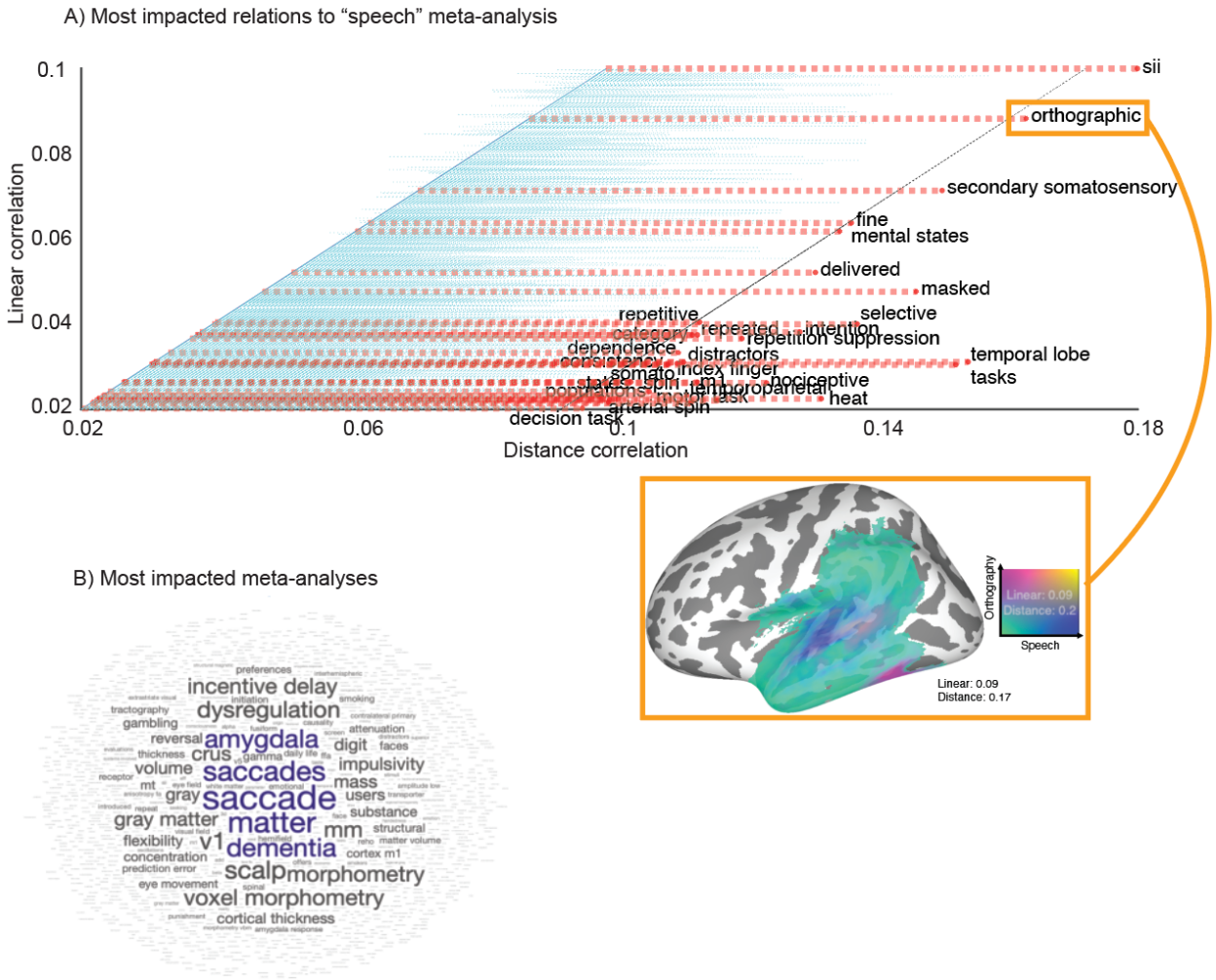

Figure 11: Non-linear relationships between meta-analyses. Panel A presents a scatter plot between linear correlation and distance correlation for the meta-analysis for speech, and every other meta-analysis in Neurosynth. The diagonal dark blue line represents identical results for pearson correlation and distance correlation. Each horizontal line connected to this diagonal dark blue line represents the distance of an actual point on the scatter plot to this hypothetical identical result. Lines to the right of the diagonal show stronger distance correlation than linear correlation. Horizontal lines depicted in red represent a selection of the most impacted relationships to speech. The meta-analysis for speech and orthography is presented in the yellow box just below, illustrating that these two meta-analyses share a region of overlapping activity. Panel B shows the same kind of analysis extended to all meta-analyses in Neurosynth (i.e., not only for speech). This word cloud shows which meta-analyses were overall more impacted by using distance correlation. The size of the word corresponds proportionally to increase in mean association strength between that meta-analysis and all other meta-analyses (as a result of using distance correlation). The handful of the most impacted meta-analyses are colored in blue.

4.2.4 Grouping meta-analyses

The similarity matrix comprised of pairwise distance correlations between meta-analyses was clustered with affinity propagation (Frey & Dueck, 2007) to produce functional categories comprised of groups of phrases that show similar statistical dependencies. These functional categories were then used to define lower-dimensional functional networks by reducing the dimensionality of all of the meta-analyses belonging to a functional group (see Figure 13 for visualization).

Affinity propagation is a clustering approach that is exemplar-based, which allowed us to define the most representative phrase (i.e., meta-analysis) within each functional category (i.e., group of meta-analyses), thereby facilitating the interpretation of the groups of phrases that were defined by the clustering. Although other exemplar-based clustering approaches exist (e.g., k-medoids), they are typically more sensitive to initialization parameters (e.g., Frey & Dueck, 2007; Lashkari & Golland, 2007). Indeed, the reliability analysis that we performed of different clustering approaches in Chapter 3 confirms this. The procedure for affinity propagation involves passing two types of real-valued messages between data samples in order to group samples according to their most representative exemplar. Initially, all nodes are considered candidates for exemplars but over many iterations a number of core exemplars emerges. For each iteration of the message-passing process, responsibilities are sent from samples to potential exemplars to determine how well each exemplar is suited to represent the sender. Following this, availabilities are sent from the potential exemplars to samples in order to gather evidence for how appropriate it would be for each sample to choose this particular exemplar given the preferences of other

samples. Availabilities and responsibilities between each pair of samples are initialized to zero and updated over iterations using a damping factor to control for the amount of influence exerted by update. The responsibility between sample *i* and potential exemplar *k* is defined as the similarity between these two samples minus the maximum of the sum of availability and similarity of all other potential exemplars to the same sample. Availability between potential exemplar *k* and sample *i* is defined as the minimum value between zero and the responsibility that exemplar *k* assigns to itself plus the sum of responsibilities that other samples assign to *k* (other than *k* or *i*). After availabilities and responsibilities are updated across some number of iterations, the final exemplars for each sample *i* are chosen based on which exemplar produces the highest sum of availability and responsibility for that sample.

The number of clusters produced by affinity propagation is not explicitly chosen by the user. However, a preference value is assigned to each data sample by changing the diagonal on the similarity matrix, which allows for prior belief to influence which samples should serve as exemplars. In practice, this replaces the problem of determining the right number of clusters with the more complex problem of choosing the right and equally arbitrary preference that determines the right number of clusters (Brusco et al., 2018). We used an adaptive framework for affinity propagation to systematically vary preference values while ensuring convergence, which allowed us to generate a series of clustering solutions that contained between 2 and 1800 clusters (Wang et al., 2007). Cluster models were evaluated using the Bayesian Information Criterion (BIC; Schwarz, 1978) as adapted to clustering problems (Fraley & Raftery, 1998; SPSS, 2011). Although no cluster validation index has been designed explicitly for affinity propagation, the optimization model

underlying it is closely related to the p-median problem that k-medoids was proposed to solve (Brusco, et al., 2015). BIC was chosen for its quick computation time and heavy penalization of large cluster models. Further, we chose BIC because in analyses not reported here, when it is combined with the cluster model selection procedure we detail in the next section, it reasonably approximates the performance of the consensus clustering based model selection method we developed in Chapter 3 (we also tested several other methods of model selection like knee identification with the following validation measures: Calinski-Harabasz, Davies Bouldin Index, the silhouette score). Unfortunately, the fact that affinity propagation cannot identify a predetermined set of clusters made it difficult to implement this clustering approach within the framework of consensus clustering (e.g., Louvain clustering performed poorly in this framework in Chapter 3 for the same reason). Although methods exist for estimating the preference value that would determine a partition with a certain number of clusters, in practice these methods are too computationally expensive to perform over many subsamples of the data. Adaptive affinity propagation is also too computationally expensive and cannot guarantee the exact range of clusters that would be produced for each subsample.

As a further test of the impact of considering non-linear relationships between meta-analyses, we repeated the affinity propagation clustering process over Pearson correlation coefficients (see Figure 12). For cluster models that contained more than 8 clusters, affinity propagation with distance correlation noticeably outperformed affinity propagation with Pearson correlation according to BIC. This suggests that clustering over distance correlation produced groups of phrases that better fit the underlying meta-analytic data from which the clusters were produced.

4.2.5 Cluster model selection

A single cluster model was chosen from the range of clustering solutions that contained different numbers of functional categories. A BIC curve was computed across clustering solutions. A knee in the curve was selected as the relatively "optimal" model in order to avoid overtrained models reflected by monotonically increasing (or decreasing) cluster validation indices. A knee point detection algorithm designed with BIC and shown to outperform other heuristics was used to select a particular model (Zhao et al., 2008). Briefly, this algorithm highlights global trends in the BIC curve by normalizing BIC into the range of clusters being evaluated and then taking the ratio of normalized BIC and the number of clusters. This BIC ratio curve reveals the point at which changes in BIC begin to reflect the number of clusters within a model. In order to combine this information with local changes in BIC, the BIC ratio curve is then normalized and averaged with the purely normalized BIC curve. The optimal knee in BIC is then represented by the largest point on the averaged BIC curve that precedes the point at which the averaged BIC curve crosses below the normalized BIC curve. Notably, the optimal knee selected in this way can differ depending on the range of clusters being evaluated. As a result, we systematically varied our assumption about the minimum cluster size supported by the data and each time found the locally optimal knee (i.e., first testing models between 2 and 1800 clusters, then 3 and 1800 clusters, etc). The knee most frequently identified across the minimum cluster size threshold was selected as the optimal model. This procedure effectively emphasized the importance of large local changes in BIC while still considering global information about the curve to exclude overfit models. For instance, Figure 12 shows a surprisingly large

increase in BIC around the model with 270 clusters. For solutions with more clusters, BIC shows a monotonic increase (solutions above 280 not presented in this figure because linear affinity propagation was not extended to cluster models beyond 280 clusters). By adjusting the minimum cluster size threshold and then applying the described knee detection algorithm, we were able to identify this as a locally optimal solution while still considering the overall pattern of the BIC curve. This procedure was used to select the optimal number of clusters when grouping phrases in Neurosynth into functional categories.

Figure 12: Impact of non-linear distance on cluster models

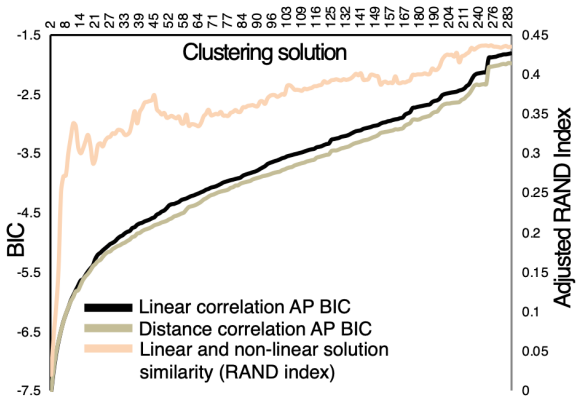

Figure 12: Impact of non-linear distance on cluster models. This figure presents BIC values for cluster models generated with affinity propagation using linear correlation as a measure of similarity (black line), and distance correlation as a measure of similarity (gold line). The relative similarity between these cluster models is also presented (pink line). BIC is shown on the left y-axis, and the RAND index adjusted for chance is shown on the right y-axis).

4.2.6 Modeling lower-dimensional functional networks

A lower-dimensional functional network was extracted from each functional category by performing dimensionality reduction over the meta-analyses of phrases assigned to each category (see Figure 13 for visualization). Non-linear principal component analysis (NLPCA) was chosen to reduce each set of meta-analyses into a one-dimensional map in order to preserve the non-linearities that helped drive the clustering of phrases into functional categories. NLPCA was performed using an autoencoder that learned an identity mapping by minimizing squared reconstruction error (Scholz, 2005). The network used for each functional category was comprised of 5 layers. The input and output layers had as many units as there were phrases/meta-analyses in the functional category. There were always 5 units in the hidden layers, and one unit in the component layer. A separate analysis was carried out for each functional category in order to find the optimal network complexity as determined by the weight-decay penalty term. This was accomplished by artificially generating missing data, applying inverse NLPCA to the training set using a variety of weight-decay complexities, then validating each of the models using the imputed values (Scholz, 2012). The process of tuning network complexity in this way ensured that NLPCA did not force a non-linear solution when a linear one would better predict the data. Thus, the lower-dimensional networks produced using NLPCA for some functional categories were equivalent to the first principal component from PCA. Inverse NLPCA models only the decoder portion of the autoencoder, estimating both the weights and the inputs simultaneously using conjugate gradient descent (Scholz et al., 2005). Critically, this process of predicting missing values in the data avoids issues associated with test set

validation, which tends to favor over-fitted models for unsupervised learning problems, including NLPCA (Scholz, 2012). The network complexity that produced the lowest median missing value estimation error across 100 runs (i.e., different assignments of missing values) was selected as the optimal complexity value. In each run, the meta-analyses for exactly half of the phrases associated with a functional category were replaced by missing values.

4.2.7 Functional parcellation

A functional parcellation of temporal and inferior parietal cortex was performed based on the lower-dimensional functional networks extracted in the previous step (see Figure 13 for visualization). The lower-dimensional networks embedded the relative importance of brain regions to groups of phrases within Neurosynth. Functional areas were then defined based on having similar patterns of association across lower-dimensional networks. That is, voxels were grouped on the basis that they loaded onto the same networks. This functional parcellation was achieved using the previously described clustering procedure and cluster model evaluation process. The eta^2 coefficient (Cohen et al., 2008) was used to generate a similarity matrix over which clustering was performed. This coefficient is largely the same as Pearson correlation but modified to account for scaling and offset when comparing network maps (Cohen et al., 2008).

4.2.8 Structural connectivity data

To validate the functional areas that were extracted from Neurosynth, a separate parcellation of the same temporal and inferior parietal cortex ROI was performed using

structural connectivity data. Structural T1-weighted and diffusion-weighted scans were downloaded for 70 subjects (38 female) from the Human Connectome Project (humanconnectome.org). Subjects were ensured to be unrelated (i.e., from the "100 unrelated subjects" database), and between the ages of 22 and 35 years. MRI data were acquired at Washington University St Louis using a Siemens Magnetom Connectome 3T scanner. A full description of the preprocessing protocols and acquisition parameters can be found within previous studies (e.g., Glasser et al., 2016). Whole-brain probabilistic tractography was performed with probtrackx2 (Behrens et al., 2007) by seeding from each voxel within the temporal and inferior parietal cortex, generating over 23,000 tractograms. A total of 5000 streamlines was initiated from each voxel. All streamline counts were normalized by the average length of the streamlines contributing to each count in order to avoid bias towards shorter tracts due to compounded uncertainty (Roberts et al., 2016). This procedure was repeated separately for each subject. The tractograms generated across participants were then averaged (see Figure 13 for visualization).

4.2.9 Structural parcellation

Although the structural parcellation was performed over tractograms generated for each voxel in our ROI, the overall procedure for forming the structural parcellation followed the same steps we described for the functional parcellation (see Figure 13). That is, we extracted lower-dimensional structural networks from the voxel-level tractograms, computed the similarity between voxels in our ROI based on their associations across the lower-dimensional structural networks, then clustered voxels together such that they formed groups that associated with similar structural networks.

Although we aimed to keep the methods for extracting lower-dimensional structural and functional networks as similar as possible, the larger dimensionality of the structural data made it unfeasible to use either NLPCA, or to simply perform clustering over distance correlations between tractograms. As a result, we instead modelled structural connectivity within a lower dimensional manifold that preserved local information using Laplacian Eigenmaps (Belkin & Niyogi, 2003). We then applied Independent Component Analysis (ICA) to this lower dimensional space to generate a set of independent sources. This procedure assumed that tractograms for voxels in the temporal and inferior parietal cortex were characterized by a set of independent structural connectivity patterns, or tracts, that may not be adequately represented if linear dimensionality reduction techniques are applied to the higher dimensional data. A fundamentally similar approach has been used for extracting resting state networks at the group-level (Liu et al., 2018). While ICA is less frequently applied in the context of structural data, it has been recently shown to produce anatomically consistent tracts with high reliability (O'Muircheartaigh & Jbabdi, 2018; Wu et al., 2015). Overall, the method we used to extract lower dimensional structural networks is highly similar to this prior work (O'Muircheartaigh & Jbabdi, 2018), with the exception that a non-linear dimensionality reduction method is applied as a pre-processing step for ICA.

In order to extract the lower-dimensional structural networks from the group-level tractograms, we first formed a symmetric similarity matrix based on the Pearson correlation coefficient computed between the tractograms of each pair of voxels. A k-nearest neighbors search was then performed on the similarity matrix to form an adjacency matrix. In line with prior work, we determined k by z-scoring each voxel's correlations with all of the other voxels, then averaging the number of correlation values

that survive a z-score threshold of 2 as determined separately for each voxel (Liu et al., 2018). We then computed the normalized Laplacian and found the smallest eigenvalues and eigenvectors to form the lower-dimensional embedding. The fastICA (Hyvarinen, 1999) algorithm was then applied to the Laplacian eigenmaps and the resulting independent Laplacian components were clustered to produce a structural parcellation of the ROI. Clustering was performed using affinity propagation. That is, over a similarity matrix constructed between each pair of voxels in our ROI based on the eta^2 coefficient between the Independent Laplacian Component scores for the pair of voxels.

The number of components in the ICA was always equal to the number of dimensions in the lower-dimensional embedding. A version of affinity propagation that iteratively tests different preference values was used to ensure a clustering that had the same number of clusters as the number of dimensions in the lower-dimensional embedding (e.g., Frey & Dueck, 2007). To determine the optimal number of these dimensions, this entire procedure of generating a lower dimensional embedding of voxels in the ROI, and then clustering them, was repeated using different dimensionalities. The end cluster models were then evaluated with BIC (i.e., in the same way as the functional parcellation) and the best performing model was retained. Because this procedure resulted in components that scored onto each voxel within our ROI, the tractograms that represented each independent Laplacian component were extracted by regressing the independent components on the group averaged set of tractograms (O'Muircheartaigh & Jbabdi, 2018). The resulting maps were corrected for significance using a gamma mixture model fit to the image histogram as commonly performed in resting state fMRI studies with ICA (Beckmann, 2012).

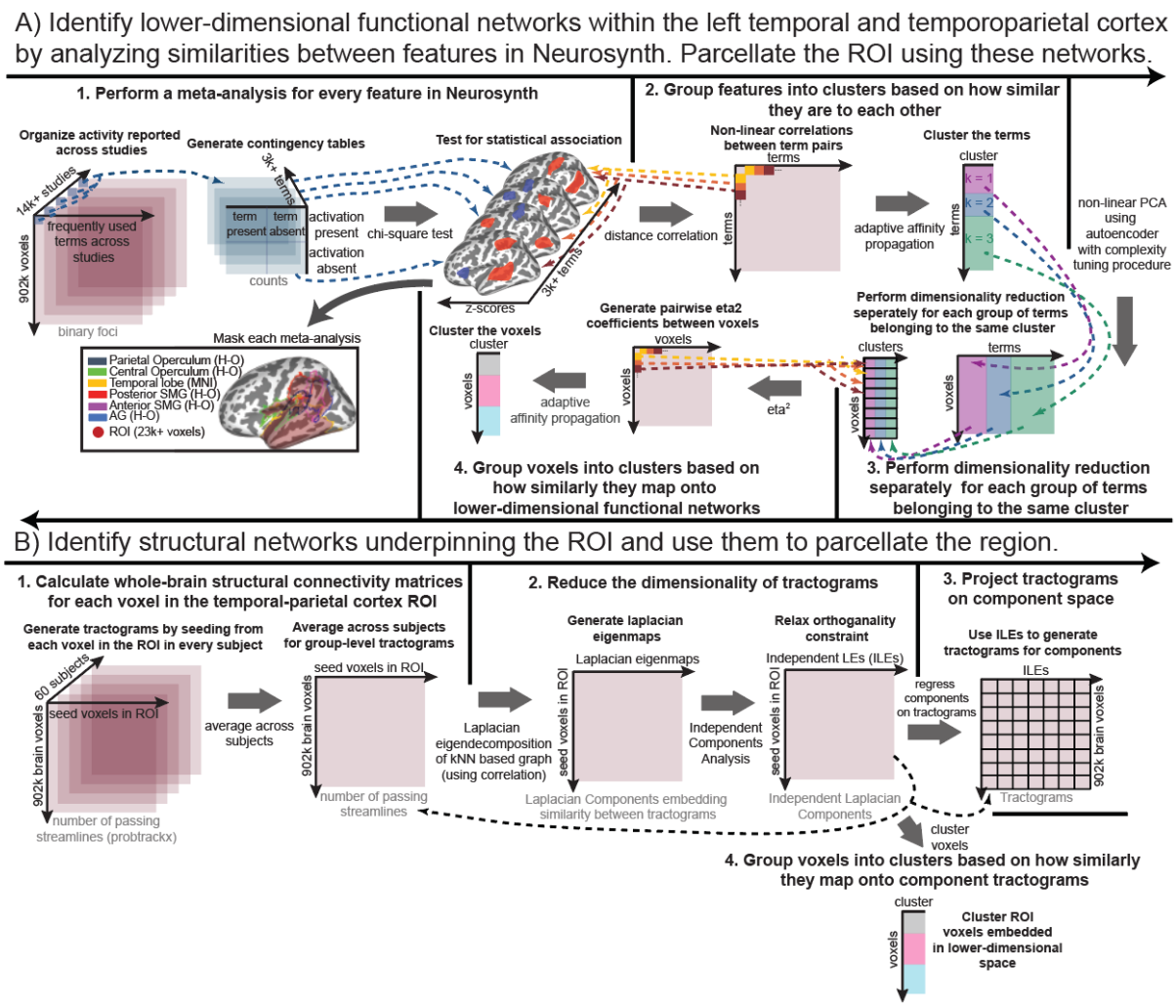

Figure 13: Overview of structural and functional parcellation analysis. Panel A presents the analysis pipeline we have employed to extract unique lower-dimensional functional networks from Neurosynth. In Neurosynth, studies are queried based on the frequency with which they use phrases (i.e., features) common to the neuroimaging literature. Each phrase can be meta-analyzed by extracting the studies that frequently use that phrase and comparing their patterns of activity to the remaining studies in the database that don't frequently use that phrase.

4.3 Results

4.3.1 Separation of functional networks involved in speech, sentence comprehension and word recognition

In the service of localizing areas involved in speech perception, we grouped together meta-analyses representing patterns of activity associated with different phrases commonly used in the functional neuroimaging literature. These phrases acted as an index of different cognitive processes, and their groupings reflected broader functional categories. For each functional category, we extracted a single lower-dimensional functional network that represents the common pattern of activity across the meta-analyses contained in the group. We then combed through the functional categories to determine if speech perception was associated with a unique functional category, which would imply that it is associated with a distinct pattern of brain activity.

Before choosing a particular grouping of phrases into functional categories to focus our analysis on for localizing the speech perception network, we examined the behavior of the clustering algorithm as a function of increasing the number of functional categories. This allowed us to determine how readily speech networks, and potentially a speech perception network, could be distinguished from other functional networks. Because our clustering approach involved selecting an exemplar for each functional category, we were able to use these exemplars to register arbitrarily assigned cluster numbers across the wide range of cluster models that were tested (i.e., containing anywhere from 2 to 500 clusters). The alluvial diagram in Figure 14 presents a small but critical slice of the

clustering algorithm's behavior, tracing which functional categories contributed items towards the formation of each newly formed functional category. Critically, speech perception and auditory processes were distinguished when very few functional categories were formed, suggesting that they express different patterns of brain activity.

First, we note that a speech perception category emerges when at least 192 functional categories are carved out of the data. To rank items within a functional category, we inspected their correlation distance to the category exemplar. For the speech perception category, this revealed 'listening', 'acoustic', 'speech perception', 'speech production', 'speech sounds', and 'phonetic' to be amongst the most highly ranked items. Thus, we knew that the functional category exemplar for a unique speech perception network was the pattern of activity for the phrase 'speech'. We tracked which cluster the activity pattern for that phrase was placed in as a function of increasing the number of functional categories carved out of the data. When three functional categories were carved out, the auditory network organized its own cluster of phrases. At this early point of very few clusters, the speech perception network was better represented by the 'auditory' meta-analysis (i.e. pattern of activity) than the other functional categories (i.e., motor, amygdala, and inferior frontal). The auditory meta-analysis served as the exemplar for the speech perception meta-analysis until 14 functional categories were defined. At this point, the speech meta-analysis was better represented by the meta-analysis for "superior temporal", which was the case until 192 functional categories were created. At that point, speech emerged as a unique exemplar for a functional category that clearly represented speech perception.

Note that exemplars were selected on the basis of statistical dependency between meta-analyses. Anatomical terms often served as exemplars because they represented the area of contact between many differently configured functional networks, which was captured by the distance correlations between phrases/meta-analyses. For instance, when much fewer than 192 functional categories were investigated, "superior temporal" served as a suitable exemplar for a range of processes/terms unrelated to speech, including several disorders (schizophrenia, Parkinson's), decision making, uncertainty, expectation, task difficulty, and other anatomical regions (e.g., variations of STS, STG, and MTG but also dorsolateral prefrontal cortex). We demonstrated that this behavior was directly attributable to the distance correlation by repeating the clustering process using the Pearson correlation coefficient. Analysis of linear relationships between networks caused speech to emerge when only 25 functional categories were generated. Thus, the fact that speech and audition were virtually always separated into different functional categories shows that they produce unique patterns of activity.

The alluvial diagram also shows that functional categories organized around processing words and language emerged as unique categories when relatively few clusters were defined. A language functional category was produced when only 6 clusters were created, and a word functional category emerged at 24 clusters. From 6 to 23 clusters, the language category included the "words" network. When words emerged as a unique cluster, the language category represented comprehension at a sentence-level that made contact with speech representations (i.e., sentence comprehension, reading, semantic, syntactic and phonological processing), and the words category represented word recognition and mapped onto lexical-semantics, word form processing, as well as tasks that

required executive and attentional processes (e.g., lexical decisions, judgment tasks, naming, priming). Overall, these patterns suggest that it may be possible for us to extract other networks involved in speech perception at a higher-level (e.g., phonological processing, word recognition, comprehension, etc). Finally, we note that clustering phrases into functional categories based on their networks was capable of isolating noise within the Neurosynth database (i.e., frequently used phrases that did not have functional significance). For instance, a cluster of commonly used phrases for BOLD response was observed, as well as clusters representing methods (e.g., ICA, resting state, ASL), and journals (e.g., hum mapp).

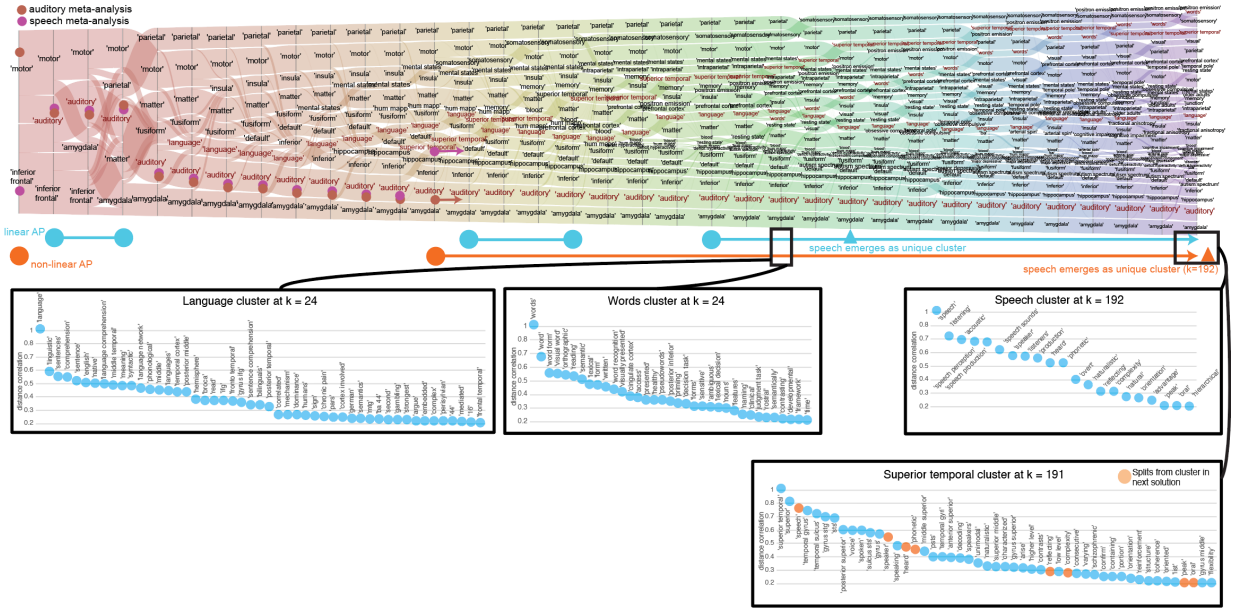

Figure 14 caption: Cluster splitting and merging behavior. An alluvial diagram showing cluster splitting and merging behavior is presented. Each vertical line represents the addition of a new cluster. The number of items that separate from a cluster in the previous solution and merge with a cluster in the new solution is depicted by a connected curve proportional to cluster size. The color scheme from left to right maps on to the number of clusters in the solution but is arbitrary in this case. Each cluster is labeled by its exemplar. Brown dots on the alluvial diagram represent the cluster in which the meta-analysis for "auditory" resided. The purple dots on the alluvial diagram represent the cluster in which the meta-analysis for "speech" resided. Clusters relevant to speech are labeled in red.

Below the alluvial diagram are orange and blue horizontal segments. Each segment represents the span of clustering solutions in which "speech" and "auditory" meta-analyses were placed in different clusters. The blue line represents clustering over linear correlation and the orange line represents clustering over distance correlation. Triangles represent the point at which speech emerged as a unique exemplar. Note, the alluvial diagram presents results from the distance correlation clustering. The contents of select language-related clusters are extracted in boxes below. Items within a cluster are ordered according to the strength of their statistical dependency on the cluster. The superior temporal cluster is included to show the contents of the cluster from which the "speech" meta-analysis emerged as a unique exemplar/cluster.

4.3.2 Model selection for functional categories, functional parcellations, and structural parcellations

In the previous section, we established that brain response associated with speech perception is readily distinguished from auditory processing, even when a different number of functional categories are carved out of the Neurosynth database. Our next aim was to determine if a relatively optimal clustering of Neurosynth meta-analyses would distinguish speech perception as its own unique functional category, indicating that it is associated with a unique pattern of brain activity. The process of choosing the number of functional categories to focus on involved selecting an optimal cluster model from a set of other models with varying numbers of functional categories. Once an optimal model was selected, the meta-analyses within each functional category were collapsed into a lower-dimensional functional network using a nonlinear dimensionality reduction technique. Extracting lower-dimensional functional networks allowed us to localize the pattern of brain activity associated with speech perception. While we inspected the functional networks related to speech more closely, we also wanted to more systematically appreciate how the areas within speech networks related not just to each other, but to the lower-

dimensional networks associated with other functional categories. To accomplish this, we performed a functional parcellation, clustering the temporal and inferior parietal cortex region that we meta-analyzed using the lower-dimensional functional networks. This produced functional areas that had similar associations to lower-dimensional functional networks, which in turn represented functional categories. Clustering for the functional parcellation also involved selecting an optimal cluster model. Finally, we parcellated the same temporal and inferior parietal cortex using its structural connectivity as a validation of the functional boundaries we could delineate with the meta-analyses. This too involved selecting an optimal cluster model. Here, we go through the process of selecting an optimal cluster model for all three analyses—determining the number of functional categories, the number of parcels in the functional parcellation, and the number of parcels in the structural parcellation.

The full extent of the procedures we employed to select the right cluster models both for functional category definition and functional parcellation are discussed in section 4.2.5. Briefly, we used a knee-point accentuation and detection algorithm designed for BIC. Because this algorithm is sensitive to the range of clustering solutions that are evaluated, we systematically changed the minimum cluster size threshold to be between 2 and 400 and recomputed the knee. This strategy allowed us to identify knees in the BIC curve that were not easily predicted by changing the range of clusters being evaluated, and therefore constituted a meaningful local change in BIC. For clustering functional categories, this approach revealed 243 clusters to be the optimal model, and for clustering functional networks the optimal model was determined to be 35 clusters (Figure 15A). We note that, in the end, this more complicated approach to identifying a knee in the BIC curve ended up

selecting the same models as a simpler approach of programmatically identifying a knee in the curve by connecting the first and last points on the curve by a line and finding the point at which the distance between the curve and the line is maximal.

The number of structural networks that we defined was based on the ability of the set of modeled structural networks to create a good structural parcellation evaluated with BIC. That is, we varied the number of structural networks that were modeled, then clustered the set of networks and measured BIC (Figure 15B). This was repeated for 10 dimensionalities/cluster models containing between 11 and 101 dimensions. The resulting BIC curve did not show any substantial local changes that would warrant applying the more careful approach taken with the functional data to identify the knee. Thus, we used the simpler procedure that was described to converge with the more complicated one in the selection of functional models. This revealed an optimal 31 structural networks and parcels.

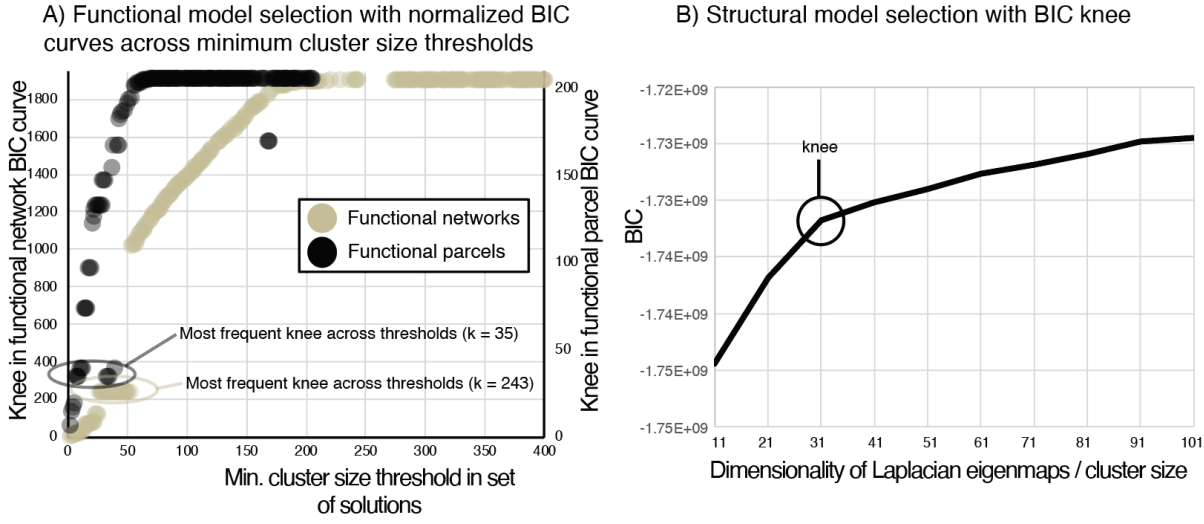

Figure 15: Model-order selection for functional networks and parcels. Panel A presents model selection performed to determine the number of functional categories in

Neurosynth, and the number of functional parcels in the parcellation of temporal and inferior parietal cortex using the lower-dimensional functional networks associated with functional categories. The knee in BIC is identified using a knee-point accentuation and detection algorithm designed for BIC. This algorithm is repeated across different assumptions about the minimum cluster size supported by the data (i.e., minimal cluster size threshold) because the algorithm is sensitive to the range of solutions evaluated. The optimal model is selected as the one that was most frequently identified as the knee across different minimum cluster size thresholds. The maximum number of clusters evaluated was often identified as the knee—these were not considered to be potential optimal models. Panel B presents model selection performed for the structural networks and structural parcels. The number of lower-dimensional structural networks was systematically changed, and clustering was repeated using a number of predetermined clusters based on the dimensionality of the set of lower-dimensional structural networks. A clear knee in the BIC curve was observed so the knee accentuation and detection method was not utilized.

4.3.3 Identifying and analyzing functional networks related to speech

Functional categories related to speech were identified based on the items that belonged to each functional category (i.e., meta-analyses for different phrases). The importance of certain phrases to a functional category was analyzed by regressing the functional network for that category on the set of meta-analyses from which the network was produced. These associations are presented for functional categories related to speech processing in Figure 16. Note that in the previous analysis inspecting the emergence of new functional categories across different cluster models, we found that sentence comprehension and word recognition emerged as unique functional categories when relatively few functional categories were carved out of the data.

The specific cluster model we selected generated a number of additional insights into dissociable patterns of brain activity related to speech processing. First, both early

acoustic and phonetic processing was associated with a unique set of phrases representing different patterns of activity. The auditory cluster was strongly associated with sounds and auditory cortex, but also included meta-analyses for noise, tones, music, spectral, audiovisual, and vocal processing. The exemplar for this cluster was the meta-analysis for auditory, confirming its role in early acoustic analysis. Note, that within the figures, clusters are referred to by their exemplar as a convenience, while in the text they are referred to by their functional interpretation. Thus, the "early auditory" or "early acoustic" cluster within text, is represented as "auditory" in the figures. The cluster representing phonetic processing was organized around the "speech" meta-analysis. This cluster referred to early speech analysis, and was associated with speech sounds, speech perception, and phonetics. Interestingly, this cluster appeared to be associated with studies using naturalistic stimuli and was also closely associated with speech production. This likely reflects the fact that speech production studies are often overt, and as a result, auditory feedback activates the region involved in representing speech sounds. Indeed, the meta-analysis for "overt" was grouped into this phonetic cluster as well. The association with speech production can also be explained in the case that this functional category symbolizes representations for speech sounds, which would need to be accessed in order to map constructed phonological representations onto sound structures that can be used to guide speech production.

