

UCLA

UCLA Electronic Theses and Dissertations

Title

Intracranial electrophysiology of mentalizing

Permalink

<https://escholarship.org/uc/item/571126q4>

ISBN

9798193223283

Author

Tan, Kevin Maluenda

Publication Date

2026-06-12

Supplemental Material

<https://escholarship.org/uc/item/571126q4#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Intracranial electrophysiology of mentalizing

Probing the cognitive, computational, and systems architecture of the social brain

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Psychology

by

Kevin Maluenda Tan

2026

ABSTRACT OF THE DISSERTATION

Intracranial electrophysiology of mentalizing

by

Kevin Maluenda Tan

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2026

Professor Matthew D. Lieberman, Chair

Functional magnetic resonance imaging (fMRI) has built the prevailing picture of the social brain. However, fMRI's hemodynamic basis collapses two dimensions of neural activity that electrophysiology preserves: the millisecond timing of activity and its distribution across the frequency spectrum. This dissertation uses human intracranial electrophysiology (iEEG; $N=34$) to recover those dimensions for mentalizing – reasoning about mental states – and to reconcile the resulting picture with the fMRI literature it appears at points to contradict.

Paper I, using high-frequency broadband (HFB) activity as a proxy for population spiking, shows that the default mode network's apparent simultaneity during mentalizing is in fact a posterior-to-anterior sequence, from visual cortex through temporoparietal DMN to medial prefrontal cortex. Self- and other-mentalizing engage near-identical populations distinguished by the timing – not the magnitude – of their activity, recasting the "other-selective" dmPFC of fMRI as a difference in duration mistaken for a difference in intensity.

Paper II decomposes this activity across the full spectrum, revealing a dual-phase

architecture shared near-identically by mentalizing and autobiographical memory and jointly distinct from arithmetic. Arithmetic is the spectral outlier: it suppresses HFB in dmPFC while driving sustained theta in the same sites – a signature consistent with frontal-midline theta, the canonical correlate of cognitive control. Because BOLD couples to HFB and is largely blind to low-frequency power, the medial-prefrontal "deactivation" reported during cognitive tasks is the HFB-suppression half of a dissociation whose low-frequency half moves oppositely; the deactivation is faithfully measured but does not amount to disengagement.

Paper III uses spectral cross- classification to show that mentalizing recruits autobiographical- like computation along the same hierarchy, graded (Autobiographical > Self $\approx$ Other > Cognitive) and short of full autobiographical engagement, with self and other differing in timing rather than content.

Together, the papers show that what fMRI represents as one activation is a sequence, a spectrum, and a computation – dimensions that, as paper IV argues, source-localized EEG and MEG can recover non-invasively and at scale. Restoring them does not overturn the fMRI view of the social brain; it recovers what the map leaves out and, in one case, reverses the sign of the inference the map supports.

The dissertation of Kevin Maluenda Tan is approved.

Carolyn M. Parkinson

Ajay Satpute

Lucina Qazi Uddin

Matthew D. Lieberman, Committee Chair

University of California, Los Angeles

2026

Dedication

I would like to thank my family and friends for their warm and unyielding support throughout my graduate studies. My path through the PhD was not always graceful or straightforward. There were periods of financial hardship, mental and emotional exhaustion, and life circumstances that made continuing feel nearly impossible. In those moments, my family and friends gave me a place to land, reasons to keep going, and the love and patience I needed to find my footing again. Without them, I doubt I could have finished this PhD at all.

I also want to acknowledge how deeply impressed I am by the tenacity, resolve, and moral courage of UCLA students. My time at UCLA has been marked by profound upheaval, both globally and within our own campus community. Through it all, I have seen students meet uncertainty with care, conviction, and an unwavering commitment to one another and to the dignity of people they may never know.

Finally, I would like to thank my advisors and committee members, along with the faculty, teachers, and mentors who supported my development as a scientist throughout my educational journey. Their guidance, feedback, and advocacy helped shape not only this dissertation, but also the foundational knowledge, research projects, and collaborations that brought me to this point.

Contents

ABSTRACT OF THE DISSERTATION	ii
Dedication	v
List of Figures	ix
Acknowledgements	x
Vita	xi
Introduction	1
Paper I: A common neurocognitive sequence for mentalizing about the self and others.....	3
Abstract	3
Introduction	4
Results	7
Data and design	7
Mentalizing activations propagate from visual cortex to mPFC	11
A spatiotemporal gradient of mentalizing-specificity from visual cortex to mPFC	12
Behavioral responses are best predicted by TPJ and dmPFC activity	14
Self- and other-mentalizing share a common neuroanatomical basis	16
Other-mentalizing evokes slower and lengthier activations across successive ROIs	17
Summary of evoked neuronal activity	20
Controlling for stimulus visual dissimilarity (VD)	20
Behavioral results.....	21
Discussion	21
Methods.....	30
Participants	30
Behavioral Task	31
Electrocorticography data acquisition.....	32
Defining regions of interest (ROIs) and brain networks	33
ECoG preprocessing	34
Statistics	34
Within-site analyses	35
Trial-averaged analysis (within-site)	36
Single-trial analysis (within-site)	36
Paper II: A spectrum of neurocognitive processing for mentalizing	41
Abstract	41
Introduction	42

Results	44
Data and Design	44
Mentalizing and autobiographical memory share a dual-phase architecture	45
Mentalizing and autobiographical memory diverge from arithmetic but not each other	47
Self- and other-mentalizing evoke near-identical spatiotemporospectral dynamics	49
Whole-brain spectral dynamics recapitulate the dual-phase architecture	52
Discussion	52
Methods	59
Participants	59
Behavioral Task	60
iEEG data acquisition	61
Defining regions of interest (ROIs)	61
iEEG preprocessing	63
Statistics	64
Within-site analyses	65
ROI condition contrasts	65
Whole-brain analysis	66
Paper III: Autobiographical computations in mentalizing	67
Abstract