The early auditory and phonetic clusters were further distinguished from meta-analyses that represented a phonological-orthographic network (Figure 16). The exemplar network for this functional category was orthography. However, it mapped more strongly onto phonology than letters or characters and was also associated with naming. Additionally, associations with phrases like "visually presented" suggests that this

functional category represents phonological processing driven by visual stimuli. Critically, this phonological-orthographic functional category was separated from a category representing word recognition, for which the “words” meta-analysis acted as a catalyst. This category mapped strongly onto lexical-semantic access, word recognition, reading, and word form processing. It was also associated with lexical decision tasks, judgment tasks, and pseudowords. Processing multiple words was associated with its own cluster, formed by the exemplar meta-analysis of “word pairs”. Speech comprehension was also dissociated, forming a unique cluster represented by “comprehension”. This cluster mapped onto sentences, language comprehension, syntax, and context. Further, it was associated with discrimination tasks. A different cluster was formed around the “language” meta-analysis. However, what this cluster represents is difficult to interpret. It appears to contain noisy meta-analyses representing vague phrases commonly used in speech studies (e.g., “English”, “linguistic”, “language network”, “broca”), but was also associated with semantics and meaning. Another functional category more specifically represented concepts, which were organized around the meta-analysis for “abstract”. Verbal fluency, a spontaneous language production task that involves finding semantically or phonemically similar items was separated into its own cluster as well. Finally, word class was distinguished into a unique cluster exemplified by the meta-analysis for “verbs”; however, the cluster also loaded strongly on verbs, nouns, and “class”.

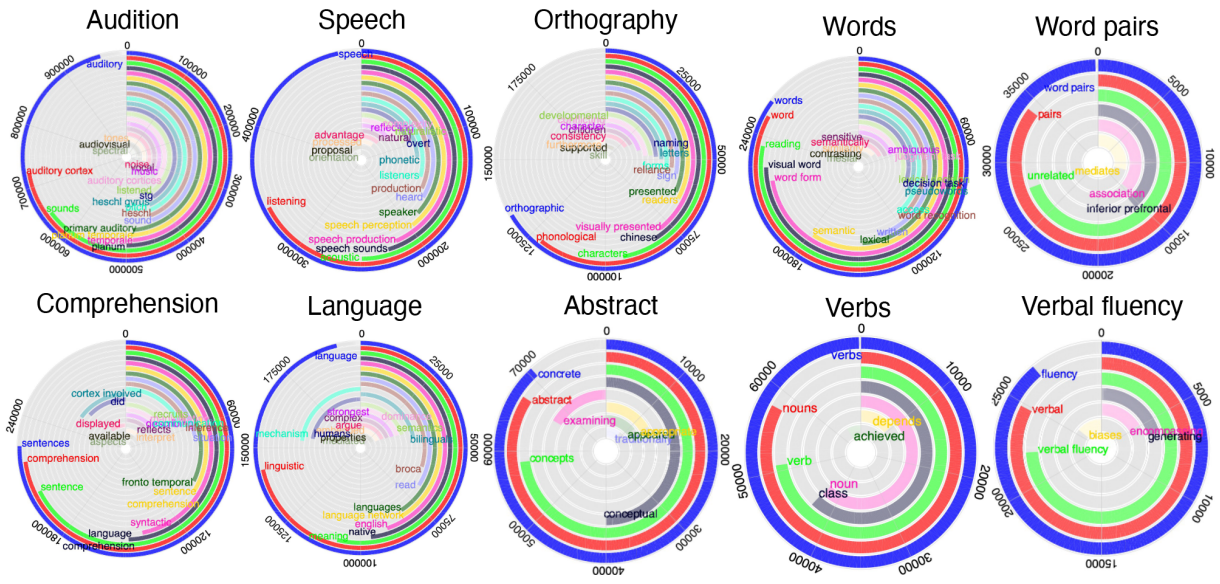

Figure 16: Functional categories related to speech. Item-to-category loadings are presented for each functional category that appeared related to speech. The curved bars going clockwise show positive loadings, while curved bars going counterclockwise show negative loadings.

Having identified functional categories related to speech, we next directly compared these functional networks to each other. Because each functional category represents a unique pattern of brain activity, we expected these speech categories to demonstrate unique networks that shared regions. In particular, we focused on a set of speech related functional categories that appeared to form, conceptually, a speech processing hierarchy—early auditory processing, phonetic processing, phonological-orthographic processing, word recognition, and speech comprehension clusters formed a clear speech processing hierarchy.

First, we inspected acoustic, phonetic, and phonological-orthographic processing. These three lower dimensional networks are projected onto the same surface in Figure 17 (panel A) using a 3-dimensional colormap to highlight areas that are relevant for all of these functional categories, and areas that are uniquely relevant to each category. This

analysis revealed a processing gradient transitioning from: i) primarily acoustic but also phonetic processing in Heschl's gyrus and sulcus, to ii) primarily phonetic processing but also phonological-orthographic processing in the ventral STG, stretching from mid-to-posterior ventral STG, to iii) stronger association with phonological-orthographic processing in the ventral bank of the pSTS, and iv) uniquely phonological-orthographic processing in the fusiform gyrus. The overlap in association patterns for phonetic and phonological-orthographic processing in superior temporal cortex suggests that this activity in the phonological-orthographic network may represent phonological processing, whereas the activity in inferior temporal cortex may represent orthographic processing. This is supported by the fact that the phonological-orthographic network is much more strongly associated with the inferior temporal cortex and at the same time its functional category both loads slightly more strongly onto orthographic processing and contains many references to orthography (e.g., characters, letters, and forms). At the same time, the phonological-orthographic network shows a stronger association with an area of pSTS relative to the phonetic network, suggesting that this area may be involved in the more abstract processing of speech sounds.

To test whether this pSTS region was relevant for higher-level speech networks, we projected three more functional networks onto a single surface: phonological-orthographic processing, word recognition, and speech comprehension (Figure 17, panel B). The dorsal bank of the posterior superior temporal sulcus showed a strong association for all three functional networks. This region was strongly associated with comprehension, and moderately associated with both word recognition and phonological-orthographic processing. Curiously, all three functional networks also showed some extent into more

posterior portions of the ventral STG region associated with phonetic processing. A strong unique relationship was observed between comprehension and anterior MTG. Further, both word recognition and speech comprehension associated with the posterior MTG. That this region was not involved in phonological-orthographic or phonetic processing suggests that it may be involved in lexical-semantics. The posterior inferior temporal gyrus mainly showed strong associations with phonological-orthographic processing and word recognition but showed a complex pattern of loadings that formed concentric circles around a central area associated purely with phonological-orthographic processing. The area around the center was associated with both phonological-orthographic processing and word recognition, and the area in the periphery was associated purely with word recognition.

In summary, the highest level of speech processing (speech comprehension), recruited the mid-to-posterior ventral STG involved in phonetics, the abutting dorsal pSTS involved in phonological processing, the neighboring pMTG involved in word recognition, as well as regions of anterior temporal lobe. Word recognition involved the same regions with the exception of the anterior temporal areas. Phonological-orthographic processing involved the same regions as word recognition, but not the pMTG, presumably because lexical-semantic information is not accessed yet at this stage. Phonetic processing did not strongly involve the pSTS or pMTG, reflecting a much lower-level speech sound analysis. Indeed, it was the only other area to load moderately onto auditory cortex. Finally, acoustic processing only modestly involved the mid-to-posterior ventral STG implicated in speech sound analysis.

To provide more transparency into the 3-dimensional surface projections (Figure 17A), we also visualized each of the speech networks individually (Figure 17B), confirming our observations about a speech processing hierarchy. Further, we ensured that this hierarchy is preserved in the original meta-analyses. That is, we verified that the meta-analyses that served as the exemplars for each of the speech functional categories showed the same patterns of activity. In addition, we confirmed that the patterns that we saw in the individual meta-analyses were robust against publication bias using an adapted Fail-Safe N statistic. Briefly, Fail-Safe N measures the number of null studies that would be required to overturn a result. Recall that generating a meta-analysis in Neurosynth involves performing a statistical test at each voxel over a contingency table that compares the presence versus absence of activity, and the presence versus absence of a phrase. We iteratively added to the number of studies that use a phrase but do not report activity in the voxel until the resulting statistic was insignificant for that voxel. This procedure was repeated across voxels to generate a Fail-Safe N map. This map was used to threshold out voxels in the meta-analysis that would be insignificant if at least 100 unpublished studies showed a null effect. Recent estimates suggest that there is somewhere between 1-12 unpublished fMRI studies for every 100 published fMRI studies in the language domain (Smartsidis et al., 2020). We note that sample sizes for the *exemplars* of each functional category were: auditory, 1252 studies; speech, 642 studies; orthographic: 372 studies; comprehension, 424 studies; word recognition, 948 studies. We report that this evaluation of publication bias did not impact the speech processing hierarchy that we observed across these meta-analyses. Finally, we note that the meta-analyses for category exemplars were overall incredibly similar to the lower-dimensional functional networks, however, the

meta-analyses showed slightly lower specificity. For example, the meta-analysis for auditory assigns a stronger role to ventral superior temporal gyrus than suggested by the latent auditory functional network.

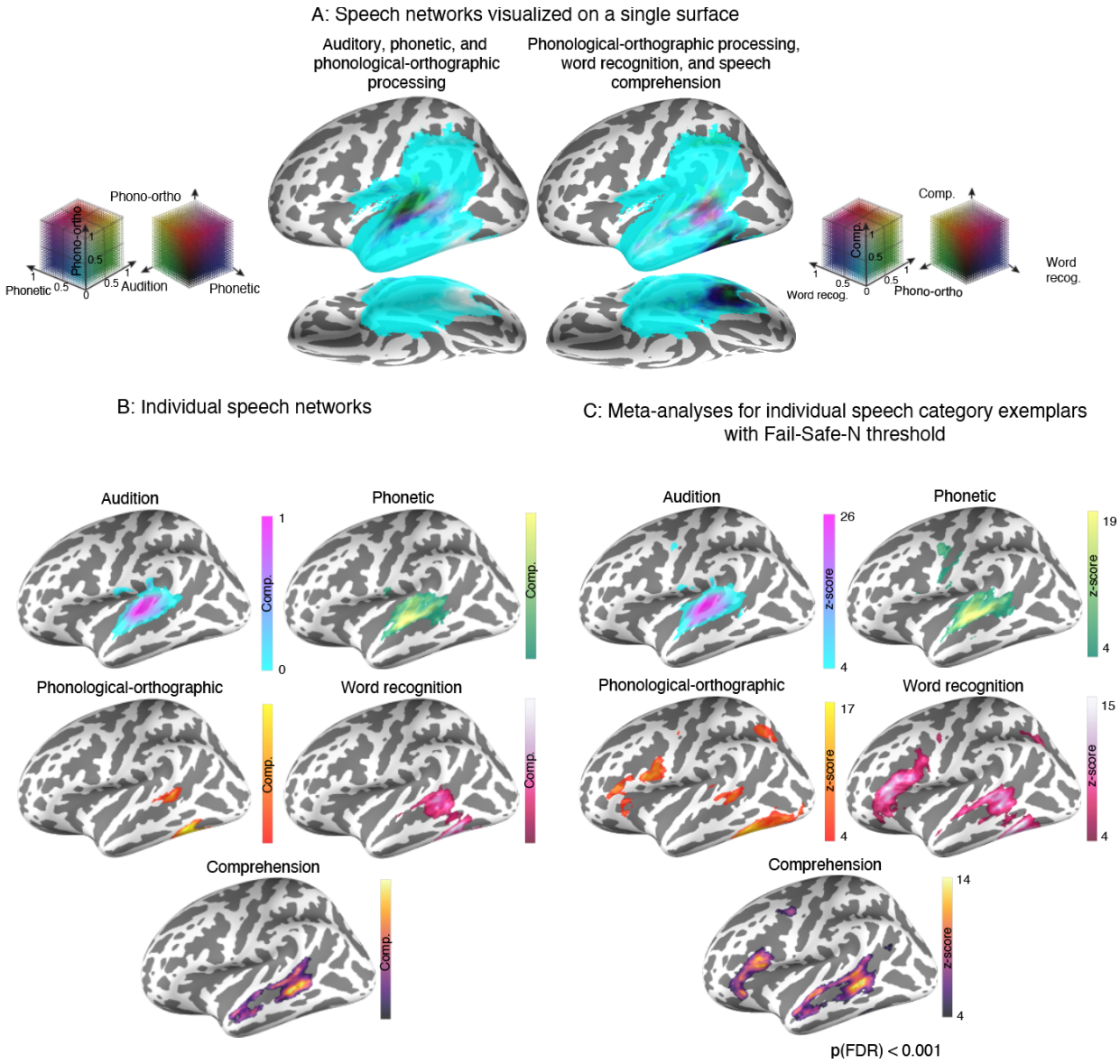

Figure 17: Functional networks related to speech. The lower-dimensional functional networks for functional categories related to speech processing are visualized in panel A. For both surface visualizations in this panel, three functional networks are projected on a single surface by mapping their values onto a 3D colorcube presented on the far left. Distinct colors represent unique loading patterns across the three networks being compared (unthresholded). In the surface on the left, dark green (Heschl's gyrus)

corresponds to a high association with auditory processing but moderate association to phonetic processing, purple (mid-to-posterior ventral STG) represents a high association with phonetic processing but a moderate association with auditory processing, a very light pink (pSTS) represents high association with phonological-orthographic processing but a moderate association with phonetic processing. White (fusiform gyrus; FG) represents a high association with phonological-orthographic processing alone. In the surface on the right, red (pSTS) represents a high association with phonological-orthographic processing, word recognition, and comprehension. Pink (pMTG) represents a high association with word recognition and comprehension. Black (FG) represents high association with word recognition and phonological-orthographic processing. A muddier white color (ATL) represents medium to high association with comprehension. Individual networks are presented separately in panel B to highlight all of these trends (the loadings for these "components" range between 0 and 1 just as in panel A; a 95-percentile threshold was applied to highlight some of the trends, however this slightly underestimates the extent of the ventral STG area in the phonological-orthographic network). In addition, the meta-analysis for the exemplar of each functional category is presented in panel C. Each meta-analysis is corrected at p(FDR) < 0.01. However, we also estimate the number of null studies necessary to eliminate an effect at each voxel. Only voxels that require more than 100 null studies to overturn the result are shown.

4.3.4 Functional and structural parcellations

In the previous section, we have mapped a speech processing hierarchy across a handful of latent speech networks that involves the mid-to-posterior ventral STG in phonetic processing, the pSTS in phonological processing, the pMTG in lexical-semantics, and anterior temporal areas in comprehension. Nevertheless, the extent to which these areas are involved in non-speech functions remains unclear. Our strategy for uncovering the other functions in which these areas participate was purely data driven. We defined functional areas by grouping voxels together such that they had a similar loading pattern onto the same functional networks. In doing this, we hoped to be able to define the same speech areas we identified by directly comparing latent speech networks to each other. Then, we could further investigate the functions associated with each region. A structural connectivity-based parcellation of the same temporal and inferior parietal region was then

carried out to corroborate the functional dissociations. In other words, we assume that areas that align more closely across modalities are more likely to reflect distinct functional modules.

Functional and structural parcellations are presented in Figure 18. Clusters were matched across parcellations with the Kuhn-Munkres algorithm (Kuhn, 1955). Each parcel within the structural parcellation was color-coded based on the functional parcel it most closely resembled based on maximizing spatial overlap. Qualitatively, these parcellations showed a striking amount of similarity, but also showed substantial differences. Critically, the functional parcellation contained 35 clusters while the structural parcellation contained 31. To ensure that the correspondence between functional and structural parcellations was statistically significant, we used a spatial permutation testing procedure that performed a randomization test of the correspondence between parcellations, by sampling uniformly from the space of possible rotations of a spherical representation of the cortical surface (c.f., Alexander-Bloch et al., 2018; Lefèvre et al., 2018; Fryzlewicz, 2016). This sampling procedure was used to build up a null distribution of spatial correspondence based on the unique region that was clustered. The degree of similarity between the rotated and original parcellations was computed using normalized variance of information (NVI). NVI is highly similar to normalized mutual information but has the property of being a true metric. This testing revealed a significant correspondence between the functional and structural parcellations ($p < 0.05$; Figure 18C). Notably, we also tested the structural parcellation against the Harvard-Oxford atlas, however, the correspondence was insignificant ($p > 0.05$). This suggests that the structural parcellation captured the

functional areas extracted from Neurosynth much better than regions defined purely on macroscopic anatomical landmarks.

We next identified the clusters within the functional parcellation that corresponded to the regions that we found to be shared across the lower-dimensional speech networks investigated earlier. As we identified these functional clusters related to speech, we also compared them to the structural parcels with which they have been matched to evaluate whether structural data appears to produce the same boundaries. Some of the functional clusters that were carved out exhibited qualitatively good alignment with the speech regions we investigated previously (Figure 18A and C). In particular, the region implicated in phonetic processing was isolated in both parcellations and appeared to show close correspondence (functional and structural cluster 26). In addition, the comprehension network was associated uniquely with involvement of anterior areas, all of which were isolated in the functional parcellation, and some of which were isolated in the structural parcellation. The more anterior comprehension area spanning aSTS, aSTG, and aMTG was isolated in both the structural and functional parcellation (functional cluster 23; structural cluster 9). The more posterior area near anteroventral STG was isolated in the functional data, but not the structural data (functional cluster 9). Curiously, the functional parcellation indicated that the pMTG region that we found to be associated with word recognition and comprehension networks could be split up into an anterior (functional cluster 30) and a posterior (functional cluster 7) cluster. However, the posterior cluster also included the region of the pSTS that was primarily associated with the phonological-orthographic processing, but which also appeared in the comprehension and word recognition networks. Thus, it appears that some of the differences in activity that we observed across speech

networks could not be resolved by the clustering algorithm, and it's unclear precisely how these functional clusters relate to the speech networks. In the structural connectivity parcellation, the pMTG region from the speech networks was not separated into two clusters (structural cluster 7), but the connectivity data could not effectively isolate the portion of the pSTS which we observed to be critical for phonological processing. This region was split across a number of different clusters (structural clusters 12, 17, 15). Notably, one of these clusters formed a subnetwork with a slightly more anteroventral portion of pMTG (structural cluster 17). However, the portion of this cluster in pSTS was comprised of few voxels and difficult to analyze. Finally, the area we found critical for the early auditory processing functional network was isolated in the functional parcels (functional cluster 27). In the structural parcellation, auditory cortex does not extend as far posteriorly into Heschl's suclus and anterior PT and appears to include large portions of the insula (structural cluster 27).

To supplement our qualitative assessment of how well clusters related to speech regions correspond across modalities, we systematically investigated how well each cluster from both parcellations could be fit to both kinds of data. To measure "fit", we computed the ratio of inter-cluster to intra-cluster similarity using the pairwise similarities between voxels (i.e., based on loadings on functional or structural networks) that were submitted to the clustering algorithm (Figure 18B). Intra-cluster similarity was taken to be the mean eta^2 coefficient between all voxels belonging to that cluster. The inter-cluster similarity for a particular cluster was then defined as the mean eta^2 coefficient between all pairs of voxels belonging to that cluster and all the voxels outside it. A score on this ratio above 1 represented a cluster that was comprised of items more similar to each other than to the

items outside it. This similarity ratio, which we call homogeneity for simplicity, was generated for each structural and functional parcel, using both structural and functional data. For each type of parcellation, we then plot structural homogeneity against functional homogeneity to reveal how well clusters fit to structural and functional data. One striking findings from this analysis was that one of the functional clusters that was best fit to the functional data corresponded to the ventral mid-to-posterior STG that we observed to be critical for phonetic processing (functional cluster 26; Figure 18). Structural data fit this parcel moderately well compared to other functional clusters. This cluster also happened to show the most qualitatively similar boundaries to its matched structural cluster (functional and structural cluster 26). Further, we found that the structural cluster associated with ventral mid-to-posterior STG fit the functional data better than any other structural cluster, suggesting that when areas in temporal and inferior parietal cortex are defined purely on their connectivity to the rest of the brain, ventral mid-to-posterior STG shows the best evidence of a functional distinction. Notably, the structural cluster associated with ventral mid-to-posterior STG was slightly more anterior relative to its functional counterpart, lying more squarely within mid-STG. However, this cluster was still mainly located posterior to HG, which is an important point in relation to previous studies that have suggested speech perception occurs anteriorly to HG. Given that structural connectivity data did not fit the functional cluster on the ventral mid-to-posterior STG as well, it appears that structural connectivity data is more sensitive to the precise boundaries of this cluster (i.e., functional data fit ventral mid-to-posterior STG cluster as well when it was defined on structure or function). Overall, this implies that the functional parcellation showed good evidence for the delineation of a region we observed to be related to phonetic

processing in the lower-dimensional speech networks, and that a region with similar boundaries can be delineated purely based on its structural connectivity relative to other areas of temporal and inferior parietal cortex.

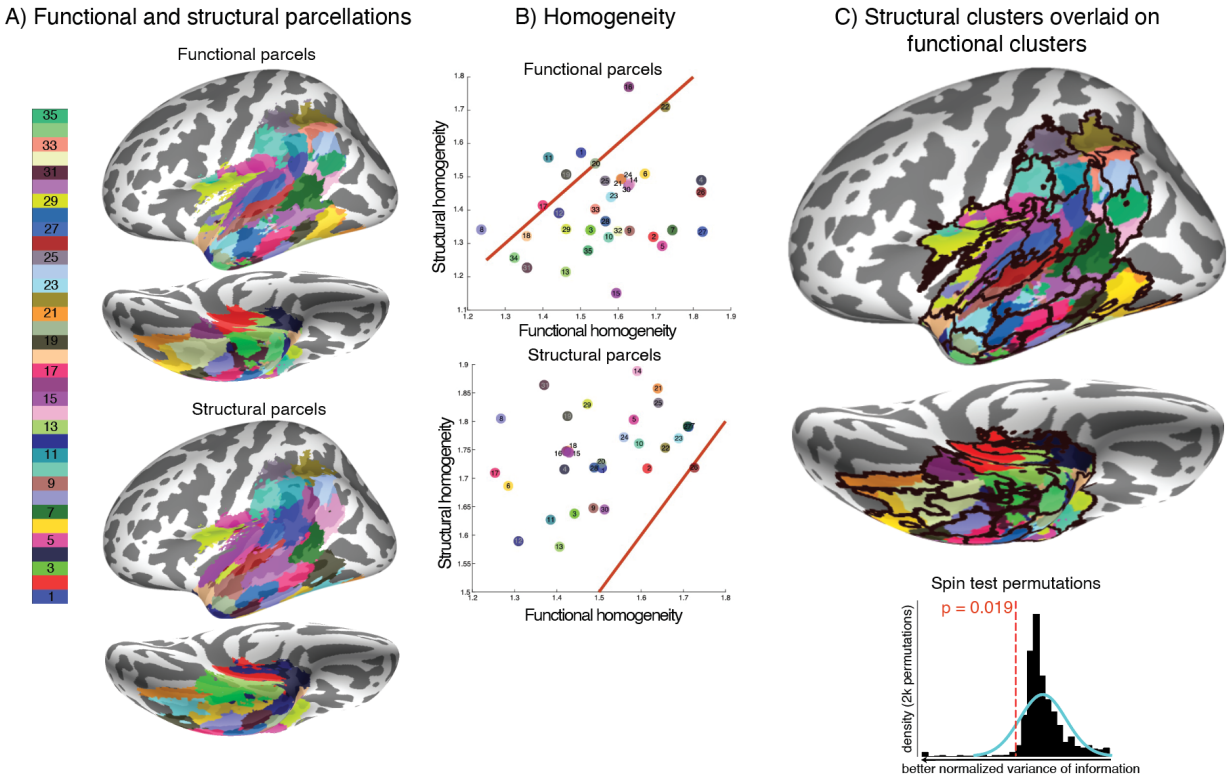

Figure 18: Functional and structural parcellations of temporal and inferior parietal cortex. Functional and structural parcellations are presented in panel A. Panel B shows structural homogeneity and functional homogeneity for structurally and functionally defined parcels. The orange line represents equal "fit" of parcels to functional and structural data. All points above the line show better fit to structural data, and all points below the line show better fit to functional data. Higher values above 1 mean that there was more intra-cluster similarity than inter-cluster similarity. Points are color coded according to the clusters they represent. In panel C, structural clusters are overlaid as outlines over colored functional clusters. The spatial correspondence between the two parcellation schemes is statistically tested using a permutation-based "spin test". The null distribution was built up by computing the normalized variance of information (NVI) between spatially permuted parcellations and original parcellations. Note, smaller NVI values are "better" (i.e., there is more similarity between parcellations).

4.3.5 Decoding functional clusters

We have shown that the structural and functional parcellations of temporal and inferior parietal cortex are meaningfully similar overall, and that there is good cluster-level correspondence between these parcellations for a few regions we have previously shown to be important for distinguishing latent speech networks. Next, we sought to decode the functional parcels in order to more systematically determine which other functions they are associated with. That is, to understand how the latent functional networks were used to carve up these functional areas.

To decode the functional clusters, we took the exemplar meta-analyses associated with each functional category and computed the mean likelihood of seeing activity within parcels (reflected as a mean z-score across voxels surviving a $p(FDR) < 0.001$ threshold). A visual representation of this decoding is presented in the hive plot in Figure 19 (the first two axes on the right). While an understanding of the functions associated with each parcel are beyond the scope of this work, we note that the results confirmed that the cluster located near the phonetic processing area we observed in the latent speech networks was in fact associated with the phonetic functional category (i.e., "speech" was the exemplar for this category, which we've interpreted as the phonetic category). Moreover, the decoding showed that this cluster has high specificity. It was only associated with phonetics, and to some extent, early auditory processing. In contrast, the early auditory functional category (i.e., "auditory" exemplar) was exclusively associated with the functional cluster that mapped onto the auditory region we previously described in relation to the lower-dimensional speech networks. In the earlier inspection of speech networks, we also

mapped two anterior comprehension areas. The more anterior of these near aSTS and aMTG mapped onto a functional cluster that was likely to activate the following exemplars: theory of mind (representing theory of mind and intention exclusively), mental states (representing mentalization, beliefs, inferences, and other social cognitive processes), abstract (representing concepts), and comprehension. Comprehension was the most likely exemplar to produce activity in the cluster. A more posterior comprehension area was mapped in anteroventral STG when inspecting speech networks. The functional cluster for this area was associated with comprehension, and language (generic language terms) exemplars. Notably, comprehension was less likely to activate this cluster than the more anterior cluster. The area of the pMTG we found to be associated with comprehension and word recognition was split into two clusters. The anterior cluster was associated with the exemplars for word recognition, comprehension, regulation (emotional), and middle temporal. The posterior cluster, which also contained an area in pSTS that we found to be involved in phonological-orthographic processing, was likely to activate for: comprehension, language, orthographic (i.e., the exemplar for the phonological speech network), middle temporal, and word recognition.

The results of this decoding clearly confirm the qualitative associations we drew between speech areas and functional clusters in the preceding section. It is worth noting that most of these clusters were highly functionally specific (other than aSTS/aMTG and pMTG) and related to speech processing. The only exception to this was the pMTG and aSTS/aMTG clusters. The anterior comprehension area appears to play a role in general conceptual processing, and is also involved in social cognitive processing, particularly theory of mind and mentalization. Additionally, the anterior portion of pMTG was

associated with emotional regulation. Thus, it appears that the area of the pMTG that is involved in word recognition and comprehension contains a posterior portion that shows relatively high specificity for speech, and an anterior portion that is still involved in speech, but also associated with other functions.

4.3.6 Structural networks

The main purpose of the structural parcellation was to find evidence that certain clusters extracted from the functional data, and related to speech, could be retrieved solely using structural connectivity. However, we also wanted to ensure that the structural networks that structural parcels were defined on corresponded to specific white matter pathways. In addition, we wanted to understand which pathways were associated with the functional clusters that mapped onto speech areas. To that end, the second and third axes on the hive plot in Figure 19 describe how functional parcels are related to structural parcels based on the percentage of their voxels that overlapped. The third and fourth axes in the plot show which structural networks were associated with each structural parcel. Using the connections in this plot, we could identify structural networks associated with each functional cluster. One clear trend revealed by the hive plot is that there was generally a one-to-one mapping between structural clusters and structural networks.

The classification of structural networks according to white matter pathways is presented in Figure 20. The XTRACT atlas (Warrington et al., 2020) was used to classify structural networks based on a combination of qualitative inspection for spatial overlap with tracts, and a measure of how strongly the structural network coefficients associated with the voxels of each tract. Generally, the tract along which structural network

coefficients were highest aligned with the final manual classification (Figure 20A). The fact that the structural network coefficients tended to be relatively specific about which tracts they showed greater magnitude for suggests that our structural connectivity analysis successfully identified existing white matter pathways. The specific tract associated with each structural cluster is presented in Figure 20B. Here, we focus only on the structural clusters related to the speech areas we mapped earlier. We note that most clusters close to auditory cortex showed exclusively local connectivity around the superior temporal plane, the operculum, and the insula. This finding is consistent with descriptions of sensory cortex being dominated by short-range connections (Eickhoff et al., 2018). The mid-to-posterior ventral STG cluster was uniquely distinguished by connectivity along the middle longitudinal fasciculus. The anterior comprehension cluster spanning aSTG, aSTS, and aMTG was uniquely distinguished with the inferior fronto-occipital fasciculus. The structural cluster associated with the word recognition area in pMTG was connected to the arcuate fasciculus. The ventral and dorsal clusters that flanked this cluster were also connected to the arcuate fasciculus. The specific networks associated with these clusters are shown in Figure 21.

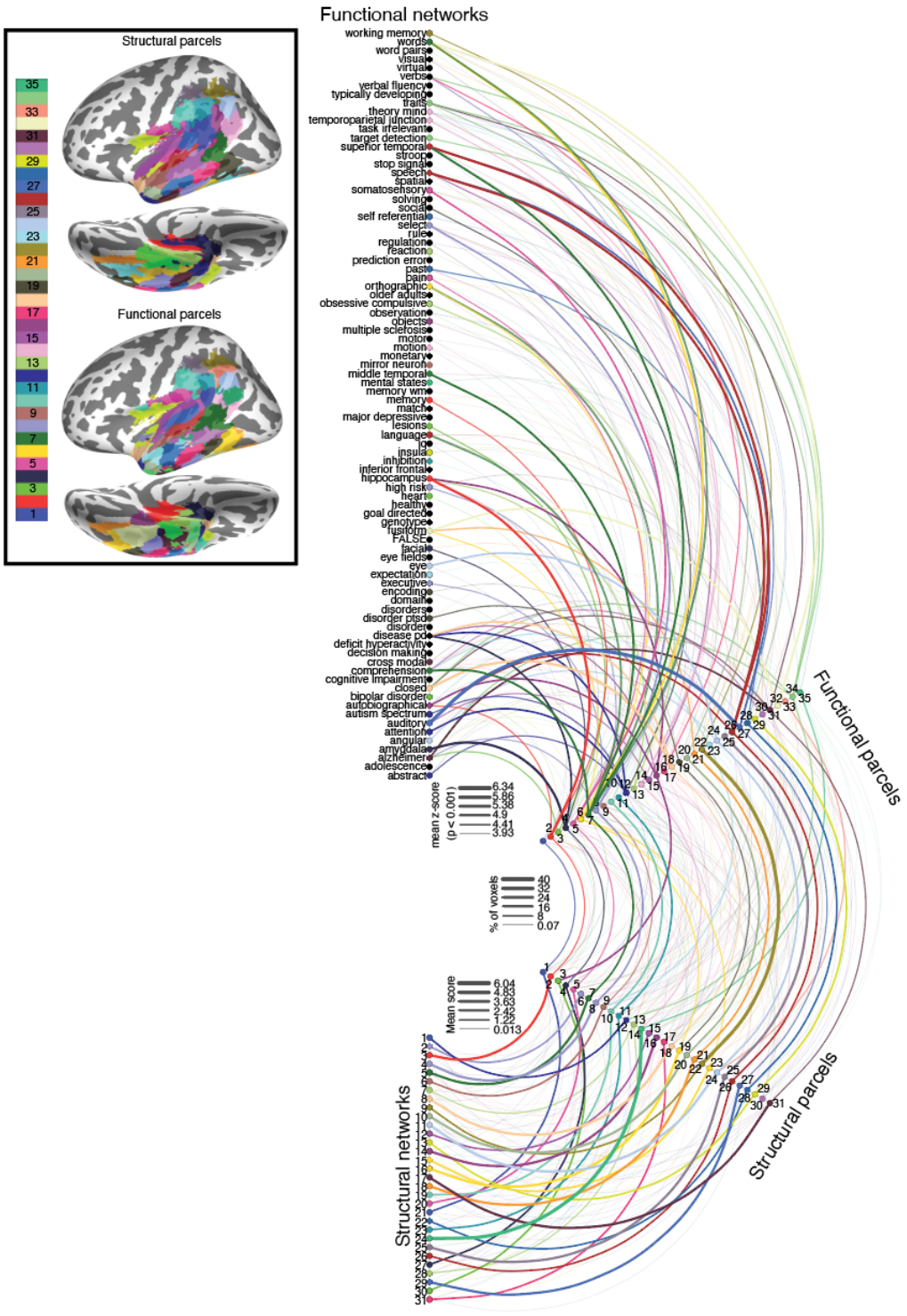

Figure 19: Relationships between parcels, functional and structural networks. A hive plot is presented showing the relationships between functional category exemplars (first axis clockwise), functional parcels (second axis), structural parcels (third axis) and structural networks (4th axis). Connections between the first and second axis show the mean likelihood of seeing activity within a functional parcel for the meta-analysis of each

functional category exemplar ($p < 0.001$). Functional category exemplar nodes on the first axis are colored based on the functional parcel that shows the highest mean likelihood of activating the meta-analysis for that exemplar. The functional parcel nodes are colored according to the parcellation scheme shown in the top left box. Connections between the second and third axis show the percentage of voxels that overlap between each functional parcel and each structural parcel. Nodes are colored according to parcellation scheme in top right box. The connections between the 3rd and 4th axis represent the highest scoring structural network on each structural parcel (a mean score is computed over the voxels in each cluster). Connection line widths vary according to the data being represented between each pair of axes (see center of hive plot).

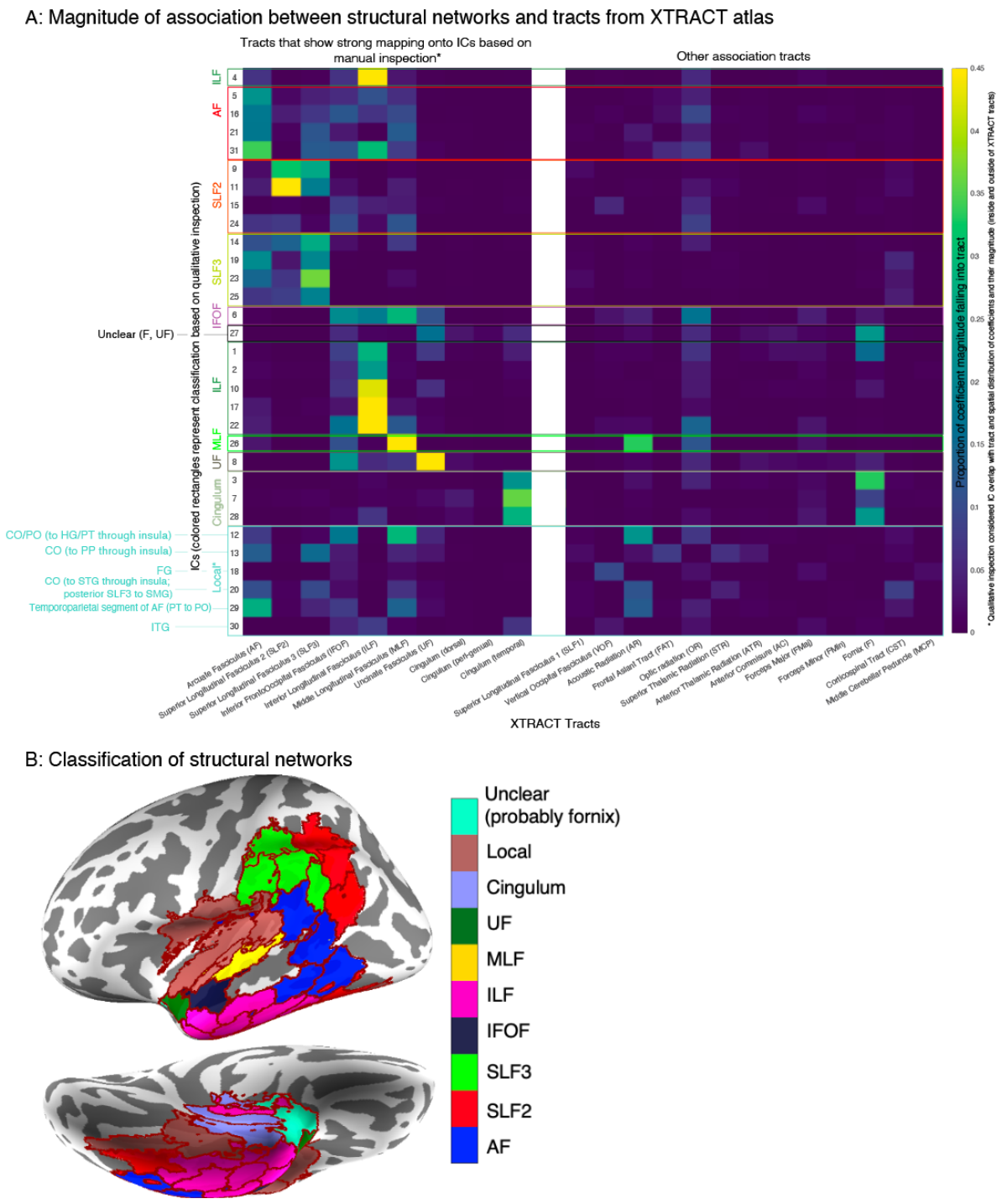

Figure 20: Categorizing structural networks. Structural networks were classified manually with reference to the XTRACT atlas. The manual classification considered both the extent of each structural networks' overlap with tracts in the atlas, and the distribution of structural network coefficients across tracts. Structural networks were corrected based on a $p < 0.001$ threshold. The magnitude of coefficients distributed among XTRACT tracts is shown in panel A (sum of coefficients within a tract were divided by the sum of all coefficients).

The rectangular colored boxes represent the result of the final, manual, classification of tracts. Structural networks were matched to their corresponding clusters to visualize the classification of tracts across all clusters (panel B). See x-axis for abbreviations.

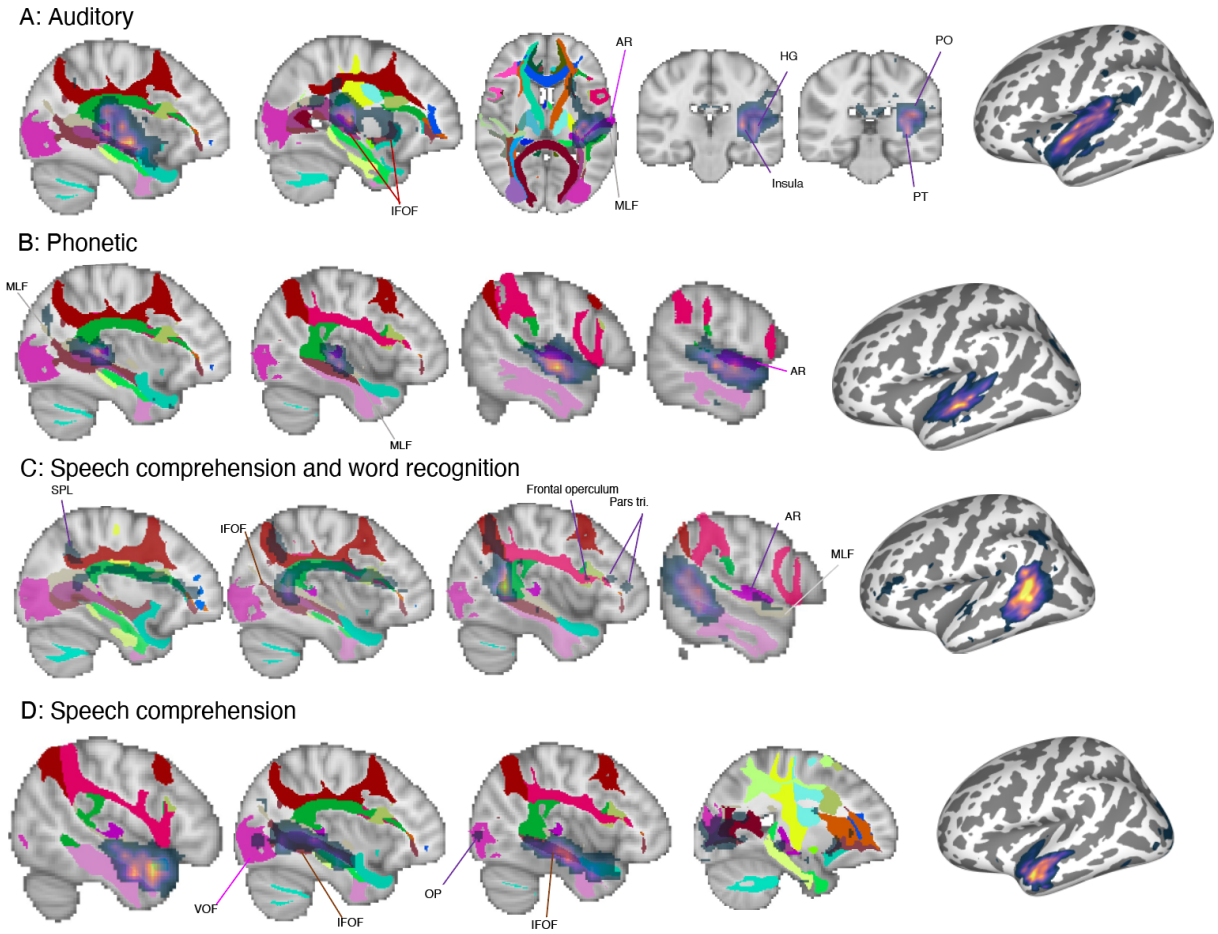

Figure 21: Structural networks related to speech. Structural networks related to the speech areas investigated by comparing latent functional networks are visualized (corrected at $p < 0.001$). Solid colors on volume-space rendering represent tracts from the XTRACT atlas. The thermal heatmap represents the structural network plotted in each panel. See Figure 20 for abbreviations of tract names. All labeled anatomical areas (e.g., PO, HG, PT, IFG pars triangularis) were confirmed using the Harvard-Oxford atlas.

4.4 Discussion

Speech processing occurs along a hierarchy, with each computational step engaging different yet interacting regions of the brain—from acoustic analysis in primary auditory cortex to phonetic processing, word recognition, and semantic analysis in neighboring higher-order areas. What remains uncertain, however, are the precise locations of these functional subregions. For instance, models of auditory processing diverge on whether speech perception involves areas anterior (Rauschecker & Scott, 2009) or posterior to Heschl's gyrus (Hickok & Poeppel, 2007). Here, we have used a purely data-driven meta-analysis that extracted unique patterns of activity associated with different groups of studies to map the networks involved in speech perception. The approach that we have taken allowed us to greatly increase the sample size associated with our meta-analysis relative to previous meta-analytic work with similar aims by relying on text-mined information to categorize studies into groups. In defining unique patterns of brain activity across all of these groups of studies, we extracted networks that cut across many different cognitive processes and tasks. This strategy allowed us to identify the networks related to speech perception in a data driven way, while at the same time mapping the full range of behavioral interventions that produce the same pattern of activity. The procedure we used to extract latent functional networks followed two steps: i) studies forming different groups were clustered together based on the relationships between their associated patterns of activity, and ii) the patterns of activity belonging each cluster, or 'functional category', were reduced in dimensionality to produce a single common pattern of activity.

In this work, we showed that studies on phonetic processing report patterns of activity that are starkly different from studies focused on early acoustic analysis, word recognition, and speech comprehension. This was evinced by the fact that these groups of studies were virtually always separated into different functional categories independent of exactly how many functional categories were carved out of the data. We then showed that a reasonable model of the number of functional categories contained in the functional neuroimaging literature separated each of these sets of studies into their own unique functional categories (i.e., where they served as exemplars), suggesting that they exhibit relatively distinct patterns of activity. In the model we investigated, we found evidence of finer structure in functional categories related to speech—additional functional categories were found that were organized around processing phonological-orthographic information, word class, word pairs, concepts, and verbal fluency. These findings demonstrate aspects of speech processing that can be distinguished based on their unique patterns of activity. Moreover, most speech related functional categories were comprised of groups of studies that appeared to be relatively homogenous, suggesting that these patterns of activity are mostly specific to speech.

We next inspected the functional networks associated with speech related functional categories that more clearly formed a processing pathway: early acoustic processing, phonetic processing, phonological-orthographic processing, word recognition, and speech comprehension. This was partly motivated because other speech categories reflected unique patterns of activity associated with highly specific tasks, or stimuli (e.g., verbal fluency [task], word pairs). This comparison between speech perception networks showed that each step in the processing hierarchy recruited additional areas that were

located increasingly more posteriorly and then ventrally in temporal cortex. First, we found that both acoustic and phonetic processing was associated with activity in Heschl's gyrus (HG), although the activity for phonetic processing was substantially weaker in this region. In contrast phonetic processing strongly involved the mid-to-posterior ventral STG, whereas acoustic processing only weakly associated with this area. That there is some overlap in activity patterns between these two sets of studies is somewhat unsurprising. The disadvantage to using frequently used phrases within studies as a proxy for cognitive processes is that the same studies may use multiple phrases at once, complicating the interpretation of the brain activity that they report. For example, some studies using the phrase 'auditory' will also use the phrase 'phonetic'. While overall there may not be much overlap between these groups of studies (i.e., most 'auditory' studies are not interested in phonetic processing), the fact that some studies within the auditory group reflect phonetic processing may explain why the latent auditory network contains some activity within the area mostly associated with phonetic processing. By grouping together studies using phrases based on their activity patterns, we were able to tease some of this overlap between groups of studies apart, but it could not be eliminated. For instance, studies using the phrases 'tones', 'audiovisual', and 'noise' may be less likely to use the phrase 'phonetic', but because their patterns of activity are similar to the group of studies using the phrase 'auditory' they may get grouped together. By allowing all of these grouped studies to generate a common network, the role of activity that is shared with some other functional category is diminished. This would predict that the latent functional networks we extract should generally show higher specificity than the individual meta-analyses that make them up. Indeed, compared to the auditory latent network, the individual meta-analysis for the

group of studies using the phrase auditory is less focal and shows relatively stronger activity in the area associated with phonetic processing.

Further, we found the same area of the mid-to-posterior ventral STG to be involved in subsequent steps along the speech processing hierarchy, including phonological-orthographic processing, word recognition, and sentence comprehension. While these other latent functional networks were all more strongly associated with activity in other areas, the fact that they all recruited the area primarily associated with phonetic processing is consistent with the fact that higher-level speech processing, including word recognition and comprehension, requires phoneme perception (e.g., Binder et al., 2017). More complex stages of speech perception involved regions outside of the latent phonetic network. Phonological-orthographic processing, word recognition, and speech comprehension all involved the dorsal bank of the posterior superior temporal sulcus. All of these aspects of speech processing involve phonological processing, implicating the posterior superior temporal sulcus in phonology. A role for the posterior superior temporal sulcus is consistent with prior work showing that this area exhibits greater response when processing larger amounts of phonemic information (e.g., Okada et al., 2006; Liebenthal et al., 2005; Specht et al., 2009). Only speech comprehension and word recognition were strongly associated with the posterior middle temporal gyrus. This is consistent with a role for the posterior middle temporal gyrus in lexical access, which involves mapping the phonological forms of words onto their meanings (Hickock & Poeppel, 2016). For instance, functional neuroimaging studies frequently report left pMTG activation in various semantic tasks (Indefrey & Levelt, 2004; Binder et al., 2009 and Vigneau et al., 2006 for review), which has led to the view that it is involved in lexical-semantic processing, facilitating

'conceptually-driven' lexical selection (e.g., Indefrey & Levelt, 2004). Notably, both word recognition and phonological-orthographic processing was associated with portions of the fusiform gyrus and these functional categories included visually presented stimuli. This implies activity in fusiform gyrus is driven by visual processing of speech stimuli, while activity in posterior superior temporal sulcus reflects phonological processing. Finally, sentence comprehension was the only aspect of speech processing that additionally recruited anterior portions of the temporal lobe. This is consistent with prior work showing that the anterior temporal lobe is generally involved in syntactic and other processes necessary for sentence-level comprehension (Rogalsky, 2016). In summary, we find evidence that speech perception occurs along a hierarchy, with each step recruiting additional regions that appear to form a core speech processing network.

Critically, we mapped a speech perception pathway that appears posteriorly directed. While past fMRI and meta-analytic work has largely supported the notion that phonetic processing occurs around mid-STG (DeWitt & Rauschecker, 2012; Tukeltaub & Coslet, 2010), some studies have suggested that processing word-level and sentence-level speech stimuli recruits more anterior areas of the superior temporal gyrus (DeWitt & Rauschecker, 2012). Our results, which synthesize tens of thousands of functional neuroimaging studies, do not support an anteriorly directed speech perception pathway. Moreover, we report phonetic processing in a relatively more posterior portion of the STG, one that is clearly posterior to HG. Thus, our findings are more in line with certain meta-analytic work that has supported sublexical speech perception within posterior superior temporal sulcus (Tukeltaub & Coslet, 2010). There may be several reasons for which some of the prior meta-analytic work has failed to capture the results that we present here,

including aforementioned issues of sample sizes. However, we point out that this work did demonstrate consistent patterns of activity across word-level and sentence-level speech stimuli in more posterior superior temporal sulcus and suggest that the stronger effects observed in anterior areas within this work may reflect response that is less functionally specific (i.e., both our individual meta-analyses and the process of forming functional categories involved steps that improve estimates of specificity). We also note that, in general, our findings are closely aligned with proposals of a ventral sensory-conceptual stream for speech perception that places phonological processing within the posterior superior temporal sulcus, lexical processing within the posterior middle temporal gyrus, and higher-level comprehension processes in the anterior temporal lobe (Hickok & Poeppel, 2007). Finally, we report that the functions associated with each of these areas as mapped in the current work appeared to be robust to effects of publication bias.

Aside from mapping latent speech networks, we also analyzed the extent to which the regions within these networks were involved in other functions. We accomplished this by performing a data-driven parcellation of temporal and inferior parietal cortex using the entire set of latent functional networks that we mapped. This analysis carved out functional areas with similar associations across latent functional networks. We found that one of the most functionally homogenous clusters delineated by this analysis was the mid-to-posterior ventral STG implicated in phonetic processing. Mapping its association to other latent functional networks revealed this area to show high specificity for the speech latent network. Further, this analysis mapped the anterior areas that we previously described to be associated with comprehension. These areas showed the strongest association with the comprehension latent network, but also with mentalization and theory of mind. Notably,

the semantic processing network and the default mode network have been shown to have substantial overlap (Binder et al., 2009). Unfortunately, the functional parcellation could not distinguish the posterior middle temporal gyrus (pMTG) and posterior superior temporal sulcus (pSTS) areas we found to be involved in word recognition and phonological processing, we suspect this was the case because only few functional categories presented activity within the pSTS and not the pMTG. Overall, the decoding of this pSTS-pMTG cluster was consistent with our observations when inspecting speech networks—these areas were relatively specific to comprehension, phonological processing, and word recognition. One intriguing finding that we report is that the large portion of the pMTG that we found to be associated with these processes could be distinguished into two areas—a more anterior area within pMTG and a posterior area (i.e., the pSTS-pMTG cluster). Curiously, both clusters were associated with comprehension, but the anterior one appeared to be functionally distinguished based on its weaker association to comprehension and stronger association to emotional regulation. While this finding is interesting, it is hard to speculate precisely what functions may be carried out within this anterior cluster.

The functional parcellation that we report was validated using an independent structural parcellation. Remarkably, we find these two parcellations to show statistically meaningful correspondence, adding to a large literature that has shown striking similarities between structural connectivity and functional data (e.g., Greicius et al., 2009; Hermundstad et al., 2013; Honey et al., 2009; Liegeois et al., 2020; Skudlarski et al., 2008). However, much of this kind of work has focused on relating functional and structural connectivity, and here we have shown this to be the case for functional areas extracted

from meta-analytic data. Overall, this correspondence illustrates that although a large number of subtle differences in the patterns of temporal and inferior parietal activity that are reported in the neuroimaging literature are difficult to interpret, they are grounded in tangible differences between structural networks in the brain. We report also that one of the most prominent similarities between the areas delineated by structural and functional data was found in the mid-to-posterior ventral STG associated with phonetic processing. The structurally defined boundaries of this area showed the best fit to functional data compared to other structural areas. This provides further evidence that our localization for early speech perception is robust. Further, the connectivity data implicates the middle longitudinal fasciculus (MLF) in auditory processing, as connectivity along this tract uniquely distinguished mid-to-posterior ventral STG from other areas. While the MLF has been proposed to be a language-related tract (Schmahmann and Pandya 2006), some research has suggested it may instead be involved in the dorsal auditory stream (Wang et al., 2013) based on the putative functions of the areas it connects. Our finding aligns with more recent work indicating that integrity of the MLF is associated with response bias in speech perception (Tremblay et al., 2019). Finally, the connectivity data was also able to marshal support for a lexical-semantic area within pMTG and indicated that this area was distinguished by connectivity to the arcuate fasciculus (AF). While our structural parcellation did not delineate the pSTS area associated with phonological processing, our findings generally align with the observation that a ventral segment of the AF that terminates in the pMTG (as opposed to pSTG) is involved in lexical-semantics (Riling et al., 2008).

Chapter 5

Predictive coding and internal error correction in speech production

5.1 Introduction

The posterior portion of middle temporal gyrus (pMTG) sits at a crucial junction in temporal cortex, nested just below superior temporal regions involved in processing speech, and at an intersection of white matter tracts involved in language (e.g., Griffis et al., 2017). Its specific pattern of connectivity (Turken & Dronkers, 2011) resembles a widely distributed network of regions previously implicated in semantic processing (Binder et al., 2009). However, the precise function of this region has remained elusive. Evidence from stroke has suggested that this region may be important for lexical processing on the basis that damage in this location results in comprehension deficits, typically at the word-level (Dronkers et al., 2004; Bates et al., 2003; Pillay et al., 2017; Thothathiri et al., 2012). This has produced one view that sees the pMTG as an area that integrates lexical and semantic information by mapping sounds onto meanings (Hickok & Poeppel, 2007). At the same time, other work has implicated the pMTG in relational semantics (Humphreys & Lambon Ralph, 2014), control of semantic retrieval (Badre et al., 2005), comprehension of actions and events (Liljestrom et al., 2008), and a number of other functions. That there is complexity in the response properties of the pMTG is highlighted by the previous study in

this dissertation, which demonstrated that a large area in the pMTG that activates in both sentence comprehension and word recognition studies could be teased apart into two areas—one that responded more specifically to comprehension and word recognition studies, and another that responded to a broader range of tasks, making it more difficult to characterize its function. Thus, efforts to synthesize the broader functional neuroimaging literature have not been able to fully describe the role of pMTG in speech processing. A recent task-fMRI study has reported *suggestive* evidence that the pMTG may be involved in error correction for speech (Okada et al., 2018). The current research investigates this intriguing possibility by attempting to replicate and extend this prior work.

5.1.1 Error correction in speech

Speaking is a deceptively complex task involving several computational stages: selecting words from a mental dictionary that can contain tens of thousands of entries, correctly accessing and coding the sequence of sounds that could have many possible permutations, and executing the motor commands to reproduce those sounds with coordinated movements of several independent articulators within the vocal tract (Levelt, 1989). Given the system's complexity, the ample opportunities for error, and the fact that speech is articulated at a rate of approximately 5 syllables per second, it is remarkable that the vast majority of words are accurately produced (estimates put the number at approximately 99.9%; Jacewicz et al., 2010; Levelt, 1992). How is this achieved? One possibility is that the mechanism is so exquisitely tuned that it simply makes very few coding errors. Another possibility is that coding errors occur more frequently, but are unconsciously detected and corrected internally, prior to ever being spoken (Hickok, 2012; Levelt, 1983; Nozari et al.,

2011). The latter possibility is consistent with the idea from the motor control literature that the brain simulates the position and trajectory of the motor effector it is controlling, a so-called forward internal model, as a mechanism for improving the speed and accuracy of movements via predictive coding (Kawato et al., 1999; Shadmehr et al., 2008; Wolpert et al., 1995). Internal predictive coding may provide a mechanism to detect and correct speech errors prior to producing them (Hickok, 2012).

Although talkers can detect and correct *overtly* produced speech errors, evidence for *internal* error correction in speech is limited to inferences based on the timing of error corrections. For example, overtly realized error corrections such as, 'v-horizontal', have been argued to occur too quickly to be accomplished using overt feedback alone, which in turn suggests the existence of at least an internal mechanism to detect, if not correct errors (Nozari et al., 2011; Nooteboom et al., 2005). Further evidence for internal error *detection* comes from electrophysiological measures, which have identified signals that predict speech errors prior to their vocalization (Möller et al., 2007). Strong direct evidence for internal error *correction* is sparse, however.

In one fMRI study, Okada et al. (2018) reported suggestive evidence for the existence of an internal error correction mechanism. In their experiment, Okada and colleagues (2018) presented participants with tongue twister sequences that were designed to bias speech errors towards either words (REEF LEECH → LEAF REACH) or non-words (WREATH LEAGUE → LEATH REEG; Alvarez et al., 2008). Behavioral research on slips of the tongue has shown that non-word errors are more rare than real word errors, the "lexical bias" effect (Baars et al., 1975; Dell, 1986; Nooteboom et al., 2005; Levelt et al., 1999). Several previous behavioral studies have shown that the lexical bias effect holds

even when subjects do not phonate their speech and self-report their errors (Alvarez et al., 1008; Corley et al., 2010; Oppenheim et al., 2010). Although not the main focus of their study, Okada, et al. (2018) reported a lexicality effect in the left pMTG, a region implicated in lexical-level processes. That is, they found greater activation in MTG when participants recited tongue twisters that were biased to produce non-word errors compared to word-errors, even *on trials in which participants responded accurately*. This is particularly interesting because no speech errors were committed on these trials - the activation differences observed in MTG reflect the potential for a word vs nonword error. The authors suggest that speech errors resulting in non-words are more readily detectable and therefore would be more likely to be internally corrected prior to speaking. The fact that they found a clear effect of lexical status of the error bias even under conditions of accurate performance demonstrates that the system detected the distinction internally, and this could only be the case if in fact an internal error was committed and then corrected prior to accurate output. Although this is a very interesting finding, the pMTG activation in their study did not reach statistical significance corrected for multiple comparisons, and as the authors note, this work requires replication. Runnqvist et al. (2020) recently reported on a study that used a very similar design to Okada et al. (2018) and while they reported evidence of internal error correction in the cerebellum, did not detect an effect in the cerebral cortex.

5.1.2 Current work

The present research builds on this prior work by attempting a direct replication of Okada et al. (2018) along with an extension to potentially render the paradigm more sensitive to

detecting evidence for an internal error correction mechanism. The experiment we introduce uses the same tongue twister stimulus list and the same tasks (silent articulation and imaging) as previous FMRI work, but additionally includes a set of stimuli designed to increase the load on internal error detection and correction, thereby increasing our chances of observing evidence for such a process. Previous work has shown that tongue twisters with the potential to induce taboo word slips (e.g., FULL BUD BUCK FUSS) elicit significantly fewer slips than neutral tongue twisters, suggesting a higher rate of internal error detection and correction (Motley et al., 1982). We hypothesized that tongue twisters biased towards taboo word errors would elicit greater activation in pMTG compared to non-taboo word errors. The basis for this is that detecting and correcting word-level speech errors presumably drives activity to areas critical for lexical processing, and that prior work has found suggestive evidence for internal error correction in pMTG using tongue twisters (Okada et al., 2018). We emphasize that the lexicality effect (i.e., non-word errors > word errors) in pMTG that trended towards significance in at least one study shows evidence for an internal error correction process at work because no external speech errors were committed while participants recited the same set of words, either arranged to bias errors towards non-words or words. Thus, the lexicality effect in pMTG reflects a neural signature associated with the potential for a non-word error, and this could only occur if a process was internally correcting errors prior to speech output. Further, a recent fMRI study used taboo words in a word-picture interference paradigm and found that taboo words lead to greater interference in naming (slowed response times) and greater activation of the pMTG (Hansen et al., 2019), consistent with our reasoning.

In addition to mapping an error correction mechanism in the brain, we aim to replicate Okada et al's finding that forward predictive signals are generated in auditory cortex in a speech production task. In their study, Okada et al (2018) had participants silently recite a sequence of tongue twisters in an FMRI experiment. Two speech production conditions were included, one in which speech was articulated without phonating (silent articulation) and one in which speech production was imagined without articulation (imagined). Both conditions were matched for acoustic input (i.e., no speech input). Previous behavioral research has shown that these two tasks engage different levels of linguistic/motor planning. Imagined speech engages lexical-level processes but not lower-level phonological processes whereas silently articulating speech engages both levels of processing. Therefore, engaging motor-phonological processes should generate a forward prediction of the acoustic consequences of the executed (silent) speech whereas engaging lexical-level processes should not. As expected, a contrast of silently articulated speech compared to imagined speech revealed activity in left inferior frontal gyrus and premotor cortex, areas involved in speech articulation. More interestingly, they found robust activity in bilateral auditory cortex when motor articulators were engaged, but not when speech was imagined, and this activation was present in the absence of external auditory stimulation. The authors suggest that these activations reflect stronger forward predictions generated in the articulation condition compared to the imagining condition.

The main goal of the present research is to examine a speech production mechanism that has been elusive thus far: neural evidence of internal error correction. To that end, we leverage a tongue twister paradigm that has been previously used to generate suggestive evidence of internal error correction during speech in the pMTG. Using the same stimulus

set and design as this prior work, we attempt to replicate evidence for internal error correction in a much larger sample of participants. Critically, we also improve our odds of finding evidence of internal error correction by introducing additional stimuli designed to tax the error correction mechanism: tongue twisters that elicit taboo word errors. We also attempt to replicate the predictive coding effect reported by this previous work. Replication is particularly important since the aforementioned study was the first FMRI experiment to show evidence of forward predictive signals involving auditory cortex, a sensory region that plays an important role in speech production.

5.2 Methods

5.2.1 Subjects

Forty participants (25 females) between 18 and 40 years of age were recruited from the University of California, Irvine community. Participants received monetary compensation for their participation. The volunteers were right-handed, native English speakers with normal or corrected-to-normal vision, no known history of neurological disease, and no other contraindications for MRI. Informed consent was obtained from each participant prior to participation in the study in accordance with guidelines from the local ethics committee that approved this study. A handful of participants were excluded from analysis for excessive head motion (N=2) and early scanning termination for a variety of reasons (N=4; e.g., claustrophobia, late start/incomplete scanning), leaving a total of 34 participants to contribute to the results.

5.2.2 Power analysis

The number of participants we recruited for this study was anticipated based on a power analysis of pilot data collected for 5 participants (1 participant excluded from analysis due to early termination of scanning). The power analysis was carried out over two regions of interest (ROIs) using the fMRIpower toolbox (fmripower.org). Anatomical ROIs for this analysis were selected from the Harvard-Oxford cortical atlas, and adjustment for Type 1 error was made by applying a Bonferroni correction based on the number of ROIs being compared ($p < 0.005$). The power analysis indicated that approximately 40 participants are necessary to achieve 80% power for detecting a predictive coding effect of size 1.0623 within Heschl's gyrus (effect sizes expressed in standard deviation units, which is analogous to Cohens D), as well as both a lexicality effect of size 0.3815 and an internal correction effect of size 0.3985 within pMTG.

5.2.3 Stimuli and Task

Scanning took place at the facility for imaging and brain research at University of California, Irvine. Participants were scanned while they recited a set of four words (e.g., lean reed reef leach) in sync with a visual metronome. Thirty-two sets of tongue twisters used in previous experiments were employed in the current study (Okada et al., 2018; Oppenheim et al., 2010). These tongue twisters are known to behaviorally elicit a lexical bias effect. Lexical bias refers to the tendency for word errors to create a real word instead of a non-word (e.g. target word is "reef" but slips to "leaf", is more likely than if target word is "wreath" and slips to "leath" because leath is a non-word). These stimuli were designed so that if an error occurred on the 3rd word of each sequence, the outcome would yield either a real word error ("lean", would induce the error "leaf" instead of "reef") or a nonword error ("lean"

would induce the error “leath” instead of “wreath”. In addition to these tongue twisters, we included 32 taboo tongue twisters (see Table 1 for example of stimuli). In a behavioral pilot, we found that these taboo tongue twisters were effective in eliciting speech errors (non-taboo error rate = 23 %, taboo word error rate = 20%).

Table 1: Stimulus Set

non-Taboo tongue twister	*Taboo tongue twister*
nod mod mock knob	cod mod mock cob
mine bikes bit mice	dine bite bike dice
bike wild wise bile	dial bile bike dies
bane gave gan bait	shave bane bit shank
name make mail nag	fake name nag fail
wing bib bit whip	shine bib bit ship
six finch fill sin	ding sin sick dill
jail cheek cheap jean	dale chip chick dean
lean reed reef leech	queen reed reef queer
yore wan wok yawn	core mud mum caught
gun bulb buck gull	fun bulb buck full
singe fib fish sip	nib singe sip knit
sing hitch his sick	ditch sing sick diss
zinc niece need zest	pink niece need pest
jog mod mock job	cot mod mock cob
job rob rock jot	call rob rock cot
van match mat verve	shack match mat shave
lull nudge buck love	full bud buck fun
zing bib bit zip	ting bib bit tip
hinge fib fit hip	shin fib fit ship
sing that them zed	shing them that shed
daft gab gas dam	dab laugh lamb dan
than bunk nuzz there	jan bid bizz jar
chicks fich fizz chin	jicks finch fizz gin
king hitch his kcik	ping hitch hiss pick
nab match mat nerve	shab volt vat shot
gun bulb but gull	con grub grunt cup
pen bunk nus pair	fend bus buck fair
zing that then zed	shang then that shed
gore wan watt gone	whole gone gore hat
goon nab nap gar	food tab tuck far
nun bulb but null	pun mull miss pull

The present study followed the experimental procedure outlined by prior work (Okada et al., 2018). On each trial, a tongue twister phrase was visually presented on screen for 3 seconds, and then subjects were cued to silently articulate the sequence or imagine saying the sequence without mouth movements (see Figure 23). The presented

cue was a cartoon face that remained on screen for 500 ms and contained a red arrow pointed either to the head or to the lips. An arrow pointing to the head cued the participants to imagine saying the word, and an arrow pointing to the lips cued the subjects to silently articulate the words. A red fixation appeared on screen 500 ms after cue offset and served as the visual metronome, flashing at a rate of 2/s. Participants recited one word per fixation in sync with the metronome. The interstimulus interval was 500 ms. After recitation, participants indicated with a button press if they were correct or incorrect on the sequence. Recall failure was treated as an incorrect response to ensure that only trials where rehearsal was successful would be analyzed. Participants were also instructed to indicate an incorrect response if overt production occurred accidentally. Prior to scanning, participants spent roughly five minutes (more when necessary) practicing silently reciting words in a way that minimized but did not eliminate articulatory movements. Continuous feedback was provided by lab staff during this practice period.

A single trial in the experiment was 8 seconds in length and there were approximately 42 trials in each session. Each session consisted of an equal number of tongue twister phrases biased to produce word errors or non-word errors and taboo errors. There were 8 experimental sessions and each session consisted of approximately 14 trials of each type, which were randomly presented along with 6 rest trials (fixation). The study started with a high-resolution structural scan. This was followed by a short practice session of the experiment using approximately 10 trials to further familiarize subjects with the task. Scanning was conducted during the practice session to acclimatize subjects to the fMRI environment, as well as to monitor head movement and provide feedback prior to the start of the experiment. Participants also received feedback about head movement in

between sessions based on qualitative assessment of the images that had been collected. The study lasted approximately 1.5 hours. Stimulus presentation and timing was controlled using PsychToolbox implemented in Matlab (Mathworks, Inc, USA).

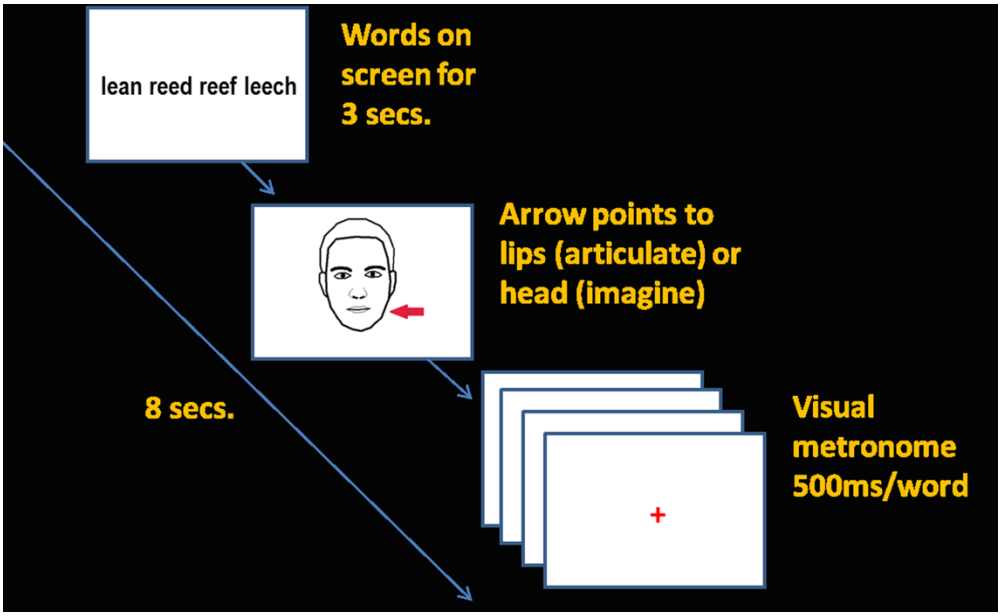

Figure 23: Example of a single trial. Subjects were presented with tongue twister sequence which remained on screen for 3 seconds, followed by a cue to either articulate the sequence or imagine the sequence. They recited each word in sync with the visual metronome.

5.2.4 Imaging

Imaging data were collected on a 3T Siemens Prisma scanner (Siemens Medical Solutions) equipped with a 32-channel RF receiver head coil. A single T1-weighted MPRAGE sequence was acquired (matrix = 256 x 256 mm, TR=2.3 s, TE = 2.32 ms, flip angle = 8°, size = 0.937 x 0.937 x 0.9 mm). An echo-planar imaging pulse sequence was collected for each of the 8 experimental sessions and the practice session (matrix = 100 x 100 mm, TR=2 s, TE = 35 ms, flip angle = 90°, size = 2.4 x 2.4 x 2.4 mm, 56 slices).

5.2.5 Data analysis

Data was preprocessed and analyzed using the FMRIB's Software Library (FSL; Smith et al., 2004). First, rigid-body motion correction was performed with FSL's intra-modal motion correction tool (MCFLIRT) using the normalized correlation cost function and the middle volume as the initial template. Participants with excessive head movement (> 0.3 mean framewise displacement across runs) were eliminated from further analysis. Mean framewise displacement amongst the remaining participants (N=34) was relatively low (M=0.15; SD=0.07). Echo planar images were high pass filtered by calculating the minimal period that retains 90% of the variance in the design matrix regressors (this amounts to a roughly 0.01 hz cutoff). These images were then spatially smoothed using an isotropic 8-mm full width half maximum (FWHM) gaussian, and the anatomical image for each subject was coregistered to their middle EPI volume. Data analysis was performed with FSL's FMRI Expert Analysis Tool and proceeded in three steps: i) modeling within-session parameter estimates for events of interest using fixed effects, ii) using these parameter estimates in a between-session analysis to model participant mean response, and iii) using mean participant response in a between-subjects analysis to model group response using mixed effects (FSL's FLAME1+2). Parameter estimates for each participant were transformed into standardized space using the MNI152 template.

Regressors for events of interest were created by convolving the predictor variables representing the time course of stimulus presentation with a gamma variate function. As in the previous experiment, regressors modeled the following experimental trial types: "Articulation: Nonword Errors", "Articulation: Word Errors", "Articulation: Taboo Errors",

"Imagining: Nonword Errors", "Imagining: Word Errors", "Imagining: Taboo Errors". All trials on which participants reported making an incorrect response, the visual presentation of words on all trials, and the 6 motion parameter estimates determined during the realignment stage of preprocessing were included in the model as nuisance regressors. Regressors were used to generate parameter estimates for each condition of speech task (i.e., imagined and silently articulated) and error type (i.e., tongue twisters biased towards: nonword, word, and taboo word errors). Parameter estimates for each speech task modeled tongue twisters of all error types and parameter estimates for each error type modeled both silently articulated and imagined tongue twisters. Multiple contrasts were set up using these parameter estimates. As the primary goal of the present research was to seek evidence for internal error correction, we first tested for the lexicality effect described by Okada and colleagues (2018) followed by a taboo effect based on the new stimuli introduced in this study, and finally the predictive coding effect also described by Okada et al.[13]. Identical to prior work, the lexicality effect was based on the contrast between parameter estimates for the nonword and word tongue twister conditions. The same logic was extended to the novel taboo stimuli and the taboo effect was based on the contrast between taboo and word tongue twister conditions. The predictive coding effect was based on the contrast between the two speech task conditions (i.e., silently articulated versus imagined). For each of these three contrasts an additional analysis was performed contrasting each condition in the pair against baseline fixation. The contrasts between each condition and baseline fixation were used to mask the contrast maps, allowing us to distinguish brain areas that show a significant difference in BOLD response between conditions but overall show below-baseline response.

Consistent with previous research, we expected to find taboo and lexicality effects in pMTG, and a predictive coding effect in auditory cortex. Because our hypotheses target specific regions, we complemented whole-brain analyses with a region of interest (ROI) approach that compared mean condition-level parameter estimates used in the contrast analyses within predefined anatomical areas. Left and right hemisphere anatomical ROIs were extracted from the Harvard-Oxford atlas (Desikan et al., 2006). The pMTG area was used to test for the lexicality and taboo effects, and the Heschl's gyrus area was used to test for the predictive coding effect. Group effects within ROIs were tested with paired t-tests carried out across participants. Tests were additionally carried out across voxels for each participant to characterize the consistency of the effects. Significance in the ROI analyses was based on a Bonferroni corrected p-value threshold of 0.05, and significance for whole brain voxelwise testing was based on a voxelwise p-value threshold of 0.01 and a GRF-based cluster size p-value threshold of 0.05.

In addition to the core analyses detailed above, we present post hoc analyses that give further context for some of our findings. We emphasize that these secondary analyses are incidental by separating them from the main results section. The aim of the post-hoc analyses is to better functionally characterize the network of regions associated with the internal error correction effects (i.e., lexicality and taboo effects), particularly the wider network capturing regions where neural response increased during the nonword or taboo conditions relative to the word condition but overall remained below-baseline. Functions associated with this network and with each of its constituent regions were probed using the Neurosynth meta-analytic database (Yarkoni et al., 2011). Note, we did not rely on the work from the second study of this dissertation to decode the results here. The motivation

for this decision was to minimize the reliance on the assumptions made in that study for decoding the current results. This was especially important since some of those modeling assumptions led us to develop this work. Thus, here, we have elected to perform decoding in Neurosynth in a much more straightforward way. However, to complement the methods that have been used to decode areas so far in the first two studies, we introduce an analysis of posterior probabilities for reverse-inference based brain-behavior relationships. Although the meta-analyses performed in the first two studies were also based on the framework of reverse-inference, we chose to analyze z-scores rather than direct estimates of posterior probability.

Network-level decoding was performed by computing the Pearson correlation coefficient between contrast activation maps from the current study and each of the meta-analyses that were generated for the 3,228 terms frequently used in the neuroimaging literature and embedded in Neurosynth. Briefly, performing a meta-analysis for each term involved separating all studies in Neurosynth into two groups: those that used a particular term frequently (minimum rate of 1/1000 words which has been shown to control for incidental word usage; Yarkoni et al., 2011) and those that didn't. Next, a search was performed for voxels where activity was more consistently reported in the set of studies that do frequently use the term relative to those that don't. This was accomplished by extracting the activation tables from these two groups of studies, creating contingency tables at each voxel that described whether activity was present and whether a phrase was used, and then performing a chi-square test. Due to their smaller size, regions were decoded in a slightly different way—by computing the mean posterior probability that a phrase was used within a study if activity was observed in each of a regions' voxels.

Posterior probability estimates assumed a uniform prior (i.e., all terms are equally likely to appear) and were generated for each term in Neurosynth. More complete details about how posterior probability was computed can be found in prior work (Yarkoni et al., 2011).

5.3 Results

5.3.1 Core findings

5.3.1.1 Internal error correction effects

Our first analysis aimed to replicate the lexicality effect described in previous work (Okada et al., 2018). Critically, all analyses we present were restricted to error-free trials. A group-level contrast between the nonword and word conditions of the experiment revealed no significant differences in brain response ($p < 0.05$; cluster corrected at $p < 0.05$). Additionally, we tested whether the mean parameter estimates within the pMTG ROI differed between these two conditions but found no significant difference across participants in either hemisphere ($p < 0.05$). Because no effect was observed at the group-level, we did not investigate how consistently the effect appeared within participants.

We next evaluated whether the novel taboo stimuli generated evidence for internal error correction by contrasting the whole-brain parameter estimates for the taboo and word conditions. As we anticipated, the taboo condition appears to have successfully increased the load on internal error detection and correction. Although no areas of the brain showed significantly greater response for the word condition than the taboo condition, relatively higher response for the taboo condition was found in a wide network

that included the posterior middle temporal gyrus (pMTG) bilaterally ($p < 0.01$, cluster corrected at $p < 0.05$; see Figure 24A). A more comprehensive description of brain regions in this network was provided by registering the contrast map to anatomical areas of the Harvard-Oxford cortical atlas (Table 2). Incidental overlap between the contrast map and anatomical areas as a result of activity spilling over an anatomical boundary in a way that is inappreciable was deemphasized by focusing only on those areas in which more than 5% of voxels showed a significant difference between conditions. For visualization, both the atlas and the contrast map were projected onto the fsaverage inflated cortical surface using a recently developed procedure that implements registration fusion with Advanced Normalization Tools (ANTs) to improve projection accuracy[29] (Figure 24B). The network of regions that responded more strongly to the taboo than the word condition spanned the bilateral frontal poles (FP), bilateral frontal medial cortex (FMC), bilateral superior frontal gyrus (SFG), right anterior cingulate gyrus (aCG), bilateral posterior cingulate gyrus (pCG), bilateral paracingulate gyrus (PaCG), right subcallosal cortex (SC), bilateral precuneous cortex (PreC), bilateral anterior middle temporal gyrus (aMTG), bilateral pMTG , left posterior inferior temporal gyrus (pITG), bilateral angular gyrus (AG), and bilateral superior lateral occipital cortex (sLOC; Figure 24A and 24B, Table 4).

The regions that responded more strongly to the taboo than the word condition were then evaluated based on whether they showed above-baseline response during the taboo condition. Our first approach was to mask the contrast map between taboo and word conditions by the contrast map for the taboo condition (i.e., taboo > word AND taboo > baseline; both contrasts set to $p < 0.01$, cluster corrected at $p < 0.05$). A 20-voxel cluster-extent threshold was applied to the resulting map, revealing that the left pMTG was by far

the largest area to show significantly greater response for taboo than word conditions while also responding significantly above-baseline to the taboo condition (see Figure 24C and Table 3). The cluster of activity that centered on left pMTG crossed only superficially over the boundary between this area and pSTG, temporooccipital middle temporal gyrus (toMTG) and temporooccipital inferior temporal gyrus (toITG) as reflected by percent overlap between these areas and the masked contrast map (see Table 3). We found additional clusters that peaked in pMTG below the cluster-extent threshold but note that we also found clusters below this threshold which peaked in right AG and left sLOC. These small clusters peaking in pMTG and sLOC overlapped with pITG and AG respectively.

In an additional analysis, we averaged the parameter estimates for the taboo and word conditions (i.e., taboo > baseline, word > baseline) within the portion of each anatomical area that exhibited a significant difference between these two conditions (i.e., the contrast map from Figure 24A/B). This areal analysis was used to estimate taboo effect size and provided an additional glimpse into which areas differentially responded to the taboo and word conditions while simultaneously showing a stronger preference for baseline fixation than the task. Overall, this analysis highlighted the same group of areas as having both greater response to the taboo than word condition and above-baseline response to the taboo condition: bilateral pMTG, left AG, left pITG, and left aMTG (see Figure 24D). Further, we report medium-to-large effect sizes (i.e., taboo versus words) in all areas, with the pITG showing the largest effect size in the left hemisphere, and PreC showing the largest effect size in the right hemisphere (Table 4). Areas that additionally responded above baseline to the taboo condition all showed large effect sizes. No hemispheric differences were found for effect size (*left:*M=1.17, SD=0.6; *right*: M=0.88,

SD=0.33; t(9)=1.04, p=0.32) or the spatial extent of activity within areas (*left:*M=22%, SD=16.18%; *right*: M=19.77%, SD=16.13%; t(9)=1.22, p=0.25), but activation maxima inside areas was typically higher in the left hemisphere (*left:*M=3.93, SD=0.37; *right*: M=3.6, SD=0.26; t(9)=3.96, p<0.01).

In the preceding whole brain analyses we showed evidence for internal error correction (i.e., taboo effect) within a portion of the pMTG that our hypotheses targeted. We next carried out a more specific test for the taboo effect across all voxels of anatomically defined pMTG. Mean parameter estimates for the taboo and word conditions were extracted from pMTG voxels for each participant and a paired t-test between conditions was performed over participants (Figure 25A). The taboo condition elicited higher parameter estimates than the word condition in the left pMTG (*taboo:* M=18.53; SD=36.77; *word*: M=9.59., SD=41.84; t(33)=3.12, Bonferroni corrected p<0.05). The same effect was on the cusp of significance in the right hemisphere (*taboo:* M=2.84; SD=25.66; *word*: M=-4.23, SD=32.61; t(33)=2.58, Bonferroni corrected p=0.05). In addition, we ensured that the left hemisphere internal error correction effect was present independently during imagined (*taboo:* M=10.24; SD=19.49; *word*: M=4.75, SD=24.53; t(33)=2.05, Bonferroni corrected p<0.05) and silently articulated tongue twister trials (*taboo:* M=2.84; SD=25.66; *word*: M=-4.23, SD=32.61; t(33)=2.99, Bonferroni corrected p<0.05).

Finally, we characterized the consistency of the taboo effect within anatomical pMTG by applying paired t-tests between the taboo and word conditions within participants (Figure 3B). This analysis yielded a significant taboo effect in roughly 76% of participants (N=26/34; Bonferroni-corrected p < 0.001). The effect occurred most often

bilaterally (N=22/26) and showed no hemispheric preference when it occasionally occurred in a single hemisphere (*left*: N=2/26; *right*: N=2/26). In four participants the effect was present, but response was greater to baseline. In all cases where the effect was not present, we saw a significant reverse effect (N=9). We considered whether portions of pMTG still showed a taboo effect in those participants that did not exhibit a mean effect across the entire pMTG by inspecting significant voxel-level differences between taboo and word conditions. No participant that showed an insignificant mean taboo effect presented with significant differences between taboo and word condition within portions of pMTG.

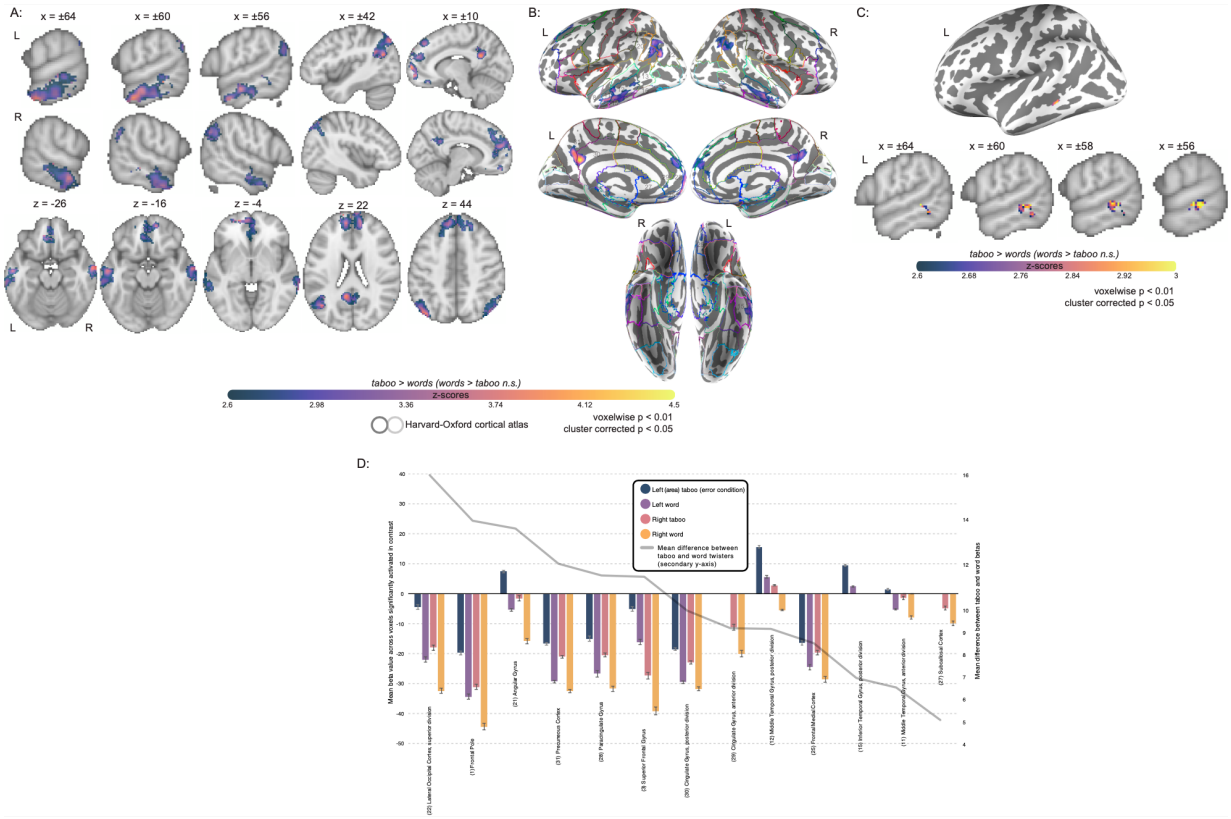

Figure 23: Group contrast between taboo and (neutral) word conditions. Group activation for the taboo condition contrasted with the word condition shows evidence for internal error correction ($p < 0.01$; cluster corr., $p < 0.05$). Z-scores are shown on volume and surface-based brain maps with purple-to-yellow colors representing significantly higher

activation for the taboo condition. No voxels showed significantly higher activation for the word condition. (A) Select slices from the volume-based group activation map. Sagittal slices cut through peak activation observed in anterior middle temporal gyrus (x=±64) and three separate peaks in posterior middle temporal gyrus (x=±60, ±56). They also show the emergence of activity in superior lateral occipital cortex (x=±56) and its splitting into more anterior and posterior foci, with the anterior activity intruding into angular gyrus (x=±42), as well as more medial activations (x=±10; ie., posterior cingulate gyrus and frontal medial cortex). Axial slices show the same patterns: peak of activation in anterior middle temporal gyrus (z=-26), two peaks of activation in posterior middle temporal gyrus (z=-14), a third peak of activation in posterior middle temporal gyrus (z=-4), and two peaks of activation in superior lateral occipital cortex (z=22,44). (B) Volume-based results are projected onto the fsaverage surface along with the Harvard-Oxford cortical atlas to better visualize overlap between activations and anatomical regions. Bilateral anatomical regions are shown as colored outlines and any region overlapping with activation is highlighted by an opaque superimposed number in one hemisphere that corresponds to the region's index within the atlas. The labels for these indices are provided in panel D. (C) The contrast between taboo and word conditions is masked by significant activity during the taboo condition (i.e., taboo > baseline) and 20 voxel cluster-extent threshold is applied, revealing a large area in pMTG. (D) Parameter estimates, or beta values for the individual conditions being contrasted, presented as an average within each anatomical region. Only voxels showing a significant effect in the contrast were included in the averages. Averages are presented separately for left hemisphere (deep violet-blue and strong reddish purple colored bars) and right hemisphere regions (deep purplish pink and yellow colored bars). Error bars represent standard error of the mean. Regions are organized on the x-axis based on descending mean difference between taboo and word parameter estimates. The magnitude of this difference is shown as a gray line that corresponds to the secondary y-axis.

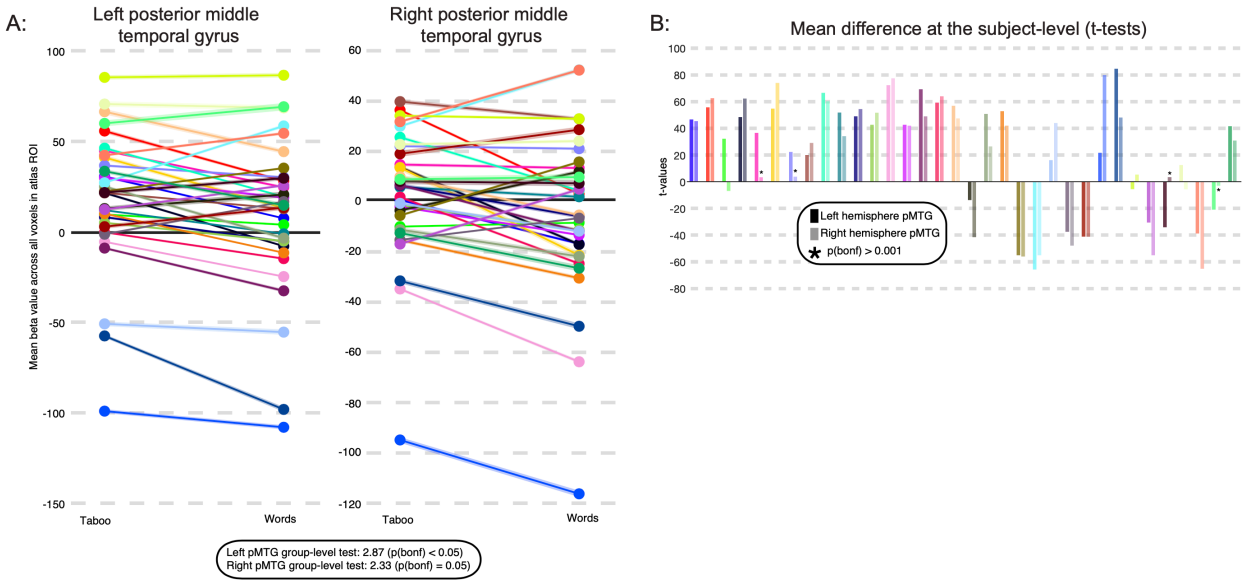

Figure 24: Taboo effect in posterior middle temporal gyrus within-participants. Anatomical ROI-based analysis of the internal error correction effect in posterior middle temporal gyrus. (A) Parameter estimates, or beta values for taboo and word conditions are averaged across all voxels of left posterior middle temporal gyrus (left) and right posterior middle temporal gyrus (right) in each subject. The shaded portions of the line plot for each subject represents standard error of the mean. Results of a paired t-test across participants is presented below, showing significantly higher response in left pMTG to the taboo condition, and a difference between the two conditions that is on the cusp of significance for right pMTG. (B) T-values showing magnitude and significance ($p < 0.001$; Bonferroni corrected) of parameter estimate differences between the two conditions are shown for each subject as a bar plot. Note the asterisk marks insignificant results. Each subject is assigned a color that is consistent between panels (and with Figure 5). Bars are presented in pairs such that the first and darker shaded bar of any pair represents the result of the t-test performed on the left hemisphere ROI, and the second and lighter shaded bar represents the result of the t-test performed over the right hemisphere ROI. Overall, 26/34 participants show a significant effect in at least one hemisphere (22 show an effect in both hemispheres).

Table 2: Clusters from the taboo versus word condition contrast

Anatomical atlas areas overlapping with cluster	Cluster size (2mm voxels)	Cohen's D	Peak z-value	Peak x-coordinate	Peak y-coordinate	Peak z-coordinate	Anatomical region overlapping with peak
Bilateral Frontal Pole; Bilateral Superior Frontal Gyrus; Bilateral Frontal Medial Cortex; Right Subcallosal Cortex; Bilateral Paracingulate; Right Cingulate Gyrus, anterior division	4238	0.52	4.2	-10	64	6	Left Frontal Pole
Left Temporal Pole; Left Superior Temporal Gyrus, posterior division; Left Middle Temporal Gyrus, anterior division; Left Middle Temporal Gyrus, posterior division; Left Middle Temporal Gyrus, temporooccipital part; Left Inferior Temporal Gyrus, posterior division; Left Inferior Temporal Gyrus, temporooccipital part	1258	0.71	3.99	-62	-8	-26	Left Middle Temporal Gyrus, anterior division
Left Supramarginal Gyrus, posterior division; Left Angular Gyrus; Left Lateral Occipital Cortex, superior division	1150	0.79	4.41	-34	-70	58	Left Lateral Occipital Cortex, superior division
Right Temporal Pole; Right Superior Temporal Gyrus, posterior division; Right Middle Temporal Gyrus, anterior division; Right Middle Temporal Gyrus, posterior division; Right Middle Temporal Gyrus, temporooccipital part	1086	1.18	3.88	72	-30	-6	Right Middle Temporal Gyrus, posterior division
Bilateral Cingulate Gyrus, posterior division; Bilateral Precuneous Cortex	1051	1.25	4.36	-8	-52	24	Left Cingulate Gyrus, posterior division
Right Angular Gyrus; Right Lateral Occipital Cortex, superior division	809	0.75	3.91	46	-70	42	Right Lateral Occipital Cortex, superior division

Table 3: Clusters from the taboo versus word conditions contrast masked by taboo > baseline contrast

Anatomical atlas areas overlapping with cluster (% of anatomical area covered)	Cluster size (2mm voxels)	Cohen's D	Peak z-value	Peak x-coordinate	Peak y-coordinate	Peak z-coordinate	Anatomical region overlapping with peak
Left Middle Temporal Gyrus, posterior division (7%); Left Superior Temporal Gyrus (2%); Left Middle Temporal Gyrus, temporooccipital part (2%); Left Inferior Temporal Gyrus, temporooccipital part (<1%)	140	1.78	3.88	-60	-38	-8	Left Middle Temporal Gyrus, posterior division
Right Angular Gyrus (>1%); Right Lateral Occipital Cortex, superior division (>1%)	19	2.13	3.06	52	-56	46	Right Angular Gyrus
Left Lateral Occipital Cortex, superior division (>1%); Left Angular Gyrus (>1%)	15	3.56	3.67	-46	-60	46	Left Lateral Occipital Cortex, superior division
Left Lateral Occipital Cortex, superior division (>1%)	10	0.81	3.5	-34	-70	56	Left Lateral Occipital Cortex, superior division
Left Middle Temporal Gyrus, posterior division (>1%); Left Inferior Temporal Gyrus, posterior division (>1%)	4	3.9	2.84	-54	-30	-17	Left Middle Temporal Gyrus, posterior division

Table 4: Areal effect sizes for taboo versus word condition contrast

	Left hemisphere							Right hemisphere						
Atlas ROI	ROI size	Cohen's D	Peak z-value	Peak x	Peak y	Peak z	% ROI active	ROI size	Cohen's D	Peak z-value	Peak x	Peak y	Peak z	% ROI active
(1) Frontal Pole	6997	0.6	4.2	-12	66	8	14	8195	0.53	3.98	12	68	22	9
(3) Superior Frontal Gyrus	3347	0.61	3.68	-10	54	26	14	3005	0.61	3.6	10	54	20	9
(11) Middle Temporal Gyrus, anterior division	512	1.62	3.99	-64	-6	-24	45	472	0.79	3.45	58	-2	-26	46
(12) Middle Temporal Gyrus, posterior division	1400	0.83	3.94	-60	-26	-8	49	1374	1.49	3.88	70	-28	-4	51
(15) Inferior Temporal Gyrus, posterior division	1175	2.38	3.3	-60	-32	-16	14	1056	-	-	-	-	-	2
(21) Angular Gyrus	1197	1.53	4.25	-50	-56	42	28	1658	1.01	3.69	48	-54	34	18
(22) Lateral Occipital Cortex, superior division	5251	0.78	4.41	-36	-68	60	15	5127	0.75	3.91	44	-68	44	10
(25) Frontal Medial Cortex	561	0.6	3.44	-6	54	-10	51	594	0.63	3.36	-2	54	-20	41
(27) Subcallosal Cortex	782	-	-	-	-	-	5	749	0.7	3.15	2	26	-20	14
(28) Paracingulate Gyrus	1720	0.71	3.73	-10	54	24	18	1650	0.71	3.59	10	52	20	26
(29) Cingulate Gyrus, anterior division	1385	-	-	-	-	-	4	1526	0.79	3.26	-2	42	-4	8
(30) Cingulate Gyrus, posterior division	1332	1.6	4.36	-10	-50	26	18	1393	1.12	3.55	-2	-52	32	14
(31) Precuneous Cortex	3047	1.62	4	-6	-54	32	11	3141	1.53	3.74	0	-60	38	9

5.3.1.2 Predictive coding effect

One aim of the current study was to replicate the predictive coding effect described in prior work (Okada et al., 2018) by identifying brain regions that respond more strongly during silently articulated than imagined tongue twisters. To that end, we confirmed that silently articulated speech activates portions of auditory cortex more strongly than imagined speech, even though both conditions lack auditory input and do not involve overt production ($p < 0.01$ cluster corrected at $p < 0.05$; see Figure 26A and 26B). Overall, silently articulated tongue twisters produced greater activity than imagined tongue twisters in and around Heschl's gyrus (HG), but also in a broad network of speech-related regions that spanned superior temporal gyrus (STG), planum temporale (PT), pMTG, temporooccipital middle temporal gyrus (toMTG), precentral gyrus (preCG), postcentral gyrus (postCG), insula, inferior frontal gyrus (IFG), aCG, and the cerebellum (Figure 26A and 26B, Table 5). We also found that silently articulated tongue twisters yielded greater activity in other brain regions, most of which have also been observed to activate during speech processing (Figure 26A and B, Table 5). These regions were found in temporal cortex (i.e., bilateral planum polare, bilateral temporal pole; PP, TP), inferior temporal and neighboring portions of occipital cortex (i.e., left temporooccipital inferior temporal gyrus, bilateral lingual gyrus, bilateral temporal occipital fusiform cortex, bilateral occipital fusiform gyrus; toITG, LG, toFC, oFG), parietal cortex (i.e., bilateral parietal operculum, bilateral anterior and posterior supramarginal gyrus, bilateral superior parietal lobule; PO, aSMG, pSMG, SPL), and frontal cortex (i.e., bilateral supplementary motor cortex, bilateral middle frontal gyrus, bilateral frontal and central operculum, bilateral superior frontal gyrus, left frontal orbital cortex; SMC, MFG, FO, CO, SFG; Figure 26A and 26B, Table 5).

One point of difference between our results here and prior work[13] is that we also report regions that express greater activity for imagined than silently articulated word lists in several areas including bilateral anterior parahippocampal gyrus (aParaHG) and posterior parahippocampal gyrus (pParaHG), bilateral posterior temporal fusiform cortex (pTFC), bilateral AG, bilateral sLOC, bilateral FP, bilateral FMC, bilateral SC, bilateral paracingulate gyrus (PaCG), right PreC, and right occipital pole (OccP; Figure 26A and 26B, Table 5).

Just as for the taboo effect, the contrast between silently articulated and imagined tongue twisters was masked by areas that showed significant above-baseline response to each of these two conditions, revealing below-baseline response almost exclusively in areas that produced significantly higher response to imagined than silently articulated tongue twisters (Figure 26C). Indeed, all areas that showed greater response to imagined tongue twisters also showed below-baseline response during the imagined condition, while only a few small areas in inferior temporal and occipital cortex that showed greater response to silently articulated tongue twisters exhibited below-baseline response during the silent articulation condition (c.f., Table 5 with Table 6; Figure 26B with 26C). The latter areas included bilateral LG, bilateral oFG, and right OccP (c.f., Table 5 with Table 6; Figure 26B with 26C).

Areal parameter estimates for silently articulated and imagined tongue twister conditions confirmed that areas with stronger response to the silently articulated condition all showed above-baseline response (Figure 26D). However, this analysis also revealed areas with stronger response to the imagined condition that showed above-baseline response, mainly left AG, left FP, and left aParaHG (Figure 26D). Notably, parameter

estimates for these three areas were relatively low for the imagined condition. The contrast between silently articulated and imagined conditions that was masked by above-baseline imagined response did not pick out voxels in these areas due to the more stringent statistical significance criteria in that analysis.

Areal effect sizes for the contrast between silently articulated and imagined conditions were large for left aParaHG, but small for left AG, and negligible but significant for the left FP (Table 7). For other areas which showed greater response during the imagined than silently articulated condition effect sizes were medium-to-large except in the right FP, bilateral sLOC, right AG, and right OccP, where effect sizes were small. The remaining areas all showed relatively greater response to silently articulated speech. A large effect was confirmed in HG bilaterally (Table 7). The second largest effect size we observed overall was in right HG and the largest was in neighboring right PT. In the left hemisphere, effect sizes were larger in portions of temporal cortex (STG, MTG, ITG, and PT) as well as postCG and CO than in HG (Table 7). Most areas that showed a stronger response for silently articulated word lists exhibited medium-to-large effect sizes. However, we report small effect sizes in bilateral TOFC, left SFG, and bilateral aCG. Areal effect sizes were larger in the right hemisphere (*left:*M=0.97, SD=26.22; *right*: M=1.2, SD=0.75; $t(35)=3.56$, $p<0.01$). However, the contrast map between silently articulated and imagined tongue twisters covered a larger portion of areas in the left hemisphere (*left:*M=43.21%, SD=28.28%; *right*: M=38.46%, SD=26.22%; $t(35)=2.8$, $p<0.01$), and local maxima within areas was not significantly different between hemispheres (*left:*M=5, SD=0.77; *right*: M=4.94, SD=0.81; $t(35)=1.24$, $p=0.22$). In other words, the likelihood of finding activity in a particular area and the maximum activity observed inside that area were comparable

across hemispheres, even though a greater portion of each area activated in the left hemisphere, and the smaller activations in right hemisphere areas tended to show a larger effect.

A test for the predictive coding effect was also carried out across all voxels of HG. Mean parameter estimates for the silently articulated and imagined conditions were extracted from HG voxels for each participant and a paired t-test between conditions was performed over participants (Figure 27A). A predictive coding effect for mean response in HG was present at the group level, with silently articulated tongue twisters producing higher mean parameter estimates across participants than imagined tongue twisters in both the left hemisphere (*articulated:* M=67.4; SD=79.84; *imagined*: M=36.39, SD=58.1; t(33)=3.62, Bonferroni-corrected p<0.01) and the right hemisphere (*articulated:* M=36.6; SD=74.46; *imagined*: M=4.15, SD=59; t(33)=4.52, Bonferroni-corrected p<0.01).
The predictive coding effect based on mean response in HG was present in roughly 76% (N=26/34) of participants (Bonferroni-corrected p < 0.001; Figure 27B). Most of these participants (N=20/26) presented with a bilateral effect, but in one participant the effect was only present in the left hemisphere, and in 5 participants it was only present in the right hemisphere. Four participants that showed the effect also showed a stronger response to baseline than silent articulation. Most participants that showed no predictive coding effect in either hemisphere presented with a significant effect in the reverse direction (i.e., imagined > silently articulated across HG; N=7/8), and a few participants with an established right hemisphere predictive coding effect presented with a reverse effect in the contralateral hemisphere (N=4). Inspecting significant voxel-level differences between speech task conditions within participants revealed that the majority of

participants that failed to show a mean effect in HG bilaterally still had voxels within that area that showed greater response to silently articulated than imagined speech ($p< 0.05$, cluster corrected $p < 0.05$; N=6/8). Notably, however, this activity often peaked or at least covered areas bordering HG, including the PT, the pSTG, the aSTG, and the PP (N=4/6; see participants 2,14, 22,32 in Figure 27C). In at least one additional participant we found evidence of stronger response to silently articulated speech, but in the posterior superior temporal sulcus (N=1/8). By masking participant-level contrast maps (i.e., silently articulated vs imagined) by condition-level activation maps (i.e., silently articulated vs baseline), we also found that half of the participants that presented with a mean effect in HG but showed below baseline response during articulated speech (bilaterally) showed some voxels in HG that responded more strongly to articulated speech and produced above-baseline response (N=2/4; see participants 16 and 23 in Figure 27C).

In general, we found at least one of the two effects that we sought (i.e., internal error correction and predictive coding) in 94% of participants (N=32/34) and both effects in 59% of participants (N=20). No significant relationship was documented between mean framewise displacement and participant-level t-scores for contrasts between conditions (*silently articulated versus imagined speech in left HG*: $r(33) = 0.21$, $p = 0.21$; *silently articulated versus imagined speech in right HG*: $r(33) = 0.02$, $p = 0.93$; *taboo errors versus word errors in left pMTG*: $r(33) = -0.14$, $p = 0.43$; *taboo errors versus word errors in right pMTG*: $r(33) = -0.11$, $p = 0.53$).

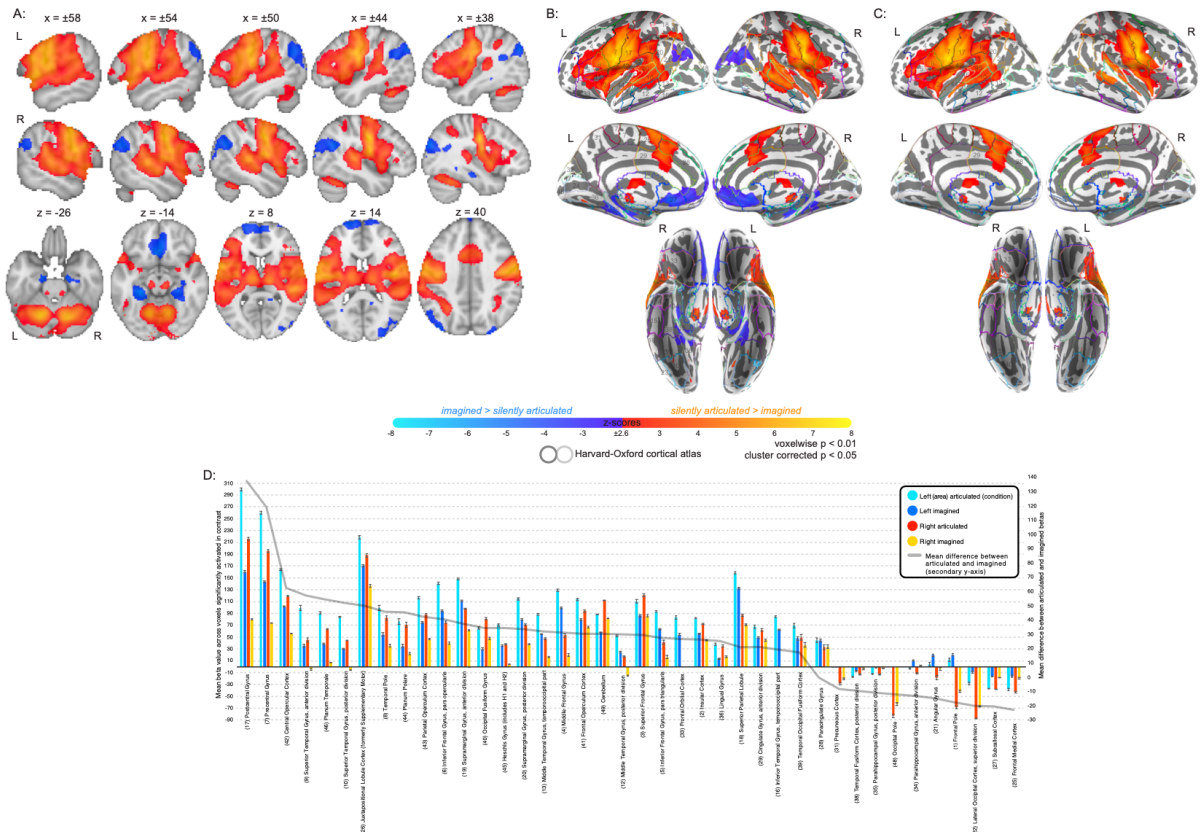

Figure 25: Group contrast between silently articulated and imagined speech. Group activation during silently articulated word lists contrasted with imagined word lists shows evidence for predictive coding ($p < 0.01$; cluster corr., $p < 0.05$). Z-scores are shown on the volume-based and surface-based brain maps with red-to-yellow colors representing significantly higher activation during silently articulated word lists than imagined word lists and blue-to-cyan colors representing significantly higher activation during imagined word lists than silently articulated word lists. (A) Select slices from the volume-based group activation map. Sagittal slices cut through peak activation for the contrast in precentral and postcentral gyri (x=±58) and follow Heschl's gyrus (x=±50, ±44, and ±38). Axial slices show the most ventral activations in the contrast, including the cerebellum and neighboring mesial structures (e.g., anterior parahippocampal gyrus and brainstem; z=-26), more dorsal activations in inferior temporal cortex (z=-14), activations along middle and superior temporal cortex (including Heschl's gyrus; z=8 and 14), and the most posterior activation in the contrast within posterior supramarginal gyrus (z=40). (B) Volume-based results are projected onto the fsaverage surface along with the Harvard-Oxford cortical atlas to better visualize overlap between activations and anatomical regions. Bilateral anatomical regions are shown as colored outlines and any region overlapping with activation is highlighted by an opaque superimposed number in one hemisphere that corresponds to the region's index within the atlas. The labels for these indices are provided in panel D. (C) The contrast between silently articulated and imagined word lists is masked by significant activity during either condition (i.e., silent articulation > baseline OR imagined > baseline), revealing that areas which showed relatively stronger

response to imagined speech were deactivated during the imagined speech task. (D) Parameter estimates, or beta values for the individual conditions being contrasted, presented as an average within each anatomical region. Only voxels showing a significant effect in the contrast were included in the averages. Averages are presented separately for left hemisphere (blue and cyan colored bars) and right hemisphere regions (red and yellow bars). Error bars represent standard error of the mean. Regions are organized on the x-axis based on descending mean difference between articulated and imagined betas. The magnitude of this difference is shown as a gray line that corresponds to the secondary y-axis.

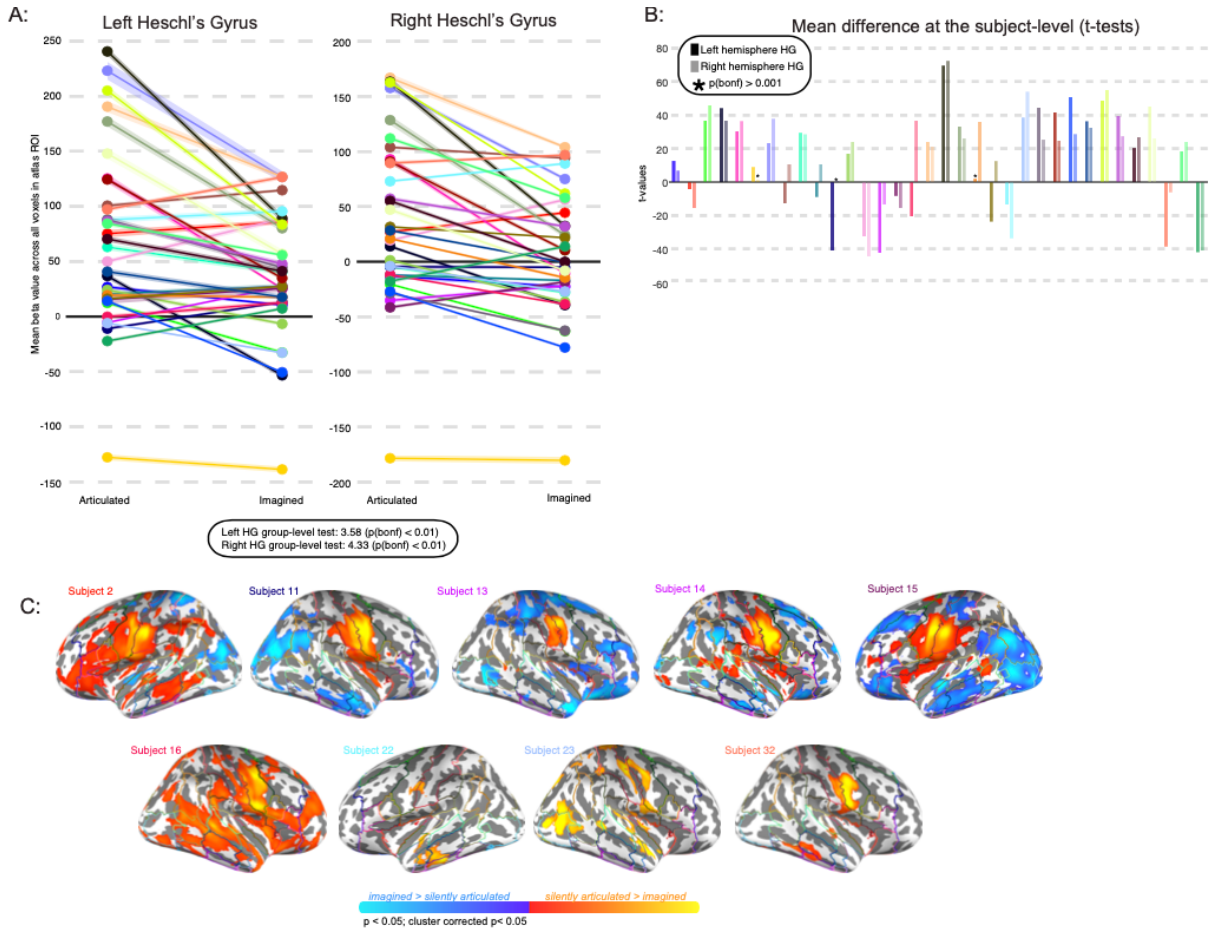

Figure 26: Predictive coding effects in Heschl's gyrus within-participants. Atlas ROI-based analysis of the predictive coding effect in Heschl's gyrus. (A) Beta values for silently articulated word lists and imagined word lists are averaged across all voxels of left Heschl's gyrus (left) and right Heschl's gyrus (right) in each subject. The shaded portions of the line plot for each participant represents standard error of the mean. The bottom portion of the panel also reports the result of a t-test for a group-level difference between conditions (i.e., paired t-test across subjects), which shows significantly greater response during silently articulated speech in Heschl's gyri. (B) T-values showing magnitude and significance (p <

0.001; Bonferroni corrected) of beta differences between the two conditions are shown for each participant as a bar plot. Note the asterisk marks insignificant results. Each participant is assigned a color that is consistent between panels (as well as Figure 3). Bars are presented in pairs such that the first and darker shaded bar of any pair represents the result of the t-test performed on the left hemisphere ROI, and the second and lighter shaded bar represents the result of the t-test performed over the right hemisphere ROI. Overall, 26/35 participants show a significant effect in at least one hemisphere (20 show an effect in both hemispheres). Overall, 26/34 participants show a significant effect in at least one hemisphere. (C) Contrast maps for example participants that show no mean predictive coding effect in Heschl's gyrus. Z-scores are shown on the brain with red-to-yellow colors representing significantly higher activation during silently articulated word lists than imagined word lists and blue-to-cyan colors representing significantly higher activation during imagined word lists than silently articulated word lists ($p < 0.05$ with a cluster significance threshold of $p < 0.05$). Maximum and minimum z-scores vary between participants. Participants 2 ,13,14,15,22, and 32 show higher response for imagined than silently articulated speech across Heschl's gyrus bilaterally. Participant 11 shows higher response to imagined speech in left Heschl's gyrus and no significant difference in the right hemisphere. Note there is still higher response to silently articulated speech within Heschl's gyrus for all of these participants, except participant 13, who still shows higher response to silently articulated speech in the posterior superior temporal sulcus. Participants 16 and 23 all show higher response during silently articulated speech than imagined speech in Heschl's gyrus bilaterally, but also present with response during silently articulated speech that is below baseline fixation. These participants' contrast maps are masked by activity that is above baseline during silent articulation. Each participant shows higher activity for silently articulated than imagined speech in areas bordering Heschl's gyrus.

Table 5: Clusters from the articulated versus imagined speech contrast

Anatomical atlas areas overlapping with cluster	Cluster size (2mm voxels)	Cohen's D	Peak z-value	Peak x-coordinate	Peak y-coordinate	Peak z-coordinate	Anatomical region overlapping with peak
Bilateral Postcentral gyrus; Bilateral Precentral gyrus; Bilateral Central Opercular Cortex; Bilateral Superior Temporal Gyrus, anterior division; Bilateral Planum Temporale; Bilateral Superior Temporal Gyrus, posterior division; Bilateral Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex); Bilateral Temporal Pole; Bilateral Parietal Operculum Cortex; Bilateral Planum Polare; Bilateral Inferior Frontal Gyrus, pars opercularis; Bilateral Supramarginal Gyrus, anterior division; Bilateral Middle Temporal Gyrus, posterior division; Bilateral Heschl's Gyrus (includes H1 and H2); Bilateral Middle Temporal Gyrus, temporooccipital part; Bilateral Supramarginal Gyrus, posterior division; Bilateral Middle Frontal Gyrus; Bilateral Frontal Operculum Cortex; Bilateral Superior Frontal Gyrus; Bilateral Inferior Frontal Gyrus, pars triangularis; Bilateral Insular Cortex; Left Frontal Orbital Cortex; Bilateral Superior Parietal Lobule, Bilateral Cingulate Gyrus, anterior division	41602	0.66	7.01	42	-8	34	Left Precentral Gyrus
Left Inferior Temporal gyrus, temporooccipital part; Bilateral Lingual Gyrus; Bilateral Temporal Occipital Fusiform Cortex; Bilateral Occipital Fusiform Gyrus; Right Occipital Pole; Bilateral Cerebellum	7137	0.93	6.96	-18	-62	-24	Left Cerebellum
Bilateral Frontal Pole; Bilateral Frontal Medial Cortex; Bilateral Superior Frontal Gyrus; Bilateral Paracingulate Gyrus; Bilateral Cingulate Gyrus, anterior portion; Bilateral Subcallosal Cortex	3586	-0.79	-4.53	-2	20	-8	Left Subcallosal Cortex
Right Angular Gyrus, Right Lateral Occipital Cortex, superior division, Right Lateral Occipital Cortex, inferior division, Right Occipital Pole	2191	-0.58	-4.3	24	-92	28	Right Occipital Pole
Right Cingulate Gyrus, posterior division; Right Lingual Gyrus; Right Parahippocampal Gyrus, anterior division; Right Parahippocampal Gyrus, posterior division; Right Temporal Occipital Fusiform Cortex; Right Temporal Fusiform Cortex, posterior division; Right Precuneous	1248	-0.79	-4.25	18	-12	-24	Right Parahippocampal Gyrus, anterior division
Left Lateral Occipital Cortex, superior division; Left Angular Gyrus; Left Lateral Occipital Cortex, inferior division	1113	-0.79	-4.03	-42	-80	34	Left Lateral Occipital Cortex, superior division
Left Parahippocampal Gyrus, posterior division; Left Parahippocampal Gyrus, anterior division; Left Temporal Fusiform Cortex, posterior division; Left Temporal Occipital Fusiform Cortex; Left Lingual Gyrus	1062	-0.67	-4.51	-26	-34	-14	Left Parahippocampal Gyrus, posterior division

Table 6: Clusters from the articulated versus imagined speech contrast masked by articulated > baseline or imagined > baseline contrasts

Anatomical atlas areas overlapping with cluster	Cluster size (2mm voxels)	Cohen's D	Peak z-value	Peak x-coordinate	Peak y-coordinate	Peak z-coordinate	Anatomical region overlapping with peak
Bilateral Postcentral gyrus; Bilateral Precentral gyrus; Bilateral Central Opercular Cortex; Bilateral Superior Temporal Gyrus, anterior division; Bilateral Planum Temporale; Bilateral Superior Temporal Gyrus, posterior division; Bilateral Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex); Bilateral Temporal Pole; Bilateral Parietal Operculum Cortex; Bilateral Planum Polare; Bilateral Inferior Frontal Gyrus, pars opercularis; Bilateral Supramarginal Gyrus, anterior division; Bilateral Middle Temporal Gyrus, posterior division; Bilateral Heschl's Gyrus (includes H1 and H2); Bilateral Middle Temporal Gyrus, temporooccipital part; Bilateral Supramarginal Gyrus, posterior division; Bilateral Middle Frontal Gyrus; Bilateral Frontal Operculum Cortex; Bilateral Superior Frontal Gyrus; Bilateral Inferior Frontal Gyrus, pars triangularis; Bilateral Insular Cortex; Left Frontal Orbital Cortex; Bilateral Superior Parietal Lobule, Bilateral Cingulate Gyrus, anterior division	38279	0.7	7.01	42	-8	34	Left Precentral Gyrus
Left Inferior Temporal gyrus, temporooccipital part; Bilateral Temporal Occipital Fusiform Cortex; Bilateral Cerebellum	6389	1.05	6.96	-18	-62	-24	Left Cerebellum

Table 7: Areal effect sizes for articulated versus imagined word lists

	Left hemisphere							Right hemisphere						
Atlas ROI	ROI size	Cohen's D	Peak z-value	Peak x	Peak y	Peak z	% ROI active	ROI size	Cohen's D	Peak z-value	Peak x	Peak y	Peak z	% ROI active
(1) Frontal Pole	6997	-0.1	4.82	-44	48	20	19	8195	-0.46	4.36	40	36	10	7
(2) Insular Cortex	1463	0.95	5.28	-36	-8	18	68	1465	1.09	5.37	34	-6	18	58
(3) Superior Frontal Gyrus	3347	0.36	5.3	-10	0	76	23	3005	0.79	4.79	0	28	52	13
(4) Middle Frontal Gyrus	3280	0.52	4.97	-48	30	34	27	3141	0.65	5.65	44	10	48	10
(5) Inferior Frontal Gyrus, pars triangularis	836	1.12	5.33	-48	28	12	62	745	0.56	4.4	46	34	8	28
(6) Inferior Frontal Gyrus, pars opercularis	915	0.91	6.02	-62	12	4	96	847	0.68	5.19	56	14	24	94
(7) Precentral Gyrus	4981	1.22	6.59	-62	0	24	50	4822	1.45	7.01	40	-6	36	48
(8) Temporal Pole	2465	0.72	5.56	-56	14	-2	13	2490	0.83	4.68	58	12	2	14
(9) Superior Temporal Gyrus, anterior division	340	1.37	5.17	-66	0	6	47	348	1.47	5.02	64	0	8	47
(10) Superior Temporal Gyrus, posterior division	1019	1.93	4.72	-66	-8	4	94	1155	2.2	5.84	48	-28	8	92
(12) Middle Temporal Gyrus, posterior division	1400	1.39	4.06	-54	-40	-2	16	1374	2.02	4.67	50	-20	-4	25
(13) Middle Temporal Gyrus, temporooccipital part	1019	1.71	3.96	-54	-42	8	29	1287	1.93	5.28	60	-48	12	15
(16) Inferior Temporal Gyrus, temporooccipital part	808	1.35	4.83	-52	-58	-12	19	907	-	-	-	-	-	<1
(17) Postcentral Gyrus	4046	1.55	6.38	-60	-4	26	36	3627	2.22	6.46	52	-6	40	35
(18) Superior Parietal Lobule	1731	0.9	4.72	-44	-38	46	15	1669	0.93	3.58	30	-50	46	7
(19) Supramarginal Gyrus, anterior division	1128	0.77	4.93	-66	-30	26	82	959	1.64	4.85	48	-32	54	45
(20) Supramarginal Gyrus, posterior division	1357	0.8	4.87	-52	-40	24	46	1516	1.04	4.91	58	-42	16	34
(21) Angular Gyrus	1197	-0.35	3.83	-38	-50	44	17	1658	-0.29	4.33	42	-46	52	20
(22) Lateral Occipital Cortex, superior division	5251	-0.24	4.54	-28	-64	54	20	5127	-0.48	-4.24	28	-86	34	27
(25) Frontal Medial Cortex	561	-0.93	-3.72	-2	34	-14	25	594	-1.18	-3.93	2	34	-12	42
(26) Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex)	874	0.79	5.28	-8	2	64	68	834	0.9	4.76	2	6	74	57
(27) Subcallosal Cortex	782	-2.52	-4.53	-4	22	-6	48	749	-2.56	-4.45	-2	22	-4	44
(28) Paracingulate Gyrus	1720	0.01	4.9	-12	14	38	41	1650	-0.01	4.41	10	16	40	38
(29) Cingulate Gyrus, anterior division	1385	0.34	4.95	-12	12	38	42	1526	0.3	4.37	2	18	38	40
(31) Precuneous Cortex	3047	-	-	-	-	-	0	3141	-	-	-	-	-	<1

(33) Frontal Orbital Cortex	1943	0.76	4.42	-50	20	-6	12	1681	-	-	-	-	-	4
(34) Parahippocampal Gyrus, anterior division	709	-1.41	-4.19	-16	-10	-22	12	766	-1.59	-4.25	16	-10	-22	11
(35) Parahippocampal Gyrus, posterior division	475	-0.8	-4.51	-28	-32	-12	46	378	-1.07	-3.72	26	-32	-10	39
(36) Lingual Gyrus	1934	0.67	5.93	-18	-62	-12	20	2037	0.43	5.35	14	-66	-12	26
(38) Temporal Fusiform Cortex, posterior division	1001	-0.69	-3.92	-28	-34	-16	14	812	-0.99	-3.53	34	-20	-18	8
(39) Temporal Occipital Fusiform Cortex	758	0.4	5.67	-22	-62	-14	29	937	0.2	4.79	20	-60	-14	22
(40) Occipital Fusiform Gyrus	1157	1.17	5.46	-20	-66	-14	19	1072	1.16	5.41	14	-70	-14	19
(41) Frontal Operculum Cortex	441	1.03	5.29	-52	14	-2	98	391	0.97	4.58	48	12	2	64
(42) Central Opercular Cortex	1079	1.41	6.36	-50	-8	20	90	990	2.05	6.52	42	-6	20	95
(43) Parietal Operculum Cortex	657	0.95	5.08	-54	-36	28	79	618	2.06	4.7	58	-32	26	46
(44) Planum Polare	444	0.91	5.01	-58	6	2	32	489	1.28	5.06	64	-2	8	25
(45) Heschls Gyrus (includes H1 and H2)	378	1.27	4.28	-56	-12	10	66	331	2.69	5.04	44	-24	10	82
(46) Planum Temporale	626	1.41	5.09	-66	-14	12	100	530	2.96	5.71	50	-26	10	98
(48) Occipital Pole	2759	-	-	-	-	-	2	2372	-0.32	5.12	14	-92	-12	25
(49) Cerebellum	11899	1.03	6.96	-20	-60	-22	22	11999	0.96	6.5	12	-64	-18	23

5.3.2 Exploratory findings

5.3.2.1 Comparing the default mode and taboo networks

The wider network that responded more strongly to taboo than word conditions while sometimes showing below-baseline response (referred to as TN throughout) bears a strong resemblance to the default mode network (DMN). Despite the stronger response to baseline fixation in some portions of this network, the fact that the class of the error word—the main difference between conditions—appears to be driving changes in activity entirely across the DMN is intriguing. Areas of the DMN commonly activate during mentalization and theory of mind tasks (Mars et al., 2011), suggesting that taboo words

may be tapping into social cognitive systems. Here, we sought to describe the functional relationship between areas of the TN and the DMN.

First, we demonstrated that areas of the TN are in fact recruited by the DMN. To define the boundaries of the DMN, we performed a meta-analysis for the phrase "default mode" in Neurosynth (FDR-corrected $p < 0.01$). The resulting meta-analytic map was found to cover canonical areas of the DMN: the anterior medial prefrontal cortex, pCG, AG, MTG, and parahippocampal gyrus (paraHG; Mars et al., 2012). We refer to this map as the DMN throughout the results. The boundaries of the DMN were superimposed over the TN in Figure 28A, qualitatively showing that roughly the same regions activated across the two networks, but that activity was more focal in the TN and often formed a subset of each of the regions that was active in the DMN (Figure 28A). The major exception to this pattern was the paraHG, which appeared in the DMN but not in the TN. Although we saw particularly strong overlap between the two networks in the MTG, the TN contained large portions of the pMTG that did not appear in the DMN. We calculated that 45% of the voxels in the left TN and 36% of the voxels in the right TN overlapped with the DMN (Figure 28B). The finding of greater overlap in the left TN than the right was invariant to network significance thresholds (Figure 28B).

To better understand the relationship between the TN and the DMN, we used Neurosynth to probe for the main functional associations of the TN and then compared the extent to which the same functions were associated with the DMN. That is, we executed the same meta-analysis procedure that we used to define the DMN over the entire set of 3,228 terms frequently used in the neuroimaging literature that are embedded in Neurosynth. We then estimate the Pearson correlation coefficient between each of these meta-analytic

networks and the TN. Networks were not thresholded prior to correlation because our goal was to determine purely whether the likelihood of seeing activity in a group of neuroimaging studies related to the likelihood of observing activity in the TN. However, the resulting correlation coefficients were corrected for multiple comparisons, and we focused only on the top 100 out of 2,952 significant correlations (FDR-corrected $p < 0.001$). Finally, each of the 100 networks most strongly correlated to the TN were in turn correlated to the DMN (Figure 28C). The DMN (i.e., the term default mode) marked the strongest relationship to the TN, explaining roughly 20% of activity in the TN, and validating our early assessment of the areas that form the TN. The set of unique terms that explained the greatest amount of variance in TN activity ($r>0.35$, $r^2 >0.12$) additionally included theory of mind, mentalizing, social [processing], social cognition, autobiographical [memory], self-referential, and beliefs. In this set, associations were higher to the DMN only for the following terms: default mode, self-referential and autobiographical [memory]. For the other terms in the set, association to the TN was on-average modestly higher (mean difference in $r=0.07$, standard error of the mean=0.01), except for the terms social [processing] (DMN $r=0.27$, TN $r=0.42$) and social cognition (DMN $r=0.23$, TN $r=0.36$), which showed a difference that was about twice that of the mean.

A host of other terms were overall significantly associated with the TN as well. Broadly, these covered other memory systems (e.g., episodic, semantic memory), generic memory processes (e.g., retrieval, recall, remembering, recollection), emotional processes and stimuli (e.g., emotions, affective, valence, emotional regulation), references to resting state (e.g., intrinsic, resting), other social cognitive processes (e.g., social interaction, inference, intentions, morals, personality, future), decision-making (e.g., judgments, value),

language (e.g., semantic, sentence, comprehension; note semantic memory does not show a strong difference in association strength to the TN and the DMN) and disorders (e.g., depression, spectrum).

Across all terms that more strongly associated with the TN than the DMN, the language-related category of terms (e.g., semantics and sentence comprehension) showed the largest difference in association strength (i.e., between TN and DMN), followed by emotion processes and social cognition (see the bilateral plots in Figure 28D, which show the terms from Figure 28C organized by association difference). Of the terms with stronger association to the DMN, resting state terms and self-reference showed some of the largest differences with the TN. When we decomposed differences in association to the DMN and TN by hemisphere, we found that the high association difference for language-related terms was being entirely driven by activity in the left hemisphere, whereas the association difference for resting state terms and self-reference was moderately higher in the right hemisphere (Figure 28D).

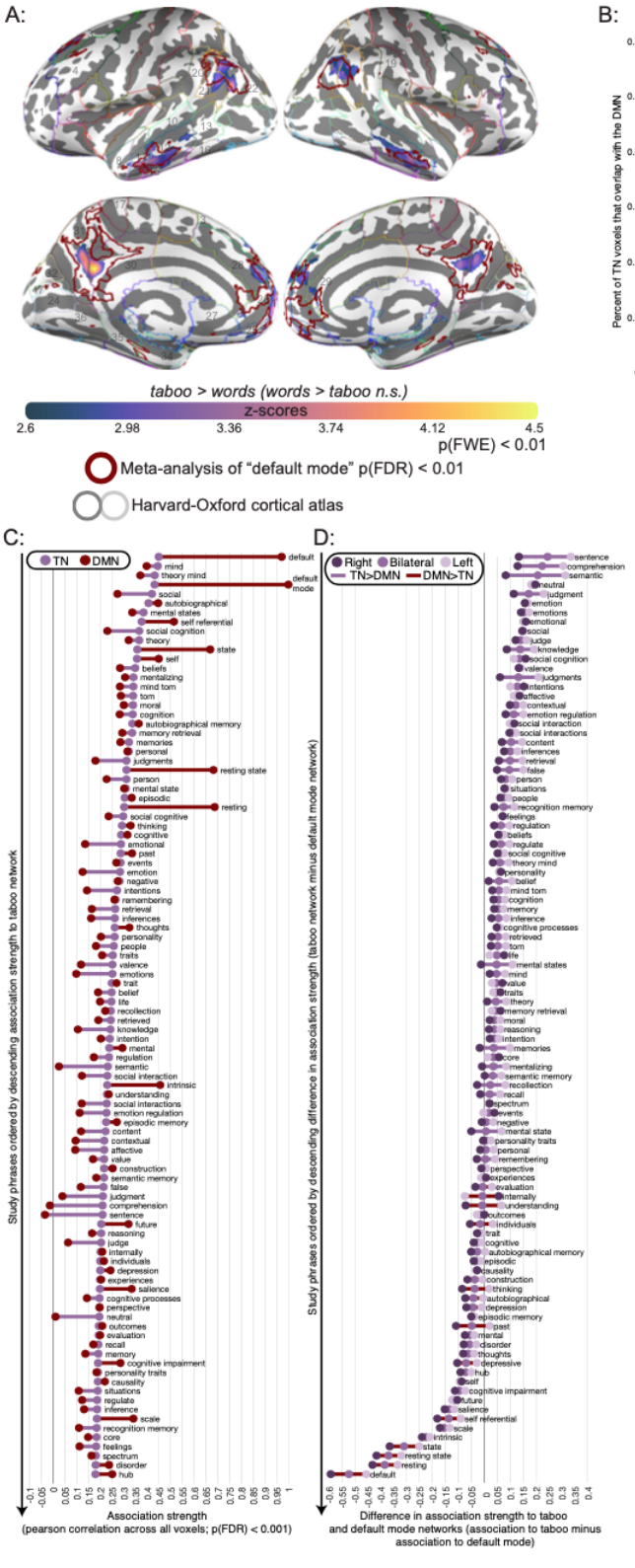
A:
taboo > words (words > taboo n.s.)
z-scores
2.6
2.98
3.36
3.74
4.12
4.5
p(FWE) < 0.01
Meta-analysis of "default mode" p(FDR) < 0.01
Harvard-Oxford cortical atlas
B:
Percent of TN voxels that overlap with the DMN
Left hemisphere
Right hemisphere
Network threshold (z-score)
C:
TN
DMN
Study phrases ordered by descending association strength to taboo network
Association strength
(pearson correlation across all voxels; p(FDR) < 0.001)
D:
Right
Bilateral
Left
TN>DMN
DMN>TN
Study phrases ordered by descending difference in association strength (taboo network minus default mode network)
Difference in association strength to taboo
and default mode networks (association to taboo minus
association to default mode)

Figure 27: Relationship between taboo and default mode networks. Decoding the relationship between taboo and default mode networks. (A) Spatial overlap between cortical areas significantly associated with the taboo network (TN) from Figure 2 and cortical areas significantly associated with the Default Mode Network (DMN). The TN is represented by purple to yellow colors on the cortical surface and the DMN is represented by a maroon outline. The DMN is defined using the Neurosynth meta-analytic database, by analyzing activity associated with studies frequently using the term "default mode". The resulting map represents areas significantly likely to activate across the set of studies frequently using this term (FDR corrected $p < 0.01$). (B) Percentage of TN voxels that overlap with the DMN, plotted as a function of network threshold. Overlap in the left hemisphere is plotted as a peach line and overlap in the right hemisphere is plotted as a teal line. (C) Groups of studies associated with the TN are found by correlating TN z-scores across the whole brain with the likelihood of seeing activity for studies frequently using each of the other terms contained in the Neurosynth database. The top 100 significant correlations (FDR corrected $p < 0.001$; $r > 0.18$) are presented as purple dots in this panel and organized by descending association strength. For each of these terms, the correlation between the likelihood of seeing activity for that term and the likelihood of seeing activity for the "default mode" term (i.e., DMN) is also shown by red dots. A line is drawn between the purple and red dots to emphasize the magnitude of difference between a term and its association to the TN and DMN. The color of the lines shows which of the two networks showed the strongest association to the term. (D) Terms associated with the TN are sorted by descending difference in their association strength to the TN minus their association strength to the DMN. Associations between each term and the two networks and the differences between them are computed separately for the left hemisphere (pale purple colored dots), right hemisphere (deep violet dots), and bilaterally (i.e., matching the data shown in panel C; medium violet dots). Hemisphere-dependent variability is highlighted by connecting each set of dots by a line. The color of the line shows whether the term was overall more strongly associated with the TN (violet line) or DMN (maroon line).

5.3.2.2 Decoding regions of the taboo network

The finding that activity in the left TN more closely resembled language-related terms than activity in the left DMN was remarkable because spatial overlap between the TN and DMN was higher in the left hemisphere. This suggested that what small differences existed in the distribution of activity between the left TN and DMN were functionally relevant for language. We tested whether the largest of such differences, the segment of the left pMTG that does not show overlap with the DMN was driving the taboo network-level association

with language. To do this, we separated the left MTG region in the TN into two segments based on its overlap with the DMN (see top panel of Figure 29B) and then decoded the functions of both segments using Neurosynth. We note that the portion of the pMTG in the TN that did not overlap with the DMN was nearly identical to the largest region from the contrast between taboo and word conditions that also showed above-baseline response. In the interest of better characterizing the functional contributions of each region in the TN, we decoded the other regions of the TN as well, including the unsegmented left MTG (see top panel of Figure 29A). We decoded regions by generating a map for each term in Neurosynth that captures the posterior probability that a study uses a term if activity was observed in a particular voxel, assuming a uniform prior (i.e., all terms are equally likely to appear). Mean probabilities were calculated across voxels of each region in the TN, and we presented all terms that showed at least a 65% mean likelihood of being used when activity was observed in one of the regions in the TN.

Decoding each region of the TN (Figure 29A) broadly showed that different functional categories were relatively more likely to be referenced in the context of different portions of the TN. The majority of language-related terms, including sentence comprehension and lexical-semantics were more likely to be used when there was activity in the left MTG compared to other regions of the TN (e.g., see language network, reading, syntactic, phonological, lexical, word, words, sentence, sentences, concepts, semantic, meaning, aphasia in Figure 29A). Auditory comprehension terms were most likely to be used when the right MTG was active (e.g., see voice, speaker, speech, acoustic, comprehension, listening, linguistic in Figure 29A). The default mode network and social cognition, including theory of mind and self-reference were most likely to be referenced

when activity was in bilateral pCG (e.g., see thoughts, moral, personal, theory of mind, episodic memory, autobiographical memory, self-referential, social interactions, default mode in Figure 29A). Beliefs were most likely to be referenced when either of parietal regions in the TN were active, but the right parietal region was more strongly associated with multiple beliefs. The term mentalization, but curiously also read, readers and written were most likely to be used when the right parietal region was active as well. Word form, retrieval, recollection, reappraisal, semantics, judgment tasks and native English are other terms that were more likely to be used when activity was observed in the left parietal region. No terms were more likely to be used when activity was found in the bilateral frontal cortex region instead of anywhere else in the TN.

Despite the relative differences in term use probabilities between regions of the TN in different functional domains, it was overwhelmingly social cognitive and memory terms that were most likely to be used when activity was found in any region of the TN. For example, the three terms that were most likely to appear when activity was found in each of the regions of the TN (Figure 5A) were largely the same: autobiographical (left MTG, right MTG, left parietal, bilateral pCG, bilateral frontal cortex), autobiographical memory (right MTG, bilateral pCG), theory of mind (right MTG, left parietal, bilateral frontal cortex), mind (left MTG, bilateral frontal cortex), mental states (left MTG), mentalizing (right parietal), beliefs (right parietal), and default network (right parietal, bilateral pCG), and default mode (left parietal). However, in segmenting the MTG region we found that language-related terms were more likely to be used than social cognitive terms when the posterior segment of the left MTG region was active (Figure 29B; e.g., sentences, sentence, language network, language comprehension, comprehension, linguistic, speaker, voice). In

the context of activating the posterior segment of the left MTG, the highest likelihood of using a term from the social cognitive category was 67.6% for the term autobiographical. This likelihood was 2.4% lower than the likelihood of using the term semantic (70%), 3.7% lower than the likelihood of using the term lexical (71.3%), 7.5% lower than using the term comprehension (75.1%), and 10.1% lower than using the term sentences (77.7%). These particular language-related terms were highlighted here on the basis that they aligned with the same terms associated with activity across the entire TN (i.e., sentence comprehension and semantics), and that they additionally implicated the posterior segment of the left MTG region in lexical processes.

Language-related terms were also more likely to be used when activity was observed in the posterior segment of the left MTG than the anterior segment of the left MTG. When the posterior segment of the left MTG was active, the likelihood of using the term semantic was 1.9% higher (posterior left MTG: 70%; anterior left MTG: 68.1%), the likelihood of using the term lexical was 8.5% higher (posterior left MTG: 71.3%; anterior left MTG: 62.8%), the likelihood of using the term comprehension was 13.2% higher (posterior left MTG: 75.1%; anterior left MTG: 61.9%), and the likelihood of using the term sentences was 19% higher (posterior left MTG: 77.7%; anterior left MTG: 58.7%). In general, the likelihood that language-related terms were used when the posterior segment of the MTG was active was strikingly high. The top percentile (>75.4%) of all probabilities that a term was used when activity was found in any region of the TN was comprised of four terms for the pCG (i.e., autobiographical, autobiographical memory, default network) and five terms for the posterior segment of the MTG (linguistic, speaker, sentence, sentences, language network). Meanwhile, social cognitive and memory terms were still

most likely to be used when the anterior segment of the MTG that overlapped with the DMN was active (e.g., see autobiographical, mind, mental states, theory of mind, default mode in Figure 29B). Splitting the left MTG region into anterior and posterior portions typically increased the likelihood that social cognitive terms were used in the anterior portion and decreases the likelihood that they were used in the posterior portion (e.g., see autobiographical, mind, mental states, theory of mind, default mode, social cognitive in Figure 29B).

Finally, the range of language-related terms associated with the posterior segment of the left MTG was much broader than the range of terms associated with the whole left MTG. This included terms related to auditory comprehension that were more strongly associated with the right MTG than the left MTG region, but not more strongly associated with the right MTG than the posterior segment of the left MTG. Other language-related terms that followed this pattern included visual word form, English, semantics, and judgment tasks (previously more strongly associated with the left parietal region than other regions), and read, readers and written (previously more strongly associated with the right parietal region than other regions). In addition, several language-related terms were only likely to be used when activity was found in the posterior segment of the left MTG, including the terms orthographic, visual word, audiovisual and speech perception. Despite the strong association between language-related terms and activity in the posterior segment of the left MTG, we did document a couple of language-related terms that were more likely to be used when activity was found in the anterior segment of the left MTG than the posterior segment. These were concepts (posterior left MTG: 59.6%; anterior left MTG: 67%) and semantic memory (posterior left MTG: 64.4%; anterior left MTG: 66.3%).

Notably, other regions were likely to activate these terms as well. The term semantic memory was about as likely to be used when the bilateral pCG was active (66.2%), and only slightly less likely to be used when the right MTG (62.7%) or the left parietal region of the TN (62.6%) were active. The term concepts also had a relatively moderate likelihood of being used when the left parietal region of the TN was active (61%).

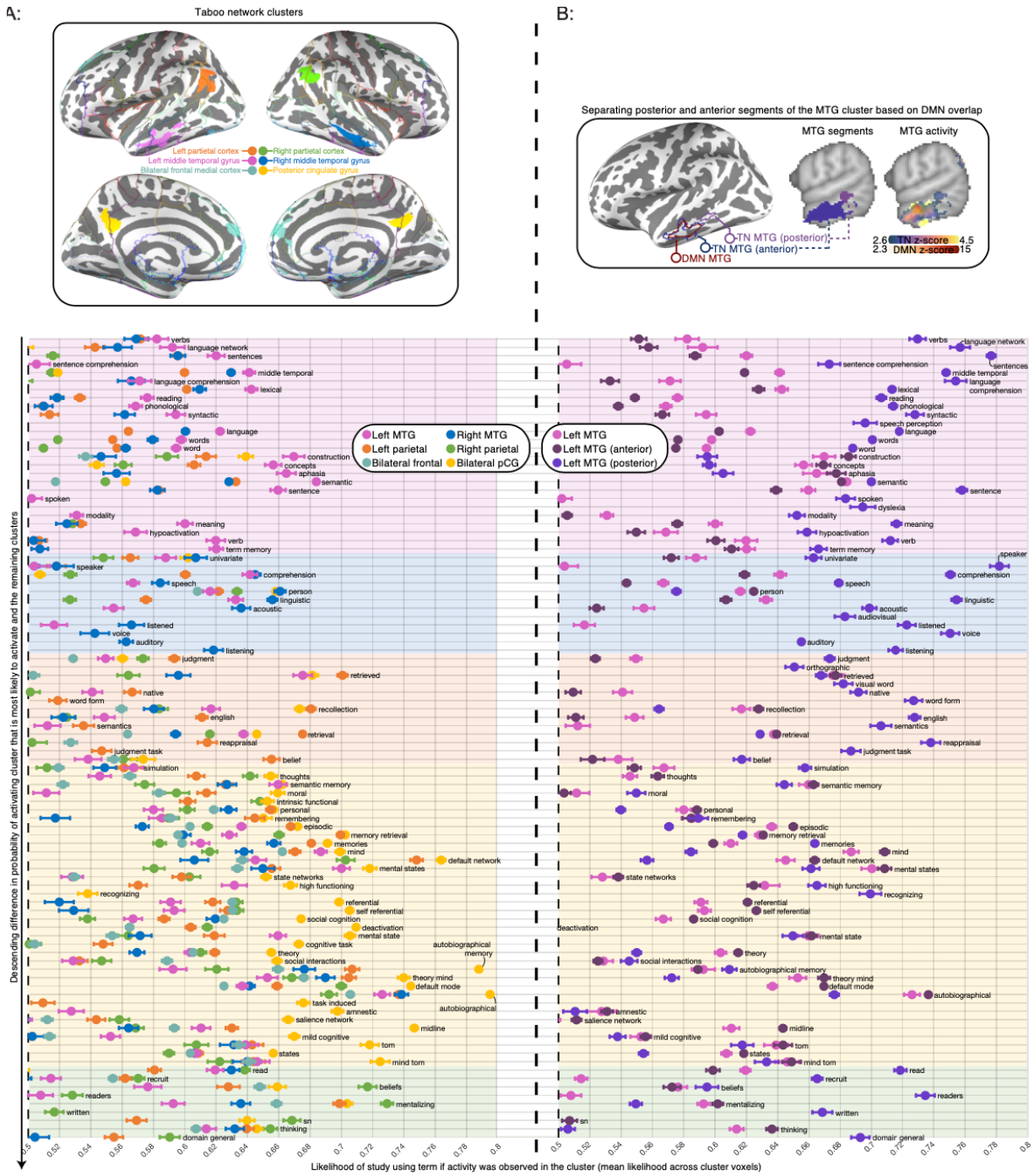

Figure 28: Decoding clusters of the taboo network. (A) The association between clusters in the taboo network (TN) from Figure 2A and groups of different studies in Neurosynth is found by computing the probability of studies in the database using a specific term when activity is reported in each of the voxels belonging to a cluster, then taking the mean of these probabilities across voxels of the cluster. Each cluster in the TN is assigned a unique color as shown in the black box at the top of this panel (clusters are defined in volume space but projected onto the cortical surface to facilitate visualization). The figure below

this box shows mean probabilities for each cluster in a set of terms that has a greater than 65% mean probability of activating voxels in at least one of the six TN clusters from this panel, or two additional TN clusters from panel B. The terms shown on the x-axis are presented in groups based on which cluster was associated with the highest mean probability of using a term. These groupings are represented by the background colors assigned to each area of the chart, which match the color of the cluster exhibiting the highest probability of activating the term. Within each of these groups, terms are further organized by the difference in probability for the most likely cluster associated with a term, and the mean probability of the remaining clusters. Bars around each point represent standard error of the mean across voxels. Clusters with ambiguous mean probabilities are not shown (< 50%). (B) The same decoding process is shown for two segments of the left middle temporal gyrus (MTG) cluster from panel A. The black box above the chart shows how the left MTG cluster from panel A was segmented into a posterior area in left anatomical MTG that has no spatial overlap with the default mode network (DMN), and an anterior area that has some spatial overlap with the DMN. The chart below this black box shows mean probabilities for the same set of terms in panel A, and the terms are organized in the same order as panel A (see panel A for additional details).

5.4 Discussion

5.4.1 Main effect of internal error correction in posterior middle temporal gyrus

The main goal of the present research was to evaluate neural evidence for internal error detection and correction in speech production. To that end, we have leveraged a tongue twister paradigm that has generated suggestive evidence of internal error correction during speech production in the pMTG. That is, prior research (Okada et al., 2018) has shown that tongue twisters designed such that slips *would* result in nonword errors generate greater activity in pMTG than tongue twisters designed such that slips would result in word errors, even though no overt errors occurred. This is intriguing because both conditions involved successfully reciting the same real words, and the only difference

between them was an increased *potential* to produce an internal and nonword speech error. The one caveat to this finding was that the lexicality effect ultimately failed to survive cluster correction. Here, we have tested for the same lexicality effect, using the same stimuli, but in a much larger group of participants. In addition, we have introduced a new class of stimuli designed to put greater demands on internal error correction mechanisms by biasing potential speech errors towards taboo words.

Although we were not able to detect a lexicality effect, we report that taboo tongue twisters successfully increased the load on internal error detection and correction processes—biasing potential speech errors towards taboo words during word list recitation generated significantly stronger response in the pMTG relative to biasing potential speech errors towards (neutral) words. In a whole brain contrast between taboo and word conditions, we demonstrated that portions of pMTG exhibited this taboo effect bilaterally. A more targeted region of interest analysis using anatomically defined pMTG showed a stronger effect in the left hemisphere. The taboo effect held across all voxels of left pMTG but was on the cusp of significance in right pMTG ($p = 0.05$). Within-participant testing for the effect showed that it was highly consistent, presenting in at least one hemisphere in 76% (N=26/34) of participants. In the majority of cases the effect was bilateral (N=22/34), suggesting that a significant group-level effect may be difficult to map in right pMTG due to individual variability in functional response.

We emphasize that the introduction of taboo tongue twisters was designed to induce the potential for errors that may be more salient to the speech system, and here we have shown that this potential is associated with greater response in pMTG, an area implicated in lexical access (Lau et al., 2019; Gow, 2012) and lexical-semantics (Gow, 2012;

Baldo et al., 2013; Devereux et al., 2013). That we were able to successfully map a taboo effect in the same area in which prior work described a lexicality effect trending towards significance suggests that it may still be possible to find a lexicality effect in pMTG using the tongue twister paradigm. Indeed, the initial pilot analysis for the current study implied a large number of participants (N=40) were necessary to achieve a small lexicality effect in middle temporal gyrus (slightly posterior to the area where we report the taboo effect). Given that we approached this number of participants in the analysis here but failed to find an effect, future work in this direction should consider relying on even larger sample sizes, but also leveraging more sensitive analysis methods (e.g., mvpa), and using behavioral data prior to scanning to inform fMRI analysis (e.g., estimating lexicality effect magnitude behaviorally). Similar lexicality effects in pMTG have been previously reported in the lexical decision task (Kotz et al., 2002).

5.4.2 Main effect of predictive coding in auditory cortex

A secondary aim of the current research was to replicate evidence for predictive coding reported by the same work that demonstrated a lexicality effect trending towards significance in the tongue twister paradigm (Okada et al., 2018). We successfully replicated this effect, showing that auditory cortex can generate forward predictive signals during speech production. That is, we found that engaging motor-phonological and lexical-level processes together during silently articulated speech generated significantly greater activity in and around auditory cortex than engaging only lexical-level processes during imagined speech. Critically, the activity we observed in auditory cortex cannot reflect auditory stimulation because no speech output was generated in either of the speech task

conditions. This predictive coding effect was apparent at the group-level based on a whole brain contrast between speech tasks, and a region of interest analysis that showed that the effect held across all voxels of anatomically defined HG in both hemispheres. Within-participant testing revealed that the predictive coding effect was about as consistent as the taboo effect and presented in the same proportion of participants (N=26/34). However, we also found that the majority of participants that did not exhibit a mean predictive coding effect in HG still presented higher brain response to silently articulated speech either in portions of HG or in bordering auditory association areas, showing that there can be substantial individual variability in the precise locus of the effect.

Response around auditory cortex can be explained in the context of a mismatch error between the auditory consequence of a planned articulatory sequence that is expected by the system, and the absence of any such signal as a result of silent articulation. Mismatch error induced by altered auditory feedback has produced a similar pattern of response in auditory cortex (Tourville et al., 2008). Speech induced suppression of auditory response as documented by electrocorticographic, electroencephalographic, and magnetoencephalographic recordings has also been interpreted in the context of a similar kind of mismatch error between a forward prediction of auditory consequences of speech and actual auditory input (Greenlee et al., 2011; Ford et al., 2010; Ventura et al., 2009; Guenther & Hickok, 2015). In models of motor control, mismatch is often extended not just to the content of speech, but its fundamental acoustic properties (e.g., pitch; Guenther & Hickok, 2015). This explains why fMRI studies have demonstrated a similar decrease in auditory response during silently articulated speech when it is coupled with hearing another person produce the same speech stimuli (Agnew et al., 2013).

5.4.3 Networks associated with the taboo effect

Comparison between the taboo and word condition revealed preferential response to the taboo condition in areas outside pMTG and commonly associated with the DMN, including in angular gyrus and adjacent portions of superior lateral occipital cortex, anterior middle temporal gyrus, posterior cingulate gyrus, paracingulate gyrus, superior frontal gyrus, and several medial frontal areas. An additional comparison between the taboo condition and baseline fixation demonstrated that most areas that showed a taboo effect were deactivated during the taboo condition. The strongest evidence for above-baseline response was found in left pMTG, although we note that an areal analysis we performed also implied above-baseline response could be found within left angular gyrus and left posterior inferior temporal gyrus at significance thresholds lower than those used in the present study for direct contrast analyses.

The finding that left pMTG showed a taboo effect and had the clearest above-baseline response during the taboo condition implicated this area in internal error correction above other areas that exhibited a taboo effect. It also implied that internal error correction processes are likely left lateralized. This is further supported by the aforementioned ROI analysis of pMTG that demonstrated a stronger effect in the left hemisphere. The broader network of areas that exhibited a taboo effect also showed evidence for left lateralization—maximum activation within areas was significantly higher in the left hemisphere. Notably, no areas of the brain showed preferential response to the word condition, consistent with our interpretation that the only substantial difference

between the taboo and word conditions is that taboo tongue twisters exercise greater demands on internal error detection and correction processes.

Recent research has also implicated left pMTG in word-picture interference using taboo distractor words (Hansen et al., 2019). This work has also demonstrated increased activity for taboo distractors in a thalamo-cortical network of areas that do not appear in the taboo effect we map using tongue twisters. The functional interpretation of this thalamo-cortical network is consistent with our results here—left pMTG is interpreted to reflect lexical processes whereas other areas of the thalamocortical network (e.g., anterior cingulate gyrus, thalamus, inferior frontal gyrus) support processing the arousing properties of taboo words and deploying attention mechanisms to suppress them. That our results did not reveal areas strongly associated with attention suggests that internal correction of speech errors, including taboo speech errors, may not require intervention from attentional mechanisms.

Although we predicted a robust response in left pMTG for taboo tongue twisters, we did not expect them to generate increased, albeit below-baseline, activity across areas of the DMN. We suggest that the taboo effect tapped into a broader lexical-semantic network and that activity in the DMN was driven by the social features of taboo words. Areas of the DMN appear to show a preference for social concepts embedded in narrative stories (Huth et al., 2016) and respond more strongly to words that connect to interpersonal interactions compared to those that don't (Lin et al., 2018; 2020). Observations like these have been used to argue that the brain areas responsible for theory of mind encode intentionality in semantic space (Binder et al., 2016). In this context, that taboo words generally elicited stronger response across the DMN is less surprising. Taboo words tend to be highly

versatile and convey very different intentions (positive, negative, and otherwise) depending on the context in which they are used (Vingerhoets et al., 2013). For example, work in natural language processing has demonstrated that detection of hate speech is considerably less successful, suffering from high false positive rates, when the pragmatic functions of taboo words aren't explicitly modeled (Davidson et al., 2017).

5.4.4 Networks associated with speech tasks

Previous research (Okada et al., 2018) has demonstrated that the contrast between silently articulated and imagined speech produces activity that is right lateralized in certain regions (e.g., sensorimotor cortex). In the present study we broke this contrast apart into anatomical areas, which showed that although generally the same areas were recruited bilaterally and maximum response within areas did not differ between hemispheres, larger portions of areas in the left hemisphere were sensitive to the difference between articulated and imagined speech, while the smaller portions of areas in the right hemisphere that were sensitive to this difference presented larger effect sizes. Overall, this describes a left hemisphere network that was more sensitive to differences between silently articulated and imagined speech and a right hemisphere network with higher specificity. In fact, the largest two effect sizes we reported were in the right planum temporale and HG, underlining the robustness of auditory response during silently articulated speech. It is worth noting too that HG was one of the few areas that broke away from the trend we described across the contrast—response in HG was substantially higher in the right hemisphere, a much larger portion of the right HG was sensitive to the contrast, and right HG also showed a larger effect size. This pattern echoed our findings at the

participant-level, where a predictive coding effect was typically present in HG for both hemispheres (N=20/26), but when the effect was occasionally unilateral it was most often found in the right hemisphere (N=5). Thus, it is possible that right auditory cortex plays a larger role in processing forward predictive models in speech.

At the group-level, greater response during silently articulated than imagined speech was found in a network that closely followed recent pooled analysis of speech production experiments (Tourville et al., 2019). The core of this network was highly consistent with the prior study we sought to replicate and primarily covered areas connected to phonological and articulatory processes, including the primary sensorimotor and somatosensory cortices (i.e., precentral and postcentral gyri), anterior cingulate gyrus, supplementary motor area, the cerebellum, the inferior frontal gyrus, the planum temporale, the supramarginal gyrus, the insula, and the adjacent central, frontal and parietal operculum (Price, 2012; Hickok, 2012; 2014; Karney & Guenther, 2019; Tourville et al., 2019;). As alluded to earlier, the silent articulation network also contained areas involved in auditory processing, which were HG and the planum polare, but also associative auditory areas in anterior and posterior segments of the superior temporal gyrus (Kearney & Guenther, 2019; Binder et al., 1996). We note that the anatomical superior temporal gyrus we used for reference included portions of the posterior superior temporal sulcus putatively involved in speech perception (Venezia et al., 2017). Stronger activity in these areas is consistent with the behavioral research this study builds on, which has demonstrated that increasing the amount of articulation in speech imagery induces higher rates of phonological speech errors that involve similarly articulated phonemes (Oppenheim & Dell, 2010). The silent articulation network also recruited areas not as

closely related to speech production. For example, some of these areas are commonly associated with executive function (i.e., the intraparietal sulcus, which crosses anatomically defined superior parietal lobule, posterior supramarginal gyrus, and superior lateral occipital cortex; Velenosi et al., 2020), orthographic processing (i.e., temporooccipital portion of inferior temporal gyrus and neighboring fusiform areas; Price, 2012), and lexical and/or semantic processing (i.e., middle temporal gyrus, temporal pole, middle frontal gyrus, superior frontal gyrus, and an area of anatomically defined orbitofrontal cortex that maps onto the pars orbitalis portion of the inferior frontal gyrus in atlases that make the distinction between pars orbitalis and pars triangularis; Desikan et al., 2006) (Price, 2012; Binder et al., 2009). Nevertheless, all of these areas play a role during speech production (Tourville et al., 2019; Price, 2012), and it is possible that the phonological-level processes that distinguished silently articulated speech from imagined speech had the effect of producing greater activity in areas that interface with phonological features.

Curiously, some areas responded more strongly to imagined than silently articulated speech, including parahippocampal gyrus, angular gyrus, superior lateral occipital cortex, temporal fusiform cortex, occipital pole, and several medial frontal areas. These are many of the same areas that exhibited stronger response to taboo tongue twisters relative to word tongue twisters, and just as in the case of the taboo effect, they were all deactivated during the imagined speech condition. However, areal effect analysis hinted that left angular gyrus, frontal pole, and anterior parahippocampal gyrus may contain above-baseline response at lower significance thresholds than those used in the current work for contrast analyses. Although prior research did not observe any areas associated with imagined speech (Okada et al., 2018), a similar network has been

documented in auditory and visual imagery tasks (Tian et al., 2016; Daselaar et al., 2010; Pearson, 2019), and functional connectivity between the angular gyrus and the other areas of this network has been shown to increase during imagined musical performance (instrumental or vocal; Tanaka & Kirino, 2019). The angular gyrus itself has been routinely implicated in semantic, episodic, and autobiographical memory, and response in this area has been related to subjective vividness during episodic memory retrieval and encoding (Tibon et al., 2019). Similarity between areas that show up in the DMN and areas involved in various mental imagery tasks (including auditory imagery) has implied a close relationship between mental imagery and the kinds of internally directed processes that have been connected to DMN areas (Daselaar et al., 2010; Pearson, 2019). Thus, the pattern of activity that we found could simply reflect the fact that imagined speech is relatively more internally oriented than articulated speech, or perhaps that imagined articulation taps into a domain general imagery network.

Given the association between many of the areas in the imagined speech network and memory, another possibility is that imagined speech recruits memory systems to facilitate imaging the auditory consequences of simulated speech under conditions where precise prediction is more difficult. Although we have shown here that forward predictions were stronger when articulatory features were salient in speech imagery (i.e., silently articulated speech), they still appear to be generated in a weaker form when such features are impoverished or absent (i.e., imagined speech; Tian et al., 2016). Further, if the quality of forward predictions produced in this memory-driven process[41] can be substantially different, it could also explain our observation that different portions of auditory cortex can simultaneously show greater activity for silently articulated and imagined speech in a

minority of participants. At least one other explanation is that the areas associated with imagined speech reflect a different process for maintaining speech images in memory. For example, some research has related activity in areas of the imagined speech network to the detail of ongoing thoughts during working memory maintenance (Sormaz et al., 2018), and other studies have implicated the parahippocampal gyrus in the maintenance of novel information during working memory (Schon et al., 2016). Because verbal memory taps into areas involved in speech production (Buchsbaum & D'Esposito, 2019), it may be the case that weaker access to articulatory features in imagined speech requires additional engagement in non-verbal memory systems for maintaining speech imagery. More research is necessary to adjudicate between these possibilities and we emphasize that no areas that generated greater response to imagined than articulated speech showed clear above-baseline response during the imagined speech condition.

5.4.5 Decoding the taboo network

We have suggested that the richness of social features in taboo words—particularly intentionality—has driven activity to areas of the DMN in the taboo contrast. We have also suggested that activity in the left pMTG does not reflect these social features, in part because this area alone shows clear above-baseline response during the recitation of taboo tongue twisters. A functional distinction between pMTG and other areas of the taboo network was further supported by a meta-analysis of studies that frequently mentioned the DMN, which showed strong overlap in most areas, but not pMTG.

Our explanation for the response pattern in the TN predicted that pMTG should be associated with language-specific processes, especially at the word-level, while other areas

of the TN should be associated with DMN functions relevant to processing intentionality. We tested these predictions by decoding the taboo network and its constituent regions in Neurosynth. Decoding the TN involved correlating activity that we mapped as part of the taboo effect (i.e., taboo > words) with the likelihood of finding activity in different groups of neuroimaging studies that frequently mentioned different neuroimaging terms (i.e., meta-analyses for different terms in Neurosynth). This unsurprisingly revealed that DMN studies were most strongly associated with TN activity, followed by studies on functions commonly ascribed to areas of the DMN (e.g., theory of mind, mentalization, self-reference, autobiographical memory, social cognition). However, activity in the TN was also related to comprehension (e.g., comprehension, semantic, sentence). When we correlated the same term-based meta-analyses with the likelihood of observing activity in the DMN (i.e., the earlier DMN meta-analysis), we found that comprehension terms showed by far the largest difference between association strength to the TN and the DMN. Moreover, this difference in association was primarily driven by activity in the left hemisphere, which also overlapped more extensively with the DMN. This pattern suggested that left hemisphere areas of the TN that fall outside the DMN play a more substantial role in comprehension, and therefore that left pMTG may be driving the network-level association between the TN and comprehension.

We directly tested whether the posterior portions of the left middle temporal gyrus (MTG) cluster in the TN was more closely connected to word-level processing by decoding the TN at a regional-level, decoding both the entire MTG cluster as well as the posterior and anterior portions of it that sit outside and inside of the DMN respectively. Decoding was made to be more sensitive to the spatial distribution of activity within smaller regions by

capturing the mean likelihood that a study uses a particular term (that is frequently used in neuroimaging studies) if there is activity in each voxel of a TN cluster. We found that although virtually all language-related terms were more likely to be used when there was activity in the MTG compared to other regions of the TN, seeing activity in any cluster of the TN, including the MTG, was most likely to signal engagement in autobiographical memory, theory of mind, mentalization, or the default mode network. This is exactly what might have been expected based on the results of the network-level decoding. However, separating the left MTG cluster based on overlap with the DMN produced a pattern that supported word processing in the posterior segment of this region. First, we found that activity in the anterior segment of MTG was still most likely to signal DMN-related functions, whereas activity in the posterior segment was most likely to signal engagement in language processes. Further, relative to the whole MTG cluster, the likelihood of using language-related terms generally increased and the likelihood of using DMN-related terms generally decreased when the posterior segment of MTG was active. The reverse of this trend was observed in the anterior segment of MTG. Closer inspection of the terms that were most likely to be used when the posterior segment of the MTG was active revealed an association with comprehension that was broad, covering both sentence and word stimuli. Terms that were likely to be used such as speaker, acoustic, lexical, meaning, comprehension, word, linguistic, and phonological were all consistent with a role for the pMTG in lexical processing. Other terms like semantic, sentences, language comprehension, syntactic, and reading may reflect the activity of lexical processes engaged during semantic tasks. In any case, these patterns clearly showed that one specific portion of the TN—the

left pMTG—was much more strongly associated with language, and at least played a relatively more significant role in lexical processing.

Regional decoding of the TN also indicated that the posterior segment of MTG was involved in semantic processes, but not necessarily semantic memory. Activity in the anterior segment of MTG was more strongly associated with concepts and semantic memory than activity in the posterior segment of MTG. Further, semantic memory was just as strongly associated with the posterior cingulate gyrus and had a remarkably weak association with the posterior segment of MTG (i.e., less than 50% likelihood of term use). The stronger association to both social cognitive terms and semantic memory in areas of the TN that happened to show stronger overlap with the DMN was consistent with our interpretation that taboo words drive response in areas of the DMN as a result of their semantic features, mainly intentionality. This interpretation aligns with a recent proposal for a componential model of semantic representation that suggests areas involved in theory of mind and mentalization are used to encode and understand words that place stronger emphasis on intentionality (Binder et al., 2016). It also aligns with a growing body of work that shows words rich in social semantic features tap into areas that overlap with the DMN (Huth et al., 2016; Lin et al., 2018; 2020; Vingerhoets et al., 2013).

5.4.6 Conclusions and limitations

In summary, the present study has provided evidence that forward predictive signals are generated in auditory cortex during speech production, and that error detection and correction in speech involves left pMTG. We have shown that silently articulated speech produces greater activity than imagined speech within early auditory areas, even though

no auditory input is present during silently articulated speech. We have also shown that biasing potential word errors towards taboo words rather than neutral words during word list recitation generates greater activity in left pMTG, even though no overt speech errors are produced in either condition. Although we found similar response in areas that that demonstrably overlapped with the DMN as well as right pMTG, these areas appeared to respond more robustly to baseline fixation than word list recitation. Further, we have provided evidence from meta-analysis that activity in left pMTG is most likely to reflect word-level processes and by extension word-level error correction. We have hypothesized, with some support from meta-analysis, that the other areas of the taboo network are involved in processing the social semantic features of taboo words, particularly intentionality.

Although we have focused on performing more straightforward meta-analyses in this study, we note that the findings are overall consistent with the functional associations we have shown in the previous study. For example, in the previous study we mapped two clusters that were associated with word processing. The only pattern of activity in this study that overlaps with these clusters is the portion of the pMTG that we have focused on. Further, this portion of the pMTG closely respects the boundary between the more anterior and posterior clusters. That is, it squarely falls into the anterior cluster. This cluster was also shown to have a broader response pattern and was more difficult to interpret functionally than its posterior neighbor. Here, we have shown that this anterior cluster may be involved in internal error correction for speech.

Nevertheless, there are several limitations to the present research. First, overt speech typically increases the amount of motion observed during scanning. Although here

we used mouthed speech, it is difficult to assess the full impact motion has had on the results. In general, we reported what can be considered as relatively low motion in our group of participants[62,63]. We have also tried to mitigate the possibility that motion has had a substantial impact on our results by showing that our primary effects of interest (i.e., the region of interest analyses for predictive coding and internal error correction) are not correlated with mean framewise displacement. One other concern is that overt speech has been shown to produce susceptibility artifacts. It is uncertain the extent to which such artifacts may have impinged on our results, but we note that past studies have found these artifacts to disproportionately affect insular and opercular areas[64], and that the current study focuses on lateral temporal cortex. Finally, our failure to find a lexicality effect in pMTG for tongue twisters suggested that evidence for such an effect trending towards significance in prior research may have reflected a false positive result. However, we suspect this is not the case and that a larger amount of data is necessary to achieve significance for such a small effect size, especially considering that our manipulation of tongue twisters to elicit a greater potential for taboo word errors did successfully produce stronger activity in pMTG.

Chapter 6

Conclusions

This dissertation has focused on extracting more detailed information about functional organization within speech networks by analyzing connectivity data and aggregated published functional neuroimaging work. The basic principle that cortical areas have distinct patterns of connectivity (Passingham et al., 2002) was exploited to parcellate regions involved in speech processing that have eluded precise functional description by mapping the functional and structural connectivity-based networks that they participate in. In tandem, aggregated published functional neuroimaging results were used to map speech networks and characterize the broader behavioral context in which they activate. In some of the work presented here, published functional neuroimaging results were the primary source of data used to investigate questions about how speech networks are configured. However, this data also served to enrich connectivity-based parcellations by providing behavioral context to purely connectivity-defined areas, and to complement more targeted task-fMRI investigations into speech processing by elucidating the functional roles of different areas that respond to the same behavioral intervention. Thus, the thrust of this work has involved using these new tools for mapping functional organization to address open questions about the networks involved speech processing, and their functions.

A strength of the approach for mapping brain areas from the bottom-up using connectivity data is that it can help elucidate functional organization within regions that have proven difficult to parcellate with task-fMRI. In the first study, this strength was

leveraged to parcellate the wider posterior perisylvian (PPS) cortical zone that task-fMRI studies have associated with the planum temporale (PT), a region posterior to auditory cortex that has been classically linked with Wernicke's area, but which has been implicated in a wide range of auditory and speech processes. Despite the diverse architecture of areas contained within this region, some functional-anatomic models developed (partly) on task-fMRI results posit that the PT serves a single broad auditory function or is generally involved in imagery. This contradicts research that has demonstrated that the PT is critical for speech production. Several task-fMRI studies have attempted to parcellate the PT, showing that it may contain portions relevant for speech production that are dissociable from general auditory processing. However, the task-fMRI approach is piecemeal, and no single study can possibly test the full breadth of tasks that have implicated the PT to systematically show which areas are functionally distinguishable. By very carefully mapping different functional networks within the PPS zone using high quality resting state functional connectivity data, the first study of this dissertation produced a map of functionally distinguishable areas within cortex that past task-fMRI work has associated with the PT. Although prior work has produced similar maps for other brain regions, this was the first connectivity-based parcellation that focused explicitly on the PT. Moreover, this work developed and advocated methods for parcellating brain regions by carefully considering the reliability of different approaches for producing parcellations. Given that the use of unsupervised learning methods for clustering functional neuroimaging data is a relatively novel approach that is gaining popularity in the field, there have been recent calls for standardizing the procedures for these methods. This work directly contributes to those efforts by showing that the methods that were advocated for (e.g., consensus

clustering, fusing participant-level parcellations, and fusing different clustering approaches) produced parcellations that were well predicted by external data. Overall, the parcellation of the PPS zone in this work exposed the functional complexity within cortex associated with the PT in prior task-fMRI work, indicating that assigning the PT a single computation is untenable. By registering this parcellation with aggregated functional neuroimaging results from past studies, this work showed that the kinds of functions previously ascribed to the entire PT, could be carried out within some—but not all—of the areas that it contains. For example, separate areas of the PT appeared to be involved in self-generated speech, which may reflect a role in speech imagery, and general auditory processing of complex stimuli. Further, this registration revealed a much more complex functional organization of the PPS zone with respect to speech than anticipated by prior work. Posterodorsal and posteroventral areas of the PPS zone were involved in more complex speech perception functions and either overlapped with the PT directly or were consistent with the extent of cortical fields underlying the PT. Aside from the PPS zone area that appeared to be associated with self-generated speech, multiple other areas within the PPS zone were mapped that may play a role in speech production. The networks that underpinned the PPS zone highlighted the importance of areas in dorsal and ventral premotor cortex for distinguishing speech related areas in the PPS zone, and the importance of motor, auditory and somatosensory regions for distinguishing anterior areas that were more strongly linked to auditory processing. Overall, this work provided a parcellation scheme of a complex cortical zone that is known to be critical for speech. This scheme was shown to be highly generalizable and provides a starting point for more thoroughly investigations into the functions of the speech networks that this zone contains.

The first study demonstrated the utility of connectivity-based parcellations for elucidating organization within speech regions that have been difficult to parcellate with task-fMRI work. However, the weakness of this approach is that it cannot say much about the specific functions that are carried out within areas. In the first study, aggregate functional data served the role of decoding areas mapped on connectivity, but presumably this data is rich enough to parcellate speech regions as well. The second study mapped the networks embedded within the functional neuroimaging literature and used them to localize areas involved in speech perception. This strategy was deliberately distinct from past meta-analytic work attempting to localize speech perception areas. Prior work has relied on conventional meta-analysis, where a group of studies is carefully selected to represent the process being localized. Unfortunately, this work has localized speech perception to different areas, providing support for two different functional-anatomic models that differ with respect to where speech perception occurs. Moreover, this prior work has relied on sample sizes now known to be insufficient and has not mapped the broader range of tasks that can produce activity in the areas putatively involved in perceiving speech. To remedy these issues, a purely data driven meta-analytic approach was taken in this second study, involving the analysis of activity patterns across temporal and inferior parietal cortex that are associated with phrases commonly used in the functional neuroimaging literature. This analysis extracted a set of latent functional networks that cut across many different cognitive processes and tasks. Relying on text-mined data about published studies allowed this work to perform a meta-analysis of speech perception on a vastly larger scale than prior work, with dramatically larger sample sizes. At the same time, the strategy of modeling latent networks provided a means by

which to localize the unique pattern of activity associated with speech perception studies while at the same time characterizing the entire behavioral context in which this pattern of activity occurs. Using this approach, the second study demonstrated that speech perception is associated with multiple unique patterns of activity that reflect the recruitment of gradually more posterior areas in temporal cortex as the complexity of speech processing grows. That is, acoustic analysis of speech sounds begins in auditory cortex while phonetic processing additionally involves the mid-to-posterior ventral superior temporal gyrus. At higher levels of the speech processing hierarchy, additional areas are recruited. Thus, phonological processing involves the posterior superior temporal sulcus and the phonetic processing region, word recognition involves both of those but also posterior middle temporal gyrus, and sentence-level comprehension involves all of these regions in addition to the anterior temporal lobe. Using a purely data driven approach, this second study demonstrated that speech perception occurs along a hierarchy that is directed posteriorly. These results suggest that a synthesis of a large portion of the functional neuroimaging literature largely confirms a specific functional-anatomic model of speech processing. Moreover, the second study went beyond capturing latent functional networks that represent the unique patterns of brain activity embedded in the functional neuroimaging literature. To determine the functional associations of individual regions along the speech processing hierarchy, all of the latent functional networks that were previously mapped were used to define brain areas with similar behavioral associations. Separately, structural brain networks were mapped to identify brain areas with similar connectivity profiles. Remarkably, this analysis confirmed that the relatively noisier and text-mined functional data could delineate approximately the same boundaries as more carefully mapped

structural connectivity data. Moreover, the structural boundaries of the mid-to-posterior ventral superior temporal gyrus implicated in phonetic processing showed the best evidence of fit to the functional data across all areas of the temporal and inferior parietal cortex. This phonetic area was revealed to be distinguished by connectivity to the middle longitudinal fasciculus. Although prior work has theorized that this tract may be associated with a range of different processes based on the putative functions of the brain regions it connects, this study presents more concrete evidence that it is involved in speech perception. Finally, the delineation of speech regions using latent functional networks suggested that most speech regions activated relatively specifically for speech. One exception to this was the posterior middle temporal gyrus, where this analysis indicated that that studies on word recognition and comprehension that observed activity there reflected response within two functionally distinguishable areas, the anterior of which activated under a broader range of conditions.

While the second study implicate the posterior middle temporal gyrus in word recognition, the precise role of this area remained ambiguous and multifaceted. Recent work with task-fMRI has suggested that this region may be involved in error correction for speech. The third study of this dissertation sought neural evidence for internal error detection and correction by leveraging a tongue twister paradigm that induces the *potential* for speech errors while simultaneously excluding any overt errors from analysis. Previous work using the same paradigm in the context of silently articulated and imagined speech production tasks has successfully demonstrated predictive coding in auditory cortex during speech, and in addition has presented suggestive evidence of internal error detection and correction in left posterior middle temporal gyrus (pMTG) on the basis that

this area tended toward showing a stronger response when potential speech errors are biased towards nonwords compared to words. The last study of this dissertation attempted to build on this prior work by replicating the predictive coding and lexicality effects within a cohort of participants nearly twice as large in size (N=40), and by introducing novel tongue twister stimuli designed to further tax internal error correction and detection mechanisms by biasing potential speech errors towards taboo words. Replication of more robust brain response in and around auditory cortex during silently articulated speech was successful. While no evidence was found for a significant difference in brain response as a function of the lexical status of the potential speech error, biasing potential speech errors towards taboo words elicited significantly greater response in left pMTG than biasing potential errors towards (neutral) words. Relatively greater activity for taboo words was also found in a network of areas that resembled the default mode network (DMN), but the pMTG alone responded above baseline during the tongue twister task and fell outside the boundaries of the DMN. In this study, aggregated functional neuroimaging data was used to understand the functional roles of the individual regions that responded to the taboo condition. Decoding the network of brain areas driven by taboo tongue twisters showed that only the portion of pMTG outside the DMN was most likely to activate in language studies, particularly in the context of lexical-semantics but not semantic memory, which was more strongly associated with other areas of this network. These findings suggest that left pMTG is involved in internal error detection and correction during speech and align with a growing body of evidence that areas of the DMN are involved in processing the social-semantic features of words, here indexed by internally generated taboo words.

References

Aarts, E., Roelofs, A., & van Turennout, M. (2009). Attentional control of task and response in lateral and medial frontal cortex: brain activity and reaction time distributions. *Neuropsychologia*, *47*(10), 2089-2099.

Abrams, D. A., Ryali, S., Chen, T., Balaban, E., Levitin, D. J., & Menon, V. (2012). Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. *Cerebral Cortex*, *23*(7), 1703-1714.

Acar, F., Seurinck, R., Eickhoff, S. B., & Moerkerke, B. (2018). Assessing robustness against potential publication bias in activation likelihood estimation (ALE) meta-analyses for fMRI. *PloS one*, *13*(11), e0208177.

Adank, P., Noordzij, M. L., & Hagoort, P. (2012). The role of planum temporale in processing accent variation in spoken language comprehension. *Human brain mapping*, *33*(2), 360-372.

Agnew, Z. K., McGettigan, C., Banks, B., & Scott, S. K. (2013). Articulatory movements modulate auditory responses to speech. *Neuroimage*, *73*, 191-199.

Akiki, T. J., & Abdallah, C. G. (2019). Determining the hierarchical architecture of the human brain using subject-level clustering of functional networks. *Scientific reports*, *9*(1), 1-15.

Alain, C., Reinke, K., McDonald, K. L., Chau, W., Tam, F., Pacurar, A., & Graham, S. (2005). Left thalamo-cortical network implicated in successful speech separation and identification. *Neuroimage*, *26*(2), 592-599.

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., De Haan, E. H., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*, *15*(2), 221-228.

Alexander-Bloch, A. F., Shou, H., Liu, S., Satterthwaite, T. D., Glahn, D. C., Shinohara, R. T., ... & Raznahan, A. (2018). On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage*, *178*, 540-551.

Alvarez, M., Nnoli, J., Carroll Jr, E. J., Hutchins-Carroll, V., Razinia, Z., & Oppenheimer, S. B. (2008). Exogenous hyalin and sea urchin gastrulation, Part II: hyalin, an interspecies cell adhesion molecule. *Zygote (Cambridge, England)*, *16*(1), 73.

Amaro Jr, E., & Barker, G. J. (2006). Study design in fMRI: basic principles. *Brain and cognition*, *60*(3), 220-232.

Amunts, K., & Zilles, K. (2015). Architectonic mapping of the human brain beyond Brodmann. *Neuron*, *88*(6), 1086-1107.

Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M. É., ... & Shah, N. J. (2013). BigBrain: an ultrahigh-resolution 3D human brain model. *Science*, *340*(6139), 1472-1475.

Amunts, K., Mohlberg, H., Bludau, S., & Zilles, K. (2020). Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*, *369*(6506), 988-992.

Amunts, K., Schleicher, A., & Zilles, K. (2007). Cytoarchitecture of the cerebral cortex—more than localization. *Neuroimage*, *37*(4), 1061-1065.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, *412*(2), 319-341.

Anderson, M. L. (2015). Mining the brain for a new taxonomy of the mind. *Philosophy Compass*, *10*(1), 68-77.

Andersson, J., Smith, S., & Jenkinson, M. (2008). FNIRT–FMRIB's non-linear image registration tool. *Human Brain Mapping*, *2008*.

Andreatta, R. D., Stemple, J. C., Joshi, A., & Jiang, Y. (2010). Task-related differences in temporo-parietal cortical activation during human phonatory behaviors. *Neuroscience letters*, *484*(1), 51-55.

Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., & Knösche, T. R. (2006). Connectivity-based parcellation of Broca's area. *Cerebral cortex*, *17*(4), 816-825.

Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., & Parisot, S. (2018). Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage*, *170*, 5-30.

Arthur, D., & Vassilvitskii, S. (2007). k-means++: the advantages of careful seeding, p 1027–1035. In *SODA'07: proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA*.

Aue, T., Lavelle, L. A., & Cacioppo, J. T. (2009). Great expectations: what can fMRI research tell us about psychological phenomena?. *International Journal of Psychophysiology*, *73*(1), 10-16.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of verbal learning and verbal behavior*, *14*(4), 382-391.

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*(6), 907-918.

Bajada, C. J., Jackson, R. L., Haroon, H. A., Azadbakht, H., Parker, G. J., Ralph, M. A. L., & Cloutman, L. L. (2017). A graded tractographic parcellation of the temporal lobe. *Neuroimage*, *155*, 503-512.

Baldo, J. V., Arévalo, A., Patterson, J. P., & Dronkers, N. F. (2013). Grey and white matter correlates of picture naming: evidence from a voxel-based lesion analysis of the Boston Naming Test. *Cortex*, *49*(3), 658-667.

Balsters, J. H., Mantini, D., Apps, M. A., Eickhoff, S. B., & Wenderoth, N. (2016). Connectivity-based parcellation increases network detection sensitivity in resting state fMRI: An investigation into the cingulate cortex in autism. *NeuroImage: Clinical*, *11*, 494-507.

Bandettini, P. A. (2012). Twenty years of functional MRI: the science and the stories. *Neuroimage*, *62*(2), 575-588.

Bandettini, Peter A. *fMRI*. MIT Press, 2020.

Barrett, D. J., & Hall, D. A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. *Neuroimage*, *32*(2), 968-977.

Barton, B., Venezia, J. H., Saberi, K., Hickok, G., & Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proceedings of the National Academy of Sciences*, *109*(50), 20738-20743.

Bastos, A. M., & Schoffelen, J. M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, *9*, 175.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature neuroscience*, *6*(5), 448-450.
Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature neuroscience*, *6*(5), 448-450.
Baumann, S., Petkov, C. I., & Griffiths, T. D. (2013). A unified framework for the organization of the primate auditory cortex. *Frontiers in systems neuroscience*, *7*, 11.
Beckmann, C. F. (2012). Modelling with independent components. *Neuroimage*, *62*(2), 891-901.
Beckmann, C. F., DeLuca, M., Devlin, J. T., & Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1457), 1001-1013.
Beckmann, M., Johansen-Berg, H., & Rushworth, M. F. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *Journal of Neuroscience*, *29*(4), 1175-1190.
Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?. *Neuroimage*, *34*(1), 144-155.
Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard III, M. A., & Greenlee, J. D. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *Neuroimage*, *109*, 418-428.
Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-312.
Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, *15*(6), 1373-1396.
Belyk, M., Brown, R., Beal, D. S., Roebroeck, A., McGettigan, C., Guldner, S., & Kotz, S. A. (2021). Human larynx motor cortices coordinate respiration for vocal-motor control. *Neuroimage*, *239*, 118326.
Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging?. *Annals of the New York Academy of Sciences*, *1191*(1), 133-155.
Bennett, C. M., & Miller, M. B. (2013). fMRI reliability: influences of task and experimental design. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(4), 690-702.
Bermudez, P., & Zatorre, R. J. (2005). Differences in gray matter between musicians and nonmusicians. *Annals of the New York Academy of Sciences*, *1060*(1), 395-399.
Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2008). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral cortex*, *19*(7), 1583-1596.
Bijsterbosch, J., Harrison, S. J., Jbabdi, S., Woolrich, M., Beckmann, C., Smith, S., & Duff, E. P. (2020). Challenges and future directions for representations of functional brain organization. *Nature neuroscience*, *23*(12), 1484-1495.
Binder, J. R. (2015). The Wernicke area: Modern evidence and a reinterpretation. *Neurology*, *85*(24), 2170-2175.
Binder, J. R. (2017). Current controversies on Wernicke's area and its role in language. *Current neurology and neuroscience reports*, *17*(8), 58.
Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive neuropsychology*, *33*(3-4), 130-174.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex*, *19*(12), 2767-2796.
Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral cortex*, *10*(5), 512-528.
Bingham, E., & Hyvärinen, A. (2000). A fast fixed-point algorithm for independent component analysis of complex valued signals. *International journal of neural systems*, *10*(01), 1-8.
Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., ... & Dogonowski, A. M. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, *107*(10), 4734-4739.
Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*, *34*(4), 537-541
Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain*, *125*(8), 1829-1838.
Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.
Boekel, W., Wagenmakers, E. J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely
Bogen, J. E., & Bogen, G. M. (1976). Wernicke's region--Where is it. *Annals of the New York Academy of Sciences*, (280), 834-843.
Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*, *32*(2), 821-841.
Bohland, J. W., Bokil, H., Allen, C. B., & Mitra, P. P. (2009). The brain atlas concordance problem:
Boongoen, T., & Iam-On, N. (2018). Cluster ensembles: A survey of approaches with recent extensions and applications. *Computer Science Review*, *28*, 1-25.
Bottenhorn, K. L., & Laird, A. R. (2021). Data Mining in the Era of Big Data: The BrainMap Database as a
Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., ... & Avesani, P. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 1-7.
Briggs, R. G., Bai, M. Y., Poologaindran, A., Young, I. M., Conner, A. K., ... & Sughrue, M. E. (2020). Parcellation-based modeling of the dorsal premotor area. *Journal of the Neurological Sciences*, *415*, 116907.
Brown, E. N., & Behrmann, M. (2017). Controversy in statistical analysis of functional magnetic resonance imaging data. *Proceedings of the National Academy of Sciences*, *114*(17), E3368-E3369.
Brusco, M. J., & Steinley, D. (2015). Affinity propagation and uncapacitated facility location problems. *Journal of Classification*, *32*(3), 443-480.
Bryce, N. V., Flournoy, J. C., Moreira, J. F. G., Rosen, M. L., Sambook, K. A., Mair, P., & McLaughlin, K. A. (2021). Brain parcellation selection: An overlooked decision point

with meaningful effects on individual differences in resting-state functional connectivity. *NeuroImage*, *243*, 118487.
Bryce, N. V., Flournoy, J. C., Moreira, J. F. G., Rosen, M. L., Sambook, K. A., Mair, P., & McLaughlin, K. A. (2021). Brain parcellation selection: An overlooked decision point with meaningful effects on individual differences in resting-state functional connectivity. *NeuroImage*, *243*, 118487.
Buchsbaum, B. R., & D'Esposito, M. (2008a). The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience*, *20*(5), 762-778.
Buchsbaum, B. R., & D'esposito, M. (2008b). Repetition suppression and reactivation in auditory–verbal short-term recognition memory. *Cerebral Cortex*, *19*(6), 1474-1485.
Buchsbaum, B. R., & D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex*, *112*, 134-148.
Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory–an aggregate analysis of lesion and fMRI data. *Brain and language*, *119*(3), 119-128.
Buchsbaum, B. R., Greer, S., Chang, W. L., & Berman, K. F. (2005). Meta-analysis of neuroimaging
Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*(5), 663-678.
Buchsbaum, B. R., Olsen, R. K., Koch, P. F., Kohn, P., Kippenhan, J. S., & Berman, K. F. (2005). Reading, hearing, and the planum temporale. *Neuroimage*, *24*(2), 444-454.
Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, *48*(4), 687-697.
Buckner, R. L., & Yeo, B. T. (2014). Borders, map clusters, and supra-areal organization in visual
Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews*
Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *Neuroimage*, *26*(4), 1119-1127.
Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), 365-376.
Buzsáki, G. (2020). The brain–cognitive behavior problem: a retrospective. *Eneuro*, *7*(4).
Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., ... & Eickhoff, S. B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage*, *81*, 381-392.
Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Human brain mapping*, *13*(1), 43-53.
Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: brain regions involved with perception and covert production. *Neuroimage*, *31*(3), 1327-1342.

Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Current opinion in behavioral sciences*, *21*, 132-137.
Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology*, *45*(2), 293-298.
Carp, J. (2012). On the plurality of (methodological) worlds: estimating the analytic flexibility of FMRI experiments. *Frontiers in neuroscience*, *6*, 149.
Carp, J. (2012). The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage*, *63*(1), 289-300.
Caspers S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., & Zilles, K. (2006). The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. *Neuroimage*, *33*(2), 430-448.
Cauda, F., D'agata, F., Sacco, K., Duca, S., Geminiani, G., & Vercelli, A. (2011). Functional connectivity of the insula in the resting brain. *Neuroimage*, *55*(1), 8-23.
Cha, J., Jo, H. J., Gibson, W. S., & Lee, J. M. (2017). Functional organization of the human posterior cingulate cortex, revealed by multiple connectivity-based parcellation methods. *Human brain mapping*, *38*(6), 2808-2818.
Chang, C. C., & Lin, C. J. (2012). LIBSVM: a library for support vector machines (2001), Software.
Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2012). Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cerebral cortex*, *23*(3), 739-749.
Chao, Y. P., Cho, K. H., Yeh, C. H., Chou, K. H., Chen, J. H., & Lin, C. P. (2009). Probabilistic topography of human corpus callosum using cytoarchitectural parcellation and high angular resolution diffusion imaging tractography. *Human brain mapping*, *30*(10), 3172-3187.
Chase, H. W., Clos, M., Dibble, S., Fox, P., Grace, A. A., Phillips, M. L., & Eickhoff, S. B. (2015). Evidence for an anterior–posterior differentiation in the human hippocampal formation revealed by meta-analytic parcellation of fMRI coordinate maps: Focus on the subiculum. *NeuroImage*, *113*, 44-60.
Chavis, D. A., & Pandya, D. N. (1976). Further observations on corticofrontal connections in the rhesus monkey. *Brain research*, *117*(3), 369-386.
Chen (2020). Pattern Recognition and Machine Learning Toolbox (https://github.com/PRML/PRMLT), GitHub. Retrieved September 4, 2020.
Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 6.
Cieslik, E. C., Zilles, K., Caspers, S., Roski, C., Kellermann, T. S., Jakobs, O., ... & Eickhoff, S. B. (2012). Is there "one" DLPFC in cognitive action control? Evidence for heterogeneity from co-activation-based parcellation. *Cerebral cortex*, *23*(11), 2677-2689.
Clarke, S., & Morosan, P. (2012). Architecture, connectivity, and transmitter receptors of human auditory cortex. In *The human auditory cortex* (pp. 11-38). Springer, New York, NY.
Clos, M., Amunts, K., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Tackling the multifunctional nature of Broca's region meta-analytically: co-activation-based parcellation of area 44. *Neuroimage*, *83*, 174-188.

Cloutman, L. L., & Lambon Ralph, M. A. (2012). Connectivity-based structural and functional parcellation of the human cortex using diffusion imaging and tractography. *Frontiers in neuroanatomy*, *6*, 34.
Co-Planar, T. J. T. P. (1988). stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. *NY: Thieme*.
Cohen, A. L., Fair, D. A., Dosenbach, N. U., Miezin, F. M., Dierker, D., Van Essen, D. C., ... & Petersen, S. E. (2008). Defining functional areas in individual human brains using resting functional connectivity MRI. *Neuroimage*, *41*(1), 45-57.
Cohen, J. R., & D'Esposito, M. (2016). The segregation and integration of distinct brain networks and their relationship to cognition. *Journal of Neuroscience*, *36*(48), 12083-12094.
Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238-251. confirmatory replication study of structural brain-behavior correlations. *Cortex*, *66*, 115-133.
Corbo, D., & Orban, G. A. (2017). Observing others speak or sing activates spt and neighboring parietal cortex. *Journal of Cognitive Neuroscience*, *29*(6), 1002-1021.
Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of experimental psychology: Learning, memory, and cognition*, *37*(1), 162.
Costafreda, S. G. (2009). Pooling FMRI data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in neuroinformatics*, *3*, 33.
Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, *29*(3), 162-173.
Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *42*(6), 1014-1018.
Craddock, R. C., James, G. A., Holtzheimer III, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, *33*(8), 1914-1928.
Crippa, A., Cerliani, L., Nanetti, L., & Roerdink, J. B. (2011). Heuristics for connectivity-based brain parcellation of SMA/pre-SMA through force-directed graph layout. *Neuroimage*, *54*(3), 2176-2184.
Cunningham, W. A., & Koscik, T. R. (2017). Balancing Type I and Type II error concerns in fMRI through compartmentalized analysis. *Cognitive Neuroscience*, *8*(3), 147-149.
Damasio, A. R., & Geschwind, N. (1984). The neural basis of language. *Annual review of neuroscience*, *7*(1), 127-147.
Damasio, H., & Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain*, *103*(2), 337-350.
Daselaar, S. M., Porat, Y., Huijbers, W., & Pennartz, C. M. (2010). Modality-specific and modality-independent components of the human imagery system. *Neuroimage*, *52*(2), 677-685.
Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017, May). Automated hate speech detection and the problem of offensive language. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 11, No. 1).

de Champfleur, N. M., Maldonado, I. L., Moritz-Gasser, S., Machi, P., Le Bars, E., Bonafé, A., & Duffau, H. (2013). Middle longitudinal fasciculus delineation within language pathways: a diffusion tensor imaging study in human. *European journal of radiology*, *82*(1), 151-157.

de Reus, M. A., & Van den Heuvel, M. P. (2013). The parcellation-based connectome: limitations and extensions. *Neuroimage*, *80*, 397-404.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283.

Dell, G. S. A spreading-activation theory of retrieval in sentence production. *Psychological review* **93**, 283-321 (1986).

Dell'Acqua, F., Bodi, I., Slater, D., Catani, M., & Modo, M. (2013). MR diffusion histology and micro-tractography reveal mesoscale features of the human cerebellum. *The Cerebellum*, *12*(6), 923-931.

Deouell, L. Y., Heller, A. S., Malach, R., D'Esposito, M., & Knight, R. T. (2007). Cerebral responses to change in spatial location of unattended sounds. *Neuron*, *55*(6), 985-996.

Derrfuss, J., & Mar, R. A. (2009). Lost in localization: the need for a universal coordinate database. *Neuroimage*, *48*(1), 1-7.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006 Jul 1;31(3):968-80.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**(3):968-80.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... & Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968-980.

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, *33*(48), 18906-18916.

Devlin, J. T., & Poldrack, R. A. (2007). In praise of tedious anatomy. *Neuroimage*, *37*(4), 1033-1041.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, *109*(8), E505-E514.

Dhanjal, N. S., Handunnetthi, L., Patel, M. C., & Wise, R. J. (2008). Perceptual systems controlling speech production. *Journal of Neuroscience*, *28*(40), 9969-9975.

Di, X., Gohel, S., Kim, E. H., & Biswal, B. B. (2013). Task vs. rest—different network configurations between the coactivation and the resting-state brain networks. *Frontiers in human neuroscience*, *7*.

Dickson, J., Drury, H., & Van Essen, D. C. (2001). 'The surface management system'(SuMS) database: a surface–based database to aid cortical surface reconstruction, visualization and analysis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1412), 1277-1292.

Dillon, K., & Wang, Y. P. (2020). Resolution-based spectral clustering for brain parcellation using functional MRI. *Journal of Neuroscience Methods*, *335*, 108628.
Dorsaint-Pierre, R., Penhune, V. B., Watkins, K. E., Neelin, P., Lerch, J. P., Bouffard, M., & Zatorre, R. J. (2006). Asymmetries of the planum temporale and Heschl's gyrus: relationship to language lateralization. *Brain*, *129*(5), 1164-1176.
Doucet, G., Naveau, M., Petit, L., Delcroix, N., Zago, L., Crivello, F., ... & Joliot, M. (2011). Brain activity at rest: a multiscale hierarchical functional organization. *Journal of neurophysiology*, *105*(6), 2753-2763.
Dronkers, N. F., Wilkins, D. P., Van Valin Jr, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1-2), 145-177.
Droutman, V., Read, S. J., & Bechara, A. (2015). Revisiting the role of the insula in addiction. *Trends in cognitive sciences*, *19*(7), 414-420.
Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in cognitive sciences*, *20*(6), 425-443.
Durnez, J., Moerkerke, B., & Nichols, T. E. (2014). Post-hoc power estimation for topological inference in fMRI. *Neuroimage*, *84*, 45-64.
Eckert, M. A., Leonard, C. M., Possing, E. T., & Binder, J. R. (2006). Uncoupled leftward asymmetries for planum morphology and functional language processing. *Brain and language*, *98*(1), 102-111.
Economo, C. F., & Koskinas, G. N. (1925). *Die cytoarchitektonik der hirnrinde des erwachsenen menschen*. J. Springer.
Edwards, L. J., Kirilina, E., Mohammadi, S., & Weiskopf, N. (2018). Microstructural imaging of human neocortex in vivo. *Neuroimage*, *182*, 184-206
Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, 54-75.
Eichenbaum, A., Pappas, I., Lurie, D., Cohen, J. R., & D'Esposito, M. (2021). Differential contributions of static and time-varying functional connectivity to human behavior. *Network Neuroscience*, *5*(1), 145-165.
Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *Neuroimage*, *59*(3), 2349-2361.
Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *Neuroimage*, *57*(3), 938-949.
Eickhoff, S. B., Constable, R. T., & Yeo, B. T. (2018). Topographic organization of the cerebral cortex and brain cartography. *Neuroimage*, *170*, 332-347.
Eickhoff, S. B., Jbabdi, S., Caspers, S., Laird, A. R., Fox, P. T., Zilles, K., & Behrens, T. E. (2010). Anatomical and functional connectivity of cytoarchitectonic areas within the human parietal operculum. *Journal of Neuroscience*, *30*(18), 6409-6421.
Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., ... & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *Neuroimage*, *137*, 70-85.
Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., ... & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *Neuroimage*, *137*, 70-85.

Eickhoff, S. B., Thirion, B., Varoquaux, G., & Bzdok, D. (2015). Connectivity-based parcellation: Critique and implications. *Human brain mapping*, *36*(12), 4771-4792.
Eickhoff, S. B., Yeo, B. T., & Genon, S. (2018). Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, *19*(11), 672-686.
Eickhoff, S. B., Yeo, B. T., & Genon, S. (2018). Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, *19*(11), 672-686.
Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, *113*(28), 7900-7905.
Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., ... & Hariri, A. R. (2020). What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. *Psychological Science*, 0956797620916786.
Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*, *49*(10), 2812-2821.
Embick, D., & Poeppel, D. (2015). Towards a computational (ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Language, cognition and neuroscience*, *30*(4), 357-366.
Evans, S. (2017). What has replication ever done for us? Insights from neuroimaging of speech perception. *Frontiers in human neuroscience*, *11*, 41.
Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., ... & Scott, S. K. (2014). The pathways for intelligible speech: multivariate and univariate perspectives. *Cerebral Cortex*, *24*(9), 2350-2361.
Fan, L., Wang, J., Zhang, Y., Han, W., Yu, C., & Jiang, T. (2013). Connectivity-based parcellation of the human temporal pole using diffusion tensor imaging. *Cerebral cortex*, *24*(12), 3365-3378.
Fegen, D., Buchsbaum, B. R., & D'esposito, M. (2015). The effect of rehearsal rate and memory load on verbal working memory. *NeuroImage*, *105*, 120-131.
Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... & Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, *18*(11), 1664-1671.
Ford, J. M., Roach, B. J., & Mathalon, D. H. (2010). Assessing corollary discharge in humans using noninvasive neurophysiological methods. *Nature protocols*, *5*(6), 1160-1168.
Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience*, *8*(9), 700-711.
Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673-9678.
Fox, P. T., & Lancaster, J. L. (2002). Mapping context and content: the BrainMap model. *Nature Reviews Neuroscience*, *3*(4), 319-321.
Frackowiak, R., & Markram, H. (2015). The future of human cerebral cartography: a novel approach. *Phil. Trans. R. Soc. B*, *370*(1668), 20140171.
Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal*, *41*(8), 578-588.

Fred, A. L., & Jain, A. K. (2005). Combining multiple clusterings using evidence accumulation. *IEEE transactions on pattern analysis and machine intelligence*, *27*(6), 835-850.
Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *science*, *315*(5814), 972-976.
Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in cognitive sciences*, *13*(4), 175-181.
Friederici, A. D., Makuuchi, M., & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *Neuroreport*, *20*(6), 563-568.
Friston, K. J., & Price, C. J. (2011). Modules and brain mapping. *Cognitive neuropsychology*, *28*(3-4), 241-250.
Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *Neuroimage*, *4*(2), 97-104.
Frühholz, S., Fehr, T., & Herrmann, M. (2009). Interference control during recognition of facial affect enhances the processing of expression specific properties—An event-related fMRI study. *Brain research*, *1269*, 143-157.
Fullerton, B. C., & Pandya, D. N. (2007). Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *Journal of Comparative Neurology*, *504*(5), 470-498.
Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. *Neuroimage*, *19*(4), 1417-1426.
Galaburda, A. M., Rosen, G. D., & Sherman, G. F. (1990). Individual variability in cortical organization: its relationship to brain laterality and implications to function. *Neuropsychologia*, *28*(6), 529-546.
Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology*, *190*(3), 597-610.
Gallen, C. L., & D'Esposito, M. (2019). Brain modularity: a biomarker of intervention-related plasticity. *Trends in cognitive sciences*, *23*(4), 293-304.
Gallen, C. L., Baniqued, P. L., Chapman, S. B., Aslan, S., Keebler, M., Didehbani, N., & D'Esposito, M. (2016). Modular brain network organization predicts response to cognitive training in older adults. *PloS one*, *11*(12), e0169015.
Gannon, P. J., Holloway, R. L., Broadfield, D. C., & Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science*, *279*(5348), 220-222.
Gao, Y., Schilling, K. G., Stepniewska, I., Plassard, A. J., Choe, A. S., Li, X., ... & Anderson, A. W. (2018). Tests of cortical parcellation based on white matter connectivity using diffusion tensor imaging. *NeuroImage*, *170*, 321-331.
Garcea, F. E., & Mahon, B. Z. (2014). Parcellation of left parietal tool representations by functional connectivity. *Neuropsychologia*, *60*, 131-143.
Garnham, A., Shillcock, R.S., Brown, G.D.A., Mill, A.I.D., & Cutler, A. in *Slips of the tongue and language production* (ed Anne Cutler) 251-263 (Mouton, 1982).
Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, *124*, 876-886.
Genon, S., Reid, A., Langner, R., Amunts, K., & Eickhoff, S. B. (2018). How to characterize the function of a brain region. *Trends in cognitive sciences*, *22*(4), 350-364.

Genon, S., Reid, A., Li, H., Fan, L., Müller, V. I., Cieslik, E. C., ... & Eickhoff, S. B. (2018). The heterogeneity of the left dorsal premotor cortex evidenced by multimodal connectivity-based parcellation and functional characterization. *Neuroimage*, *170*, 400-411.

Genon, S., Wensing, T., Reid, A., Hoffstaedter, F., Caspers, S., Grefkes, C., ... & Eickhoff, S. B. (2017). Searching for behavior relating to grey matter volume in a-priori defined right dorsal premotor regions: lessons learned. *Neuroimage*, *157*, 144-156.

Georgiou-Karistianis, N., Akhlaghi, H., Corben, L. A., Delatycki, M. B., Storey, E., Bradshaw, J. L., & Egan, G. F. (2012). Decreased functional brain activation in Friedreich ataxia using the Simon effect task. *Brain and cognition*, *79*(3), 200-208.

Girard, G., Caminiti, R., Battaglia-Mayer, A., St-Onge, E., Ambrosen, K. S., Eskildsen, S. F., ... & Innocenti, G. M. (2020). On the cortical connectivity in the macaque brain: A comparison of diffusion tractography and histological tracing data. *Neuroimage*, *221*, 117201.

Glasser, M. F., & Van Essen, D. C. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *Journal of Neuroscience*, *31*(32), 11597-11616.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Smith, S. M. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171-178

Goder, A., & Filkov, V. (2008, January). Consensus clustering algorithms: Comparison and refinement. In *2008 Proceedings of the Tenth Workshop on Algorithm Engineering and Experiments (ALENEX)* (pp. 109-117). Society for Industrial and Applied Mathematics.

Goder, A., & Filkov, V. (2008, January). Consensus clustering algorithms: Comparison and refinement. In *2008 Proceedings of the Tenth Workshop on Algorithm Engineering and Experiments (ALENEX)* (pp. 109-117). Society for Industrial and Applied Mathematics.

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral cortex*, *26*(1), 288-303.

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., ... & Hampton, J. M. (2017). Precision functional mapping of individual human brains. *Neuron*, *95*(4), 791-807.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., ... & Yarkoni, T. (2015). NeuroVault. org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in neuroinformatics*, *9*, 8

Goulas, A., Zilles, K., & Hilgetag, C. C. (2018). Cortical gradients and laminar projections in mammals. *Trends in Neurosciences*, *41*(11), 775-788.

Gow Jr, D. W. (2012). The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain and language*, *121*(3), 273-288.

Grandjean, J., D'Ostilio, K., Phillips, C., Balteau, E., Degueldre, C., Luxen, A., ... & Collette, F. (2012). Modulation of brain activity during a Stroop inhibitory task by the kind of cognitive control required. *PloS one*, *7*(7), e41513.

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., ... & Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, *98*(2), 439-452.

Graves, W. W., Grabowski, T. J., Mehta, S., & Gupta, P. (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of Cognitive Neuroscience*, *20*(9), 1698-1710.

Greenlee, J. D., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., ... & Howard III, M. A. (2011). Human auditory cortical activation during self-vocalization. PloS one, 6(3), e14744.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral cortex*, *19*(1), 72-78.

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *European Journal of Neuroscience*, *17*(12), 2735-2740.

Griffis, J. C., Nenert, R., Allendorfer, J. B., & Szaflarski, J. P. (2017). Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke. *NeuroImage: Clinical*, *14*, 552-565.

Griffiths, T. D. (2000). Musical hallucinosis in acquired deafness: phenomenology and brain substrate. *Brain*, *123*(10), 2065-2076.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in neurosciences*, *25*(7), 348-353.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in neurosciences*, *25*(7), 348-353.

Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound?. *Neuroreport*, *10*(18),

Grossman, E. D., Jardine, N. L., & Pyles, J. A. (2010). fMR-adaptation reveals invariant coding of biological motion on human STS. *Frontiers in human neuroscience*, *4*, 15.

Guenther, F. H., & Hickok, G. (2015). Role of the auditory system in speech production. *Handbook of clinical neurology*, *129*, 161-175.

Haak, K. V., Marquand, A. F., & Beckmann, C. F. (2018). Connectopic mapping with resting-state fMRI. *Neuroimage*, *170*, 83-94.

Hackett, T. A. (2015). Anatomic organization of the auditory cortex. *Handbook of clinical neurology*, *129*, 27-53.

Hackett, T. A., & Kaas, J. H. (2004). Auditory Cortex in Primates: Functional Subdivisions and Processing Streams.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *Journal of Comparative Neurology*, *502*(6), 924-952.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral cortex*, *9*(7), 697-704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*(9), 1281-1292.

Hansen, J. Y., Markello, R. D., Vogel, J. W., Seidlitz, J., Bzdok, D., & Misic, B. (2021). Mapping gene transcription and neurocognition across human neocortex. *Nature Human Behaviour*, 1-11. Sheets, J. R.,

Hansen, S. J., McMahon, K. L., & de Zubicaray, G. I. (2019). The neurobiology of taboo language processing: fMRI evidence during spoken word production. *Social cognitive and affective neuroscience*, *14*(3), 271-279.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The annals of Statistics*, *13*(1), 70-84.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage*, *62*(2), 852-855.

Haynes, J. D. (2015). A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron*, *87*(2), 257-270.

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—it's my area: or is it?. *Journal of cognitive neuroscience*, *20*(12), 2125-2136.

Hermundstad, A. M., Bassett, D. S., Brown, K. S., Aminoff, E. M., Clewett, D., Freeman, S., ... & Carlson, J. M. (2013). Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proceedings of the National Academy of Sciences*, *110*(15), 6169-6174.

Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., ... & Backlund, M. G. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome biology*, *8*(5), R76.

Hickok, G. (2009b). The functional neuroanatomy of language. *Physics of life reviews*, *6*(3),121-143.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature reviews neuroscience*, *13*(2), 135-145.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29*(1), 2-20.

Hickok, G. (2017). A cortical circuit for voluntary laryngeal control: Implications for the evolution language. *Psychonomic bulletin & review*, *24*(1), 56-63.

Hickok, G. Computational neuroanatomy of speech production. *Nature Reviews Neuroscience* **13**, 135-145, doi:10.1038/nrn3158 (2012).

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, *4*(4), 131-138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1), 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.

Hickok, G., & Poeppel, D. (2016). Neural basis of speech perception. In *Neurobiology of language* (pp. 299-310). Academic Press.

Hickok, G., & Saberi, K. (2012). Redefining the functional organization of the planum temporale region: space, objects, and sensory–motor integration. In *The human auditory cortex* (pp. 333-350). Springer New York.

Hickok, G., & Saberi, K. (2012). Redefining the functional organization of the planum temporale region: space, objects, and sensory–motor integration. In *The human auditory cortex*(pp. 333-350). Springer, New York, NY.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of cognitive neuroscience*, *15*(5), 673-682.
Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407-422.
Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, *101*(5), 2725-2732.
Hoffstaedter, F., Grefkes, C., Caspers, S., Roski, C., Palomero-Gallagher, N., Laird, A. R., ... & Eickhoff, S. B. (2014). The role of anterior midcingulate cortex in cognitive motor control. *Human brain mapping*, *35*(6), 2741-2753.
Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, *106*(6), 2035-2040.
Hopkins, W. D., Marino, L., Rilling, J. K., & MacGregor, L. A. (1998). Planum temporale asymmetries in great apes as revealed by magnetic resonance imaging (MRI). *NeuroReport*, *9*(12), 2913-2918.
Hori, Y., Schaeffer, D. J., Gilbert, K. M., Hayrynen, L. K., Cléry, J. C., Gati, J. S., ... & Everling, S. (2020). Comparison of resting-state functional connectivity in marmosets with tracer-based cellular connectivity. *Neuroimage*, *204*, 116241.
Howe, W. M., Berry, A. S., Francois, J., Gilmour, G., Carp, J. M., Tricklebank, M., ... & Sarter, M. (2013). Prefrontal cholinergic mechanisms instigating shifts from monitoring for cues to cue-guided performance: converging electrochemical and fMRI evidence from rats and humans. *Journal of Neuroscience*, *33*(20), 8742-8752.
Hua et al., Tract probability maps in stereotaxic spaces: analysis of white matter anatomy and tract-specific quantification. NeuroImage, 39(1):336-347 (2008)
Humphreys, G. F., & Lambon Ralph, M. A. (2015). Fusion and fission of cognitive functions in the human parietal cortex. *Cerebral Cortex*, *25*(10), 3547-3560.
Hung, Y., Gaillard, S. L., Yarmak, P., & Arsalidou, M. (2018). Dissociations of cognitive inhibition, response inhibition, and emotional interference: Voxelwise ALE meta-analyses of fMRI studies. *Human brain mapping*, *39*(10), 4065-4082.
Huntenburg, J. M., Bazin, P. L., Goulas, A., Tardif, C. L., Villringer, A., & Margulies, D. S. (2017). A systematic relationship between functional connectivity and intracortical myelin in the human cerebral cortex. *Cerebral Cortex*, *27*(2), 981-997.
Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunisen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453-458.
Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. MIT press.
Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *Neuroimage*, *84*, 1061-1069.
Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *Neuroimage*, *84*, 1061-1069.
Hyvarinen, A. (1999, May). Fast ICA for noisy data using Gaussian moments. In *1999 IEEE International Symposium on Circuits and Systems (ISCAS)* (Vol. 5, pp. 57-61). IEEE.
Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144.

Inui, K., Urakawa, T., Yamashiro, K., Otsuru, N., Takeshima, Y., Nishihara, M., ... & Kakigi, R. (2010). Echoic memory of a single pure tone indexed by change-related brain activity. *BMC neuroscience*, *11*(1), 135.
Isenberg, A. L., Vaden, K. I., Saberi, K., Muftuler, L. T., & Hickok, G. (2012). Functionally distinct regions for spatial processing and sensory motor integration in the planum temporale. *Human brain mapping*, *33*(10), 2453-2463.
Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., ... & Tiitinen, H. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences*, *101*(17), 6809-6814.
Jacewicz, E., Fox, R. A. & Wei, L. Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America* **128**, 839-850, doi:10.1121/1.3459842 (2010).
Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, *128*(2), 839-850.
Jackson, R. L., Bajada, C. J., Rice, G. E., Cloutman, L. L., & Ralph, M. A. L. (2018). An emergent functional parcellation of the temporal cortex. *NeuroImage*, *170*, 385-399.
Jacobs, B., Schall, M., & Scheibel, A. B. (1993). A quantitative dendritic analysis of Wernicke's area in humans. II. Gender, hemispheric, and environmental factors. *Journal of Comparative Neurology*, *327*(1), 97-111.
Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience*, *23*(29), 9541-9546.
Jakab, A., Molnár, P. P., Bogner, P., Béres, M., & Berényi, E. L. (2012). Connectivity-based parcellation reveals interhemispheric differences in the insula. *Brain topography*, *25*(3), 264-271.
Jäncke, L., & Shah, N. J. (2002a). Does dichotic listening probe temporal lobe functions?. *Neurology*, *58*(5), 736-743.
Jäncke, L., & Shah, N. J. (2004). 'Hearing' syllables by 'seeing' visual stimuli. *European Journal of Neuroscience*, *19*(9), 2603-2608.
Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, *15*(4), 733-746.
Janssen, A., Klein, C., & Slors, M. (2017). What is a cognitive ontology, anyway?.
Jaskowiak, P. A., Campello, R. J., & Costa, I. G. (2014, January). On the selection of appropriate distances for gene expression data clustering. In *BMC bioinformatics* (Vol. 15, No. S2, p. S2). BioMed Central.
Jbabdi, S., & Johansen-Berg, H. (2011). Tractography: where do we go from here?. *Brain connectivity*, *1*(3), 169-183.
Jennings JM, McIntosh AR, Kapur S, Tulving E, Houle S. Cognitive subtractions may not add up: the interaction between semantic processing and response mode. Neuroimage. 1997;5:229–39.
Ji, B., Li, Z., Li, K., Li, L., Langley, J., Shen, H., ... & Hu, X. (2016). Dynamic thalamus parcellation from resting-state fMRI data. *Human brain mapping*, *37*(3), 954-967.
Ji, J. L., Spronk, M., Kulkarni, K., Repovš, G., Anticevic, A., & Cole, M. W. (2019). Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage*, *185*, 35-57.

Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Drobnjak, I., Rushworth, M. F. S., Brady, J. M., ... & Matthews, P. M. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13335-13340.

John, C. R., Watson, D., Russ, D., Goldmann, K., Ehrenstein, M., Pitzalis, C., ... & Barnes, M. (2020). M3C: Monte Carlo reference-based consensus clustering. *Scientific reports*, *10*(1), 1-14.

Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., ... & Tzourio-Mazoyer, N. (2015). AICHA: An atlas of intrinsic connectivity of homotopic areas. *Journal of neuroscience methods*, *254*, 46-59.

Josse, G., Kherif, F., Flandin, G., Seghier, M. L., & Price, C. J. (2009). Predicting language lateralization from gray matter. *Journal of Neuroscience*, *29*(43), 13516-13523.

Kahnt, T., Chang, L. J., Park, S. Q., Heinzle, J., & Haynes, J. D. (2012). Connectivity-based parcellation of the human orbitofrontal cortex. *Journal of Neuroscience*, *32*(18), 6240-6250.

Kalyvas, A., Koutsarnakis, C., Komaitis, S., Karavasilis, E., Christidi, F., Skandalakis, G. P., ... & Stranjalis, G. (2020). Mapping the human middle longitudinal fasciculus through a focused anatomo-imaging study: shifting the paradigm of its segmentation and connectivity pattern. *Brain Structure and Function*, *225*(1), 85-119.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, *9*(6), 718-727.

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, cognition and neuroscience*, *34*(9), 1214-1229.

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage*, *14*(6), 1402-1408.

Keller, S. S., Roberts, N., García-Fiñana, M., Mohammadi, S., Ringelstein, E. B., Knecht, S., & Deppe, M. (2011). Can the language-dominant hemisphere be predicted by brain anatomy?. *Journal of Cognitive Neuroscience*, *23*(8), 2013-2029.

Kelley, T. A., Rees, G., & Lavie, N. (2013). The impact of distractor congruency on stimulus processing in retinotopic visual cortex. *Neuroimage*, *81*, 158-163.

Kelly, C., Toro, R., Di Martino, A., Cox, C. L., Bellec, P., Castellanos, F. X., & Milham, M. P. (2012). A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage*, *61*(4), 1129-1142.

Kemeny, S., Ye, F. Q., Birn, R., & Braun, A. R. (2005). Comparison of continuous overt speech fMRI using BOLD and arterial spin labeling. *Human brain mapping*, *24*(3), 173-183.

Kim, S. Y., & Won Lee, J. (2007). Ensemble clustering method based on the resampling similarity measure for gene expression data. *Statistical methods in medical research*, *16*(6), 539-564.

Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., ... & Hemberg, M. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nature methods*, *14*(5), 483-486.

Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T., & Mathiak, K. (2011). Supramodal representation of emotions. *Journal of Neuroscience*, *31*(38), 13635-13643.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage*, *36*(3), 889-900.

Klein, C. (2010). Philosophical issues in neuroimaging. *Philosophy Compass*, *5*(2), 186-198.

Klein, C. (2012). Cognitive ontology and region-versus network-oriented analyses. *Philosophy of Science*, *79*(5), 952-960.
Klein, J. C., Behrens, T. E., Robson, M. D., Mackay, C. E., Higham, D. J., & Johansen-Berg, H. (2007). Connectivity-based parcellation of human cortex using diffusion MRI: establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *Neuroimage*, *34*(1), 204-211.
Knaus, T. A., Bollich, A. M., Corey, D. M., Lemen, L. C., & Foundas, A. L. (2006). Variability in perisylvian brain anatomy in healthy adults. *Brain and language*, *97*(2), 219-232.
Kotz, S. A., Cappa, S. F., von Cramon, D. Y., & Friederici, A. D. (2002). Modulation of the lexical–semantic network by auditory semantic priming: An event-related functional MRI study. *Neuroimage*, *17*(4), 1761-1772.
Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Sound of silence activates auditory cortex. *Nature*, *434*(7030), 158-158.
Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, *93*(3), 480-490.
Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature neuroscience*, *12*(5), 535.
Krienen, F. M., Yeo, B. T., & Buckner, R. L. (2014). Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1653), 20130526.
Krug, M. K., & Carter, C. S. (2012). Proactive and reactive control during emotional interference and its relationship to trait anxiety. *Brain research*, *1481*, 13-36.
Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G. R., & Von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *European Journal of Neuroscience*, *21*(1), 230-238.
Kubit, B., & Jack, A. I. (2013). Rethinking the role of the rTPJ in attention and social cognition in light of the opposing domains hypothesis: findings from an ALE-based meta-analysis and resting-state functional connectivity. *Frontiers in human neuroscience*, *7*, 323.
Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly*, *2*(1-2), 83-97.
Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vértes, P. E., Inati, S. J., ... & 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.
Kundu, P., Inati, S. J., Evans, J. W., Luh, W. M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage*, *60*(3), 1759-1770.
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.
Kurmukov, A., Mussabaeva, A., Denisova, Y., Moyer, D., Jahanshad, N., Thompson, P. M., & Gutman, B. A. (2020). Optimizing Connectivity-Driven Brain Parcellation Using Ensemble Clustering. *Brain Connectivity*, *10*(4), 183-194.

Lacosse, E., Scheffler, K., Lohmann, G., & Martius, G. (2021). Jumping over baselines with new methods to predict activation maps from resting-state fMRI. *Scientific reports*, *11*(1), 1-15.
Laird, A. R., Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., Turner, J. A., ... & Fox, P. T. (2009). ALE meta-analysis workflows via the brainmap database: progress towards a probabilistic functional brain atlas. *Frontiers in neuroinformatics*, *3*.
Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., ... & Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human brain mapping*, *25*(1), 155-164.
Lancichinetti, A., & Fortunato, S. (2012). Consensus clustering in complex networks. *Scientific reports*, *2*, 336.
Langers, D. R., Backes, W. H., & van Dijk, P. (2003). Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. *Neuroimage*, *20*(1), 265-275.
Lashkari, D., & Golland, P. (2008). Convex clustering with exemplar-based models. In *Advances in neural information processing systems* (pp. 825-832)
Lau, E. F., & Namyst, A. (2019). fMRI evidence that left posterior temporal cortex contributes to N400 effects of predictability independent of congruity. *Brain and language*, *199*, 104697.
Lefèvre, J., Pepe, A., Muscato, J., De Guio, F., Girard, N., Auzias, G., & Germanaud, D. (2018). SPANOL (SPectral ANalysis of Lobes): A Spectral Clustering Framework for Individual and Group Parcellation of Cortical Surfaces in Lobes. *Frontiers in Neuroscience*, *12*, 354.
Lefranc, S., Roca, P., Perrot, M., Poupon, C., Le Bihan, D., Mangin, J. F., & Rivière, D. (2016). Groupwise connectivity-based parcellation of the whole human cortical surface using watershed-driven dimension reduction. *Medical image analysis*, *30*, 11-29.
Leinonen, L., Hyvärinen, J., & Sovijärvi, A. R. A. (1980). Functional properties of neurons in the temporo-parietal association cortex of awake monkey. *Experimental Brain Research*, *39*(2), 203-215.
Leonard, C. M., Puranik, C., Kuldau, J. M., and Lombardino, L. J. 1998. Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: Where is it? *Cereb. Cortex* 8: 397–406.
Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41-104.
Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1-22.
Levelt, W. J. (1993). *Speaking: From intention to articulation*(Vol. 1). MIT press.
Levelt, W. J. M. Accessing Words in Speech Production - Stages, Processes and Representations. *Cognition* **42**, 1-22, doi:Doi 10.1016/0010-0277(92)90038-J (1992).
Levelt, W. J. M. *Speaking: From intention to articulation*. (MIT Press, 1989).
Levelt, W. J. M., Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behavioral & Brain Sciences* **22**, 1-75 (1999).
Levelt, W. J. Monitoring and self-repair in speech. *Cognition* **14**, 41-104 (1983).
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*, 1-38.

Liebenthal, E., Desai, R. H., Humphries, C., Sabri, M., & Desai, A. (2014). The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. *Frontiers in neuroscience*, *8*, 289.

Lieberman, M. D., Burns, S. M., Torre, J. B., & Eisenberger, N. I. (2016). Reply to Wager et al.: Pain and the dACC: The importance of hit rate-adjusted effects and posterior probabilities with fair priors. *Proceedings of the National Academy of Sciences*, *113*(18), E2476-E2479.

Liégeois, R., Santos, A., Matta, V., Van De Ville, D., & Sayed, A. H. (2020). Revisiting correlation-based functional connectivity and its relationship with structural connectivity. *Network Neuroscience*, *4*(4), 1235-1251.

Liljeström, M., Tarkiainen, A., Parviainen, T., Kujala, J., Numminen, J., Hiltunen, J., ... & Salmelin, R. (2008). Perceiving and naming actions and objects. *Neuroimage*, *41*(3), 1132-1141.

Lin, N., Wang, X., Xu, Y., Wang, X., Hua, H., Zhao, Y., & Li, X. (2018). Fine subdivisions of the semantic network supporting social and sensory–motor semantic processing. *Cerebral Cortex*, *28*(8), 2699-2710.

Lin, N., Xu, Y., Yang, H., Zhang, G., Zhang, M., Wang, S., ... & Li, X. (2020). Dissociating the neural correlates of the sociality and plausibility effects in simple conceptual combination. *Brain Structure and Function*, *225*(3), 995-1008.

Linting, M., Van Os, B. J., & Meulman, J. J. (2011). Statistical significance of the contribution of variables to the PCA solution: an alternative permutation strategy. *Psychometrika*, *76*(3), 440-460.

Liu, C., JaJa, J., & Pessoa, L. (2018). LEICA: Laplacian eigenmaps for group ICA decomposition of fMRI data. *NeuroImage*, *169*, 363-373.

Liu, H., Qin, W., Li, W., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Connectivity-based parcellation of the human frontal pole with diffusion tensor imaging. *Journal of Neuroscience*, *33*(16), 6782-6790.

Liu, H., Qin, W., Li, W., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Connectivity-based parcellation of the human frontal pole with diffusion tensor imaging. *Journal of Neuroscience*, *33*(16), 6782-6790.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869-878.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157. doi: 10.1038/35084005

Long, X., Goltz, D., Margulies, D. S., Nierhaus, T., & Villringer, A. (2014). Functional connectivity-based parcellation of the human sensorimotor cortex. *European Journal of Neuroscience*, *39*(8), 1332-1342.

Lorca-Puls, D. L., Gajardo-Vidal, A., White, J., Seghier, M. L., Leff, A. P., Green, D. W., ... & Price, C. J. (2018). The impact of sample size on the reproducibility of voxel-based lesion-deficit mappings. *Neuropsychologia*, *115*, 101-111.

Lucas G. S. Jeub, Marya Bazzi, Inderjit S. Jutla, and Peter J. Mucha, *"A generalized Louvain method for community detection implemented in MATLAB,"*https://github.com/GenLouvain/GenLouvain (2011-2019).

Luo, W., Greene, A. S., & Constable, R. T. (2021). Within node connectivity changes, not simply edge changes, influence graph theory measures in functional connectivity studies of the brain. *NeuroImage*, *240*, 118332.
Luria, S. (1958). Brain disorders and language analysis. Language and Speech 1: 14-34. Reprinted in Aphasia: Selected Readings. M. T. Sarno, Ed. Appleton-Century-Crofts. New York, N.Y.
Maffei, C., Jovicich, J., De Benedictis, A., Corsini, F., Barbareschi, M., Chioffi, F., & Sarubbo, S. (2018). Topography of the human acoustic radiation as revealed by ex vivo fibers micro-dissection and in vivo diffusion-based tractography. *Brain Structure and Function*, *223*(1), 449-459.
Maldonado, I. L., de Champfleur, N. M., Velut, S., Destrieux, C., Zemmoura, I., & Duffau, H. (2013). Evidence of a middle longitudinal fasciculus in the human brain from fiber dissection. *Journal of anatomy*, *223*(1), 38-45.
Marie, D., Roth, M., Lacoste, R., Nazarian, B., Bertello, A., Anton, J. L., ... & Meguerditchian, A. (2018). Left brain asymmetry of the planum temporale in a nonhominid primate: Redefining the origin of brain specialization for language. *Cerebral Cortex*, *28*(5), 1808-1815.
Mars, R. B., Jbabdi, S., Sallet, J., O'Reilly, J. X., Croxson, P. L., Olivier, E., ... & Rushworth, M. F. (2011). Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *Journal of Neuroscience*, *31*(11), 4087-4100.
Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. (2012). On the relationship between the "default mode network" and the "social brain". *Frontiers in human neuroscience*, *6*, 189.
Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2011). Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cerebral cortex*, *22*(8), 1894-1903.
Masland, R. L. (1968). Some neurological processes underlying language. Ann. Otol. Rhinal. Laryngol. 77: 787-805.
Masouleh, S. K., Eickhoff, S. B., Hoffstaedter, F., Genon, S., & Alzheimer's Disease Neuroimaging Initiative. (2019). Empirical examination of the replicability of associations between brain structure and psychological variables. *elife*, *8*, e43464.
Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, *30*(3), 1481-1498.
Matthews B. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et biophysica acta. 1975;405(2):442–451.
Mazziotta et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Phil. Trans. Royal Soc. B Biol. Sci. 356(1412):1293-1322 (2001)
McAndrews, M. P., Girard, T. A., Wilkins, L. K., & McCormick, C. (2016). Semantic congruence affects hippocampal response to repetition of visual associations. *Neuropsychologia*, *90*, 235-242.
McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left?. *Trends in cognitive sciences*, *16*(5), 269-276.

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, *119*(3), 907-917.
McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: applications and advances. *Neuroimage*, *23*, S250-S263.
McKeown, M. J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: examining the assumptions. *Human brain mapping*, *6*(5-6), 368-372.
Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men?. *Science*, *317*(5834), 82-82.
Meila, M. (2016). Spectral Clustering: a Tutorial for the 2010's. In *Handbook of cluster analysis* (pp. 1-23). CRC Press.
Mennes, M., Kelly, C., Zuo, X. N., Di Martino, A., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2010). Inter-individual differences in resting-state functional connectivity predict task-induced BOLD activity. *Neuroimage*, *50*(4), 1690-1701.
Meola, A., Comert, A., Yeh, F. C., Stefaneanu, L., & Fernandez-Miranda, J. C. (2015). The controversial existence of the human superior fronto-occipital fasciculus: Connectome-based tractographic study with microdissection validation. *Human brain mapping*, *36*(12), 4964-4971.
Messé, A. (2020). Parcellation influence on the connectivity-based structure–function relationship in the human brain. *Human brain mapping*, *41*(5), 1167-1180.
Mestres-Missé, A., Bazin, P. L., Trampel, R., Turner, R., & Kotz, S. A. (2014). Dorsomedial striatum involvement in regulating conflict between current and presumed outcomes. *NeuroImage*, *98*, 159-167.
Mesulam MM (1998): From sensation to cognition. *Brain* 121:1013– 1052.
Meunier, D., Lambiotte, R., Fornito, A., Ersche, K., & Bullmore, E. T. (2009). Hierarchical modularity in human brain functional networks. *Frontiers in neuroinformatics*, *3*, 37.
Meyer, M., Zaehle, T., Gountouna, V. E., Barron, A., Jancke, L., & Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *Neuroreport*, *16*(18), 1985-1989.
Mitra, P. P. (2014). The circuit architecture of whole brains at the mesoscopic scale. *Neuron*, *83*(6), 1273-1283.
Moayedi, M., Salomons, T. V., Dunlop, K. A., Downar, J., & Davis, K. D. (2015). Connectivity-based parcellation of the human frontal polar cortex. *Brain structure & function*, *220*(5), 2603.
modality-independent components of the human imagery system. *Neuroimage*, *52*(2), 677-685.
Møller, A. M., & Myles, P. S. (2016). What makes a good systematic review and meta-analysis?.
Möller, J., Jansma, B. M., Rodriguez-Fornells, A. & Münte, T. F. What the Brain Does before the Tongue Slips. *Cerebral Cortex* **17**, 1173-1178, doi:10.1093/cercor/bhl028 (2007).
Möller, J., Jansma, B. M., Rodriguez-Fornells, A., & Münte, T. F. (2007). What the brain does before the tongue slips. *Cerebral Cortex*, *17*(5), 1173-1178.
Monti, S., Tamayo, P., Mesirov, J., & Golub, T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, *52*(1-2), 91-118.

Morosan, P., Rademacher, J., Palomero-Gallagher, N., & Zilles, K. (2005b). Anatomical organization of the human auditory cortex: cytoarchitecture and transmitter receptors. *The auditory cortex: a synthesis of human and animal research*, 27-50.
Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage*, *13*(4), 684-701.
Morosan, P., Schleicher, A., Amunts, K., & Zilles, K. (2005a). Multimodal architectonic mapping of human superior temporal gyrus. *Anatomy and embryology*, *210*(5-6), 401-406.
Motley, M. T., Camden, C. T. & Baars, B. J. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning & Verbal Behavior* **21**, 578-594, doi:10.1016/S0022-5371(82)90791-5 (1982).
Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*(5), 578-594.
Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J. P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *science*, *328*(5980), 876-878.
Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., ... & Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron*, *77*(3), 586-595.
Mumford, J. A. & Nichols, T. E., Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. *Neuroimage* **39**, 261–268. (2008).
Mumford, J. A., & Nichols, T. E. (2008). Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. *Neuroimage*, *39*(1), 261-268.
musicians. *SCIENCE-NEW YORK THEN WASHINGTON-*, 699-699.
Mustovic, H., Scheffler, K., Di Salle, F., Esposito, F., Neuhoff, J. G., Hennig, J., & Seifritz, E. (2003). Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage*, *20*(1), 429-434.
Naeser, M. A. (1987). Nancy Helm-Estabrooks N, Haas G, Auerbach S, Srinivasan M. Relationship between lesion extent in 'Wernicke's area'on computed tomographic scan and predicting recovery of comprehension in Wernicke's aphasia. *Arch Neurol*, *44*, 73-82.
Nanetti, L., Cerliani, L., Gazzola, V., Renken, R., & Keysers, C. (2009). Group analyses of connectivity-based cortical parcellation using repeated k-means clustering. *Neuroimage*, *47*(4), 1666-1677.
Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, *13*(12), 1362-1368.
Neubert, F. X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron*, *81*(3), 700-713.
Nicolini C, Bordier C, Bifone A. 2017. Community detection in weighted brain connectivity networks beyond the resolution limit. Neuroimage 146:28–39

Nielsen, F. Å. (2004). The Indispensable Brede Database and Brede Neuroinformatics Toolbox.
Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech communication*, *47*(1-2), 43-58.
Nooteboom, S. G. Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication* **47**, 43-58,
Northoff, G., Qin, P., & Nakao, T. (2010). Rest-stimulus interaction in the brain: a review. *Trends in neurosciences*, *33*(6), 277-284
Nozari, N., Dell, G. S. & Schwartz, M. F. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive psychology* **63**, 1-33, doi:10.1016/j.cogpsych.2011.05.001 (2011).
Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive psychology*, *63*(1), 1-33.
O'Muircheartaigh, J., & Jbabdi, S. (2018). Concurrent white matter bundles and grey matter networks using independent component analysis. *NeuroImage*, *170*, 296-306.
Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2006). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebral Cortex*, *17*(10), 2251-2257.
Ojemann, G. A. (1979). Individual variability in cortical localization of language. *Journal of neurosurgery*, *50*(2), 164-169.
Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and language*, *98*(1), 112-117.
Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience letters*, *452*(3), 219-223.
Okada, K., Matchin, W., & Hickok, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic bulletin & review*, *25*(1), 423-430.
Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., ... & Hickok, G. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486-2495.
Okada, K., Smith, K. R., Humphries, C., & Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport*, *14*(18), 2323-2326.
Okada, K., Venezia, J. H., Matchin, W., Saberi, K., & Hickok, G. (2013). An fMRI study of audiovisual speech perception reveals multisensory interactions in auditory cortex. *PloS one*, *8*(6), e68959.
Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & cognition*, *38*(8), 1147-1160.
Overath, T., Cusack, R., Kumar, S., Von Kriegstein, K., Warren, J. D., Grube, M., ... & Griffiths, T. D. (2007). An information theoretic characterisation of auditory encoding. *PLoS biology*, *5*(11), e288.

Overath, T., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *Journal of Neuroscience*, *28*(49), 13268-13273.
Pa, J., & Hickok, G. (2008). A parietal–temporal sensory–motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia*, *46*(1), 362-368.
Parisot, S., Arslan, S., Passerat-Palmbach, J., Wells III, W. M., & Rueckert, D. (2016). Group-wise parcellation of the cortex through multi-scale spectral clustering. *NeuroImage*, *136*, 68-83.
Park, H. J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, *342*(6158), 1238411.
Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert systems with applications*, *36*(2), 3336-3341.
Parrish, T. B., Gitelman, D. R., LaBar, K. S., & Mesulam, M. M. (2000). Impact of signal-to-noise on functional MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *44*(6), 925-932.
Paschke, L. M., Walter, H., Steimke, R., Ludwig, V. U., Gaschler, R., Schubert, T., & Stelzel, C. (2015). Motivation by potential gains and losses affects control processes via different mechanisms in the attentional network. *Neuroimage*, *111*, 549-561.
Passingham, R. E., Stephan, K. E., & Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nature Reviews Neuroscience*, *3*(8), 606-616.
Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767-776.
Pavani, F., Macaluso, E., Warren, J. D., Driver, J., & Griffiths, T. D. (2002). A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Current Biology*, *12*(18), 1584-1590.
Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624-634.
Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, *30*(30), 10127-10134.
Peelle, J. E., Johnsrude, I., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in human neuroscience*, *4*, 51.
Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., & Sams, M. (2006). Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Human brain mapping*, *27*(6), 471-477.
Penfield, W. and Roberts, L. (1959). Speech and brain-mechanisms. Princeton University Press, Princeton, N.J.
Penfield,W. ,and Roberts,L.(1959).Speech and Brain mechanisms. Princeton University Press, Princeton, N.J.
Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, *45*(1), S199-S209.
Peschke, C., Ziegler, W., Kappes, J., & Baumgaertner, A. (2009). Auditory–motor integration during fast repetition: the neuronal correlates of shadowing. *Neuroimage*, *47*(1), 392-402.

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of life reviews*, *11*(3), 400-435.
Petrides, M., & Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *Journal of Comparative Neurology*, *273*(1), 52-66
Petrides, M., & Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European Journal of Neuroscience*, *16*(2), 291-310.
Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. (2013). Good things peak in pairs: a note on the bimodality coefficient. *Frontiers in psychology*, *4*, 700.
Pillay, S. B., Binder, J. R., Humphries, C., Gross, W. L., & Book, D. S. (2017). Lesion localization of speech comprehension deficits in chronic aphasia. *Neurology*, *88*(10), 970-975.
Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience*, *32*(41), 14125-14131.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?. *Trends in cognitive sciences*, *10*(2), 59-63.
Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-
Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. *Neuroimage*, *62*(2), 1216-1220.
Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual review of psychology*, *67*, 587-612.
Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., ... & Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience*, *18*(2), 115-126.
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.
Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the new York Academy of Sciences*, *1191*(1), 62-88.
Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847.
Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*(3-4), 262-275.
Price, C. J., Crinion, J., & MacSweeney, M. (2011). A generative model of speech production in Broca's and Wernicke's areas. *Frontiers in psychology*, *2*, 237.
Price, C. J., Moore, C. J., & Friston, K. J. (1997). Subtractions, conjunctions, and interactions in experimental design of activation studies. *Human brain mapping*, *5*(4), 264-272.
publication bias in activation likelihood estimation (ALE) meta-analyses for fMRI. *PloS one*, *13*(11), e0208177.
Quadflieg, S., Gentile, F., & Rossion, B. (2015). The neural basis of perceiving person interactions. *Cortex*, *70*, 5-20.
Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., & Zilles, K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage*, *13*(4), 669-683.

Raettig, T., Frisch, S., Friederici, A. D., & Kotz, S. A. (2010). Neural correlates of morphosyntactic and verb-argument structure processing: an EfMRI study. *Cortex*, *46*(5), 613-620.
Raichle, M. E. (2010). Two views of brain function. *Trends in cognitive sciences*, *14*(4), 180-190.
Rathkopf, C. A. (2013). Localization and intrinsic function. *Philosophy of Science*, *80*(1), 1-21.
Rauschecker, A. M., Bowen, R. F., Parvizi, J., & Wandell, B. A. (2012). Position sensitivity in the visual word form area. *Proceedings of the National Academy of Sciences*, *109*(24), E1568-E1577.
Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing research*, *271*(1), 16-25.
Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience*, *12*(6), 718-724.
Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*(22), 11800-11806.
Ray, K. L., Zald, D. H., Bludau, S., Riedel, M. C., Bzdok, D., Yanes, J., ... & Laird, A. R. (2015). Co-activation based parcellation of the human frontal pole. *NeuroImage*, *123*, 200-211.
Raznahan, A. (2018). On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage*, *178*, 540-551.
Richardson, F. M., Thomas, M. S., & Price, C. J. (2010). Neuronal activation for semantically reversible sentences. *Journal of cognitive neuroscience*, *22*(6), 1283-1298.
Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *Neuroimage*, *6*(4), 288-304.
Roberts, J. A., Perry, A., Lord, A. R., Roberts, G., Mitchell, P. B., Smith, R. E., ... & Breakspear, M. (2016). The contribution of geometry to the human connectome. *Neuroimage*, *124*, 379-393.
Rogalsky, C., & Hickok, G. (2010). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, *23*(7), 1664-1680.
Rogalsky, C., Poppa, T., Chen, K. H., Anderson, S. W., Damasio, H., Love, T., & Hickok, G. (2015). Speech repetition as a window on the neurobiology of auditory–motor integration for speech: A voxel-based lesion symptom mapping study. *Neuropsychologia*, *71*, 18-27.
Roland, P. E., Geyer, S., Amunts, K., Schormann, T., Schleicher, A., Malikovic, A., & Zilles, K. (1997). Cytoarchitectural maps of the human brain in standard anatomical space. *Human brain mapping*, *5*(4), 222-227.
Rosen, B. R., & Savoy, R. L. (2012). fMRI at 20: has it changed the world?. *Neuroimage*, *62*(2), 1316-1324.
Rousseeuw, P. J., & Kaufman, L. (1990). Finding groups in data. *Hoboken: Wiley Online Library*, *1*.
Roy, A. K., Shehzad, Z., Margulies, D. S., Kelly, A. C., Uddin, L. Q., Gotimer, K., ... & Milham, M. P. (2009). Functional connectivity of the human amygdala using resting state fMRI. *Neuroimage*, *45*(2), 614-626.

Runnqvist, E., Chanoine, V., Strijkers, K., Patamadilok, C., Bonnard, M., Nazarian, B., ... & Alario, F. X. (2020). The cerebellum is involved in internal and external speech error monitoring.
Ruschel, M., Knösche, T. R., Friederici, A. D., Turner, R., Geyer, S., & Anwander, A. (2013). Connectivity architecture and subdivision of the human inferior parietal cortex revealed by diffusion MRI. *Cerebral cortex*, *24*(9), 2436-2448.
Saad, Z. S., Gotts, S. J., Murphy, K., Chen, G., Jo, H. J., Martin, A., & Cox, R. W. (2012). Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain connectivity*, *2*(1), 25-32.
Saito, Y., Ishii, K., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *Neuroreport*, *17*(18), 1893-1897.
Salehi, M., Greene, A. S., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. T. (2020). There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *NeuroImage*, *208*, 116366.
Salehi, M., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. T. (2018). An exemplar-based approach to individualized parcellation reveals the need for sex specific functional networks. *NeuroImage*, *170*, 54-67.
Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., & Nichols, T. E. (2009). Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *Neuroimage*, *45*(3), 810-823.
Samartsidis, P., Montagna, S., Laird, A. R., Fox, P. T., Johnson, T. D., & Nichols, T. E. (2017). Estimating the number of missing experiments in a neuroimaging meta-analysis. *BioRxiv*, 225425.
Samartsidis, P., Montagna, S., Nichols, T. E., & Johnson, T. D. (2017). The coordinate-based meta-analysis of neuroimaging data. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *32*(4), 580.
Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., & Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain*, *130*(3), 630-653.
Scholz, M. (2012). Validation of nonlinear PCA. *Neural processing letters*, *36*(1), 21-30.
Scholz, M. J., Steinley, D., Stevens, J., & Cradit, J. D. (2019). Affinity propagation: An exemplar-based tool for clustering in psychological research. *British Journal of Mathematical and Statistical Psychology*, *72*(1), 155-182.
Scholz, M., Kaplan, F., Guy, C. L., Kopka, J., & Selbig, J. (2005). Non-linear PCA: a missing data approach. *Bioinformatics*, *21*(20), 3887-3895.
Schon, K., Newmark, R. E., Ross, R. S., & Stern, C. E. (2016). A working memory buffer in parahippocampal regions: evidence from a load effect during the delay period. *Cerebral Cortex*, *26*(5), 1965-1974.
Schönwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences*, *106*(34), 14611-14616.
Schubotz, R. I., Anwander, A., Knösche, T. R., von Cramon, D. Y., & Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in
Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461-464.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in neurosciences*, *26*(2), 100-107.
Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400-2406.
Selnes, O. A., Knopman, D. S., Niccum, N., Rubens, A. B., & Larson, D. (1983). Computed tomographic scan correlates of auditory comprehension deficits in aphasia: a prospective recovery study. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *13*(5), 558-566.
Selnes, O. A., Knopman, D. S., Niccumm, N., & Rubens, A. B. (1985). The critical role Wernicke's area in sentence repetition. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *17*(6), 549-557.
Şenbabaoğlu, Y., Michailidis, G., & Li, J. Z. (2014). Critical limitations of consensus clustering in class discovery. *Scientific reports*, *4*(1), 1-13.
Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H., & Johnson, K. A. (2012). Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. *Journal of Neuroscience*, *32*(31), 10649-10661.
Shadmehr, R. & Krakauer, J. W. A computational neuroanatomy for motor control. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale* **185**, 359-381, doi:10.1007/s00221-008-1280-5 (2008).
Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental brain research*, *185*(3), 359-381.
Shah, L. M., Cramer, J. A., Ferguson, M. A., Birn, R. M., & Anderson, J. S. (2016). Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and behavior*, *6*(5), e00456.
Shapleske, J., Rossell, S. L., Woodruff, P. W. R., & David, A. S. (1999). The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. *Brain Research Reviews*, *29*(1), 26-49.
Sheets, J. R., Briggs, R. G., Bai, M. Y., Poologaindran, A., Young, I. M., Conner, A. K., ... & Sughrue, M. E. (2020). Parcellation-based modeling of the dorsal premotor area. *Journal of the Neurological Sciences*, *415*, 116907.
Shehzad, Z., Kelly, A. C., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., ... & Milham, M. P. (2009). The resting brain: unconstrained yet reliable. *Cerebral cortex*, *19*(10), 2209-2229.
Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*, *82*, 403-415.
Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological medicine*, *31*(2), 241.
Sherwood, C. C. (2010). Wernicke's area homologue in chimpanzees (Pan troglodytes) and its relation to the appearance of modern human language. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1691), 2165-2174.
Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex*

Shukla, S., & Naganna, S. (2014). A review on K-means data clustering approach. *International Journal of Information and Computation Technology*, *4*(17), 1847-1860.
Siegel, J. S., Seitzman, B. A., Ramsey, L. E., Ortega, M., Gordon, E. M., Dosenbach, N. U., ... & Corbetta, M. (2018). Re-emergence of modular brain networks in stroke recovery. *Cortex*, *101*, 44-59.
Simpson, T. I., Armstrong, J. D., & Jarman, A. P. (2010). Merged consensus clustering to assess and improve class discovery with microarray data. *BMC bioinformatics*, *11*(1), 1-12.
Skudlarski, P., Jagannathan, K., Calhoun, V. D., Hampson, M., Skudlarska, B. A., & Pearlson, G. (2008). Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. *Neuroimage*, *43*(3), 554-561.
Small, S. L., & Hickok, G. (2016). The neurobiology of language. In *Neurobiology of language* (pp. 3-9). Academic Press.
Smiley, J. F., & Falchier, A. (2009). Multisensory connections of monkey auditory cerebral cortex. *Hearing research*, *258*(1-2), 37-46.
Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *Journal of Comparative Neurology*, *502*(6), 894-923.
Smith, K. R., Hsieh, I. H., Saberi, K., & Hickok, G. (2010). Auditory spatial and object processing in the human planum temporale: no evidence for selectivity. *Journal of Cognitive Neuroscience*, *22*(4), 632-639.
Smith, K. R., Okada, K., Saberi, K., & Hickok, G. (2004). Human cortical auditory motion areas are not motion selective. *Neuroreport*, *15*(9), 1523-1526.
Smith, K. R., Saberi, K., & Hickok, G. (2007). An event-related fMRI study of auditory motion perception: no evidence for a specialized cortical system. *Brain research*, *1150*, 94-99.
Smith, R. E., Tournier, J. D., Calamante, F., & Connelly, A. (2015). The effects of SIFT on the reproducibility and biological accuracy of the structural connectome. *Neuroimage*, *104*, 253-265.
Smith, R. E., Tournier, J. D., Calamante, F., & Connelly, A. (2015). SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *Neuroimage*, *119*, 338-351.
Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*(1), 83-98.
Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... & Niazy, R. K. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*, S208-S219.
Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., ... & Van Essen, D. C. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, *109*(8), 3131-3136.
Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., ... & Van Essen, D. C. (2013). Functional connectomics from resting-state fMRI. *Trends in cognitive sciences*, *17*(12), 666-682.

Sormaz, M., Murphy, C., Wang, H. T., Hymers, M., Karapanagiotidis, T., Poerio, G., ... & Smallwood, J. (2018). Default mode network can support the level of detail in experience during active task states. *Proceedings of the National Academy of Sciences*, *115*(37), 9318-9323.

Sörös, P., Inamoto, Y., & Martin, R. E. (2009). Functional brain imaging of swallowing: An activation likelihood estimation meta-analysis. *Human brain mapping*, *30*(8), 2426-2439.

Sörös, P., Sokoloff, L. G., Bose, A., McIntosh, A. R., Graham, S. J., & Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *Neuroimage*, *32*(1), 376-387.

Spadone, S., Della Penna, S., Sestieri, C., Betti, V., Tosoni, A., Perrucci, M. G., ... & Corbetta, M. (2015). Dynamic reorganization of human resting-state networks during visuospatial attention. *Proceedings of the National Academy of Sciences*, *112*(26), 8112-8117.

Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *Neuroimage*, *20*(4), 1944-1954.

Specht, K., Osnes, B., & Hugdahl, K. (2009). Detection of differential speech-specific processes in the temporal lobe using fMRI and a dynamic "sound morphing" technique. *Human brain mapping*, *30*(10), 3436-3444.

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E. & Wise, R. J. Converging language streams in the human temporal lobe. *J. Neurosci.* **26**, 7328–7336 (2006).

Spocter, M. A., Hopkins, W. D., Garrison, A. R., Bauernfeind, A. L., Stimpson, C. D., Hof, P. R., & Sherwood, C. C. (2010). Wernicke's area homologue in chimpanzees (Pan troglodytes) and its relation to the appearance of modern human language. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1691), 2165-2174.

Spocter, M. A., Hopkins, W. D., Garrison, A. R., Bauernfeind, A. L., Stimpson, C. D., Hof, P. R., &
Spss, I. I. B. M. (2011). IBM SPSS statistics for Windows, version 20.0. *New York: IBM Corp*, *440*.

Stark, C. E., & Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proceedings of the national Academy of Sciences*, *98*(22), 12760-12766.

Steinmetz, H., & Seitz, R. J. (1991). Functional anatomy of language processing: neuroimaging and the problem of individual variability. *Neuropsychologia*, *29*(12), 1149-1161.

Strehl, A., & Ghosh, J. (2002). Cluster ensembles---a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, *3*(Dec), 583-617.

studies of the Wisconsin Card-Sorting task and component processes. *Human brain mapping*, *25*(1), 35-45.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D. A., Jones, D. T., Konermann, C., ... & Kool, M. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer cell*, *22*(4), 425-437.

Sweet, R. A., Dorph-Petersen, K. A., & Lewis, D. A. (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *Journal of Comparative Neurology*, *491*(3), 270-289.

Swift, S., Tucker, A., Vinciotti, V., Martin, N., Orengo, C., Liu, X., & Kellam, P. (2004). Consensus clustering and functional interpretation of gene-expression data. *Genome biology*, *5*(11), R94.
Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The annals of statistics*, *35*(6), 2769-2794.
Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, *221*, 117164.
Szycik, G. R., Jansma, H., & Münte, T. F. (2009). Audiovisual integration during speech comprehension: An fMRI study comparing ROI-based and whole brain analyses. *Human brain mapping*, *30*(7), 1990-1999.
Tanaka, S., & Kirino, E. (2019). Increased functional connectivity of the angular gyrus during imagined music performance. *Frontiers in human neuroscience*, *13*, 92.
Taylor, P. N., Wang, Y., & Kaiser, M. (2017). Within brain area tractography suggests local modularity using high resolution connectomics. *Scientific reports*, *7*, 39859.
Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., Van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Human brain mapping*, *10*(2), 74-79.
Teszner, D., Tzavaras, A., Gruner, J., & Hécaen, H. (1972). L'asymétrie droite-gauche du planum temporale: à propos de l'étude anatomique de 100 cerveaux. *Revue Neurologique*, *126*, 444-449.
Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., & Poline, J. B. (2006). Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Human brain mapping*, *27*(8), 678-693.
Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., & Poline, J. B. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, *35*(1), 105-120.
Thirion, B., Varoquaux, G., Dohmatob, E., & Poline, J. B. (2014). Which fMRI clustering gives good brain parcellations?. *Frontiers in neuroscience*, *8*, 167.
Thothathiri, M., Kimberg, D. Y., & Schwartz, M. F. (2012). The neural basis of reversible sentence comprehension: evidence from voxel-based lesion symptom mapping in aphasia. *Journal of Cognitive Neuroscience*, *24*(1), 212-222.
Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1-12.
Tian, Y., & Zalesky, A. (2018). Characterizing the functional connectivity diversity of the insula cortex: Subregions, diversity curves and behavior. *Neuroimage*, *183*, 716-733.
Tibon, R., Fuhrmann, D., Levy, D. A., Simons, J. S., & Henson, R. N. (2019). Multimodal integration and vividness in the angular gyrus during episodic encoding and retrieval. *Journal of Neuroscience*, *39*(22), 4365-4374.
Tittgemeyer, M. (2010). Anatomical and functional parcellation of the human lateral premotor cortex. *Neuroimage*, *50*(2), 396-408.
Tomassini, V., Jbabdi, S., Klein, J. C., Behrens, T. E., Pozzilli, C., Matthews, P. M., ... & Johansen-Berg, H. (2007). Diffusion-weighted imaging tractography-based parcellation of the human lateral premotor cortex identifies dorsal and ventral subregions with anatomical and functional specializations. *Journal of Neuroscience*,

Tourville, J. A., Nieto-Castañón, A., Heyne, M., & Guenther, F. H. (2019). Functional parcellation of the speech production cortex. *Journal of Speech, Language, and Hearing Research*, *62*(8S),

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. Neuroimage, 39(3), 1429-1443.

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, *9*(1), 1-12.

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and language*, *162*, 60-71.

Tremblay, P., Baroni, M., & Hasson, U. (2013). Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. *Neuroimage*, *66*, 318-332.

Tremblay, P., Deschamps, I., & Gracco, V. L. (2013). Regional heterogeneity in the processing and the production of speech in the human planum temporale. *Cortex*, *49*(1), 143-157.

Tremblay, P., Perron, M., Deschamps, I., Kennedy-Higgins, D., Houde, J. C., Dick, A. S., & Descoteaux, M. (2019). The role of the arcuate and middle longitudinal fasciculi in speech perception in noise in adulthood. *Human brain mapping*, *40*(1), 226-241.

Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and language*, *114*(1), 1-15

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage*, *16*(3), 765-780.

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human brain mapping*, *33*(1), 1-13.

Turken, A., & Dronkers, N. (2011). Hemispheric asymmetries in the resting-sate functional connectivity patterns of the brain regions critical for language comprehension. In *Front. Neuroinform. Conference Abstract: 4th INCF Congress of Neuroinformatics. doi: 10.3389/conf. fninf* (Vol. 44).

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*(1), 1-10.

Tzourio-Mazoyer, N. (2016). Intra-and Inter-hemispheric Connectivity Supporting Hemispheric Specialization. In *Micro-, Meso-and Macro-Connectomics of the Brain* (pp. 129-146). Springer International Publishing.

Tzourio-Mazoyer, N., & Mazoyer, B. (2017). Variations of planum temporale asymmetries with Heschl's Gyri duplications and association with cognitive abilities: MRI investigation of 428 healthy volunteers. *Brain structure & function*.

Van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European neuropsychopharmacology*, *20*(8), 519-534.

Van Den Heuvel, M., Mandl, R., & Pol, H. H. (2008). Normalized cut group clustering of resting-state FMRI data. *PloS one*, *3*(4), e2001.

Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of neurophysiology*, *103*(1), 297-321.

Van Elk, M., Arciniegas Gomez, M. A., van der Zwaag, W., Van Schie, H. T., & Sauter, D. (2019). The neural correlates of the awe experience: Reduced default mode network activity during feelings of awe. *Human brain mapping*, *40*(12), 3561-3574.
van Erp, S., Verhagen, J., Grasman, R. P., & Wagenmakers, E. J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in Psychological Bulletin from 1990–2013. *Journal of Open Psychology Data*, *5*(1).
Van Essen, D. C. (2009). Lost in localization—But found with foci?!. *Neuroimage*, *48*(1), 14-17.
Van Essen, D. C. (2013). Cartography and connectomes. *Neuron*, *80*(3), 775-790.
Van Essen, D. C., & Glasser, M. F. (2014). In vivo architectonics: a cortico-centric perspective. *Neuroimage*, *93*, 157-164.
Van Essen, D. C., Donahue, C. J., Coalson, T. S., Kennedy, H., Hayashi, T., & Glasser, M. F. (2019). Cerebral cortical folding, parcellation, and connectivity in humans, nonhuman primates, and mice. *Proceedings of the National Academy of Sciences*, *116*(52), 26173-26180.
Van Essen, D. C., Donahue, C., Dierker, D. L., & Glasser, M. F. (2016). Parcellations and connectivity patterns in human and macaque cerebral cortex. In *Micro-, Meso-and Macro-Connectomics of the Brain* (pp. 89-106). Springer, Cham.
Vanni, M. P., Chan, A. W., Balbi, M., Silasi, G., & Murphy, T. H. (2017). Mesoscale mapping of mouse cortex reveals frequency-dependent cycling between distinct macroscale functional modules. *Journal of Neuroscience*, *37*(31), 7513-7533.
Vartanian, O., & Skov, M. (2014). Neural correlates of viewing paintings: Evidence from a quantitative meta-analysis of functional magnetic resonance imaging data. *Brain and cognition*, *87*, 52-56.
Velenosi, L. A., Wu, Y. H., Schmidt, T. T., & Blankenburg, F. (2020). Intraparietal sulcus maintains working memory representations of somatosensory categories in an adaptive, context-dependent manner. *NeuroImage*, *221*, 117146.
Venezia, J. H., Thurman, S. M., Richards, V. M., & Hickok, G. (2019). Hierarchy of speech-driven spectrotemporal receptive fields in human auditory cortex. *Neuroimage*, *186*, 647-666.
Venezia, J. H., Vaden Jr, K. I., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, visual and audiovisual speech processing streams in superior temporal sulcus. *Frontiers in Human Neuroscience*, *11*, 174.
Venezia, J. H., Vaden Jr, K. I., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, visual and audiovisual speech processing streams in superior temporal sulcus. *Frontiers in Human Neuroscience*, *11*, 174.
Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC neuroscience*, *10*(1), 1-11.
Viceic, D., Fornari, E., Thiran, J. P., Maeder, P. P., Meuli, R., Adriani, M., & Clarke, S. (2006). Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. *Neuroreport*, *17*(16), 1659-1662.
Vigneau, Mathieu, Virginie Beaucousin, Pierre-Yves Herve, Hugues Duffau, Fabrice Crivello, Olivier Houde, Bernard Mazoyer, and Nathalie Tzourio-Mazoyer. "Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing." *Neuroimage* 30, no. 4 (2006): 1414-1432.

Vingerhoets, A. J., Bylsma, L. M., & De Vlam, C. (2013). Swearing: A biopsychosocial perspective. *Psihologijske teme*, *22*(2), 287-304.
Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, *13*(7), 994-1005.
Voxel-based lesion–symptom mapping. *Nature neuroscience*, *6*(5), 448-450.
Wager, T. D., Lindquist, M., & Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Social cognitive and affective neuroscience*, *2*(2), 150-158.
Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., ... & Cohen, J. D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science*, *303*(5661), 1162-1167.
Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. *Cleveland Clinic journal of medicine*, *75*(6), 431.
Wallace, M. N., Johnston, P. W., & Palmer, A. R. (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Experimental Brain Research*, *143*(4), 499-508.
Wang, J., Fan, L., Zhang, Y., Liu, Y., Jiang, D., Zhang, Y., ... & Jiang, T. (2012). Tractography-based parcellation of the human left inferior parietal lobule. *Neuroimage*, *63*(2), 641-652.
Wang, J., Wang, L., Zang, Y., Yang, H., Tang, H., Gong, Q., ... & He, Y. (2009). Parcellation-dependent small-world brain functional networks: A resting-state fMRI study. *Human brain mapping*, *30*(5), 1511-1523.
Wang, J., Xie, S., Guo, X., Becker, B., Fox, P. T., Eickhoff, S. B., & Jiang, T. (2017). Correspondent Functional Topography of the Human Left Inferior Parietal Lobule at Rest and Under Task Revealed Using Resting-State f MRI and Coactivation Based Parcellation. *Human brain mapping*, *38*(3), 1659-1675.
Wang, J., Yang, Y., Fan, L., Xu, J., Li, C., Liu, Y., ... & Jiang, T. (2015). Convergent functional architecture of the superior parietal lobule unraveled with multimodal neuroimaging approaches. *Human brain mapping*, *36*(1), 238-257.
Wang, K., Zhang, J., Li, D., Zhang, X., & Guo, T. (2008). Adaptive affinity propagation clustering. *arXiv preprint arXiv:0805.1096*.
Wang, R., Liu, M., Cheng, X., Wu, Y., Hildebrandt, A., & Zhou, C. (2021). Segregation, integration, and balance of large-scale resting brain networks configure different cognitive abilities. *Proceedings of the National Academy of Sciences*, *118*(23).
Wang, S. F., Ritchey, M., Libby, L. A., & Ranganath, C. (2016). Functional connectivity based parcellation of the human medial temporal lobe. *Neurobiology of learning and memory*, *134*, 123-134.
Wang, Y., & Pan, Y. (2014). Semi-supervised consensus clustering for gene expression data analysis. *BioData mining*, *7*(1), 1-13.
Wang, Y., Fernández-Miranda, J. C., Verstynen, T., Pathak, S., Schneider, W., & Yeh, F. C. (2013). Rethinking the role of the middle longitudinal fascicle in language and auditory pathways. *Cerebral cortex*, *23*(10), 2347-2356.
Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience*, *23*(13), 5799-5804.

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience*, *23*(13), 5799-5804.
Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences*, *100*(17), 10038-10042.
Warren, J. D., Zielinski, B. A., Green, G. G., Rauschecker, J. P., & Griffiths, T. D. (2002). Perception of sound-source motion by the human brain. *Neuron*, *34*(1), 139-148.
Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: auditory–motor transformations and the posterior temporal plane. *Trends in neurosciences*, *28*(12), 636-643.
Warrington, S., Bryant, K. L., Khrapitchev, A. A., Sallet, J., Charquero-Ballester, M., Douaud, G., ... & Sotiropoulos, S. N. (2020). XTRACT-Standardised protocols for automated tractography in the human and macaque brain. *Neuroimage*, *217*, 116923.
Weiskopf, N., Mohammadi, S., Lutti, A., & Callaghan, M. F. (2015). Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology. *Current opinion in neurology*, *28*(4), 313-322.
Welvaert, M., & Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PloS one*, *8*(11), e77089.
Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Cohn & Weigert.
Westbury, C. F., Zatorre, R. J., & Evans, A. C. (1999). Quantifying variability in the planum temporale: a probability map. *Cerebral Cortex*, *9*(4), 392-405.
Wig, G. S. (2017). Segregated systems of human brain networks. *Trends in cognitive sciences*, *21*(12), 981-996.
Williams, N., Arnulfo, G., Wang, S. H., Nobili, L., Palva, S., & Palva, J. M. (2019). Comparison of methods to identify modules in noisy or incomplete brain networks. *Brain connectivity*, *9*(2), 128-143.
Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage*, *171*, 62-74.
Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within Wernicke's area'. *Brain*, *124*(1), 83-95.
Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880-1882.
Wu J, Ngo GH, Greve DN, Li J, He T, Fischl B, Eickhoff SB, Yeo BTT. Accurate nonlinear mapping between MNI volumetric and FreeSurfer surface coordinate systems, Human Brain Mapping 39:3793–3808, 2018
Wu, J., Ngo, G. H., Greve, D., Li, J., He, T., Fischl, B., ... & Yeo, B. T. (2018). Accurate nonlinear mapping between MNI volumetric and FreeSurfer surface coordinate systems. *Human brain mapping*, *39*(9), 3793-3808.
Wu, L., Calhoun, V. D., Jung, R. E., & Caprihan, A. (2015). Connectivity-based whole brain dual parcellation by group ICA reveals tract structures and decreased connectivity in schizophrenia. *Human brain mapping*, *36*(11), 4681-4701.
Wu, Z., Xu, D., Potter, T., Zhang, Y., & Alzheimer's Disease Neuroimaging Initiative. (2019). Effects of brain parcellation on the characterization of topological deterioration in Alzheimer's disease. *Frontiers in aging neuroscience*, *11*, 113.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one*, *8*(7), e68910.
Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., ... & Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, *27*(2), 173-183.
Yang, Z., Zhuang, X., Sreenivasan, K., Mishra, V., Cordes, D., & Alzheimer's Disease Neuroimaging Initiative. (2019). Robust motion regression of resting-state data using a convolutional neural network model. *Frontiers in neuroscience*, *13*, 169.
Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature methods*, *8*(8), 665-670.
Yarkoni, T., Poldrack, R. A., Van Essen, D. C., & Wager, T. D. (2010). Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends in cognitive sciences*, *14*(11), 489-496.
Ye, Z., Donamayor, N., & Muente, T. F. (2014). Brain network of semantic integration in sentence reading: insights from independent component analysis and graph theoretical analysis. *Human brain mapping*, *35*(2), 367-376.
Yeni-Komshian, G. H., & Benson, D. A. (1976). Anatomical study of cerebral asymmetry in the temporal lobe of humans, chimpanzees, and rhesus monkeys. *Science*, *192*(4237), 387-389.
Yu, C., Zhou, Y., Liu, Y., Jiang, T., Dong, H., Zhang, Y., & Walter, M. (2011). Functional segregation of the human cingulate cortex is confirmed by functional connectivity based neuroanatomical parcellation. *Neuroimage*, *54*(4), 2571-2581.
Yuan, G., Liu, G., Wei, D., Wang, G., Li, Q., Qi, M., & Wu, S. (2018). Functional connectivity corresponding to the tonotopic differentiation of the human auditory cortex. *Human brain mapping*, *39*(5), 2224-2234.
Zachlod, D., Rüttgers, B., Bludau, S., Mohlberg, H., Langner, R., Zilles, K., & Amunts, K. (2020). Four new cytoarchitectonic areas surrounding the primary and early auditory cortex in human brains. *Cortex*.
Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087-1104.
Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron*, *47*(1), 9-12.
Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is' where'in the human auditory cortex?. *Nature neuroscience*, *5*(9), 905-909.
Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, *14*(4), 1908-1919.
Zatorre, R. J., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *484*, 297.
Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *Journal of cognitive neuroscience*, *8*(1), 29-46.
Zelnik-Manor, L., & Perona, P. (2005). Self-tuning spectral clustering. In *Advances in neural information processing systems* (pp. 1601-1608).

Zetzsche, T., Meisenzahl, E. M., Preuss, U. W., Holder, J. J., Kathmann, N., Leinsinger, G., ... & Möller, H. J. (2001). In-vivo analysis of the human planum temporale (PT): does the definition of PT borders influence the results with regard to cerebral asymmetry and correlation with handedness?. *Psychiatry Research: Neuroimaging*, *107*(2), 99-115.

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2018). The Impact of inserting an additional mental process. *Computational Brain & Behavior*, *1*(1), 22-35.

Zhang, Y., Fan, L., Zhang, Y., Wang, J., Zhu, M., Zhang, Y., ... & Jiang, T. (2014). Connectivity-based parcellation of the human posteromedial cortex. *Cerebral cortex*, *24*(3), 719-727.

Zhao, Q., Hautamaki, V., & Fränti, P. (2008, October). Knee point detection in BIC for detecting the number of clusters. In *International conference on advanced concepts for intelligent vision systems* (pp. 664-673). Springer, Berlin, Heidelberg.

Zhao, Q., Xu, M., & Fränti, P. (2008, November). Knee point detection on bayesian information criterion. In *2008 20th IEEE International Conference on Tools with Artificial Intelligence*(Vol. 2, pp. 431-438). IEEE.

Zündorf, I. C., Lewald, J., & Karnath, H. O. (2013). Neural correlates of sound localization in complex acoustic environments. *PLoS One*, *8*(5), e64259.

Zwaag, W., Gentile, G., Gruetter, R., Spierer, L., & Clarke, S. (2011). Where sound position influences sound object representations: a 7-T fMRI study. *Neuroimage*, *54*(3), 1803-1881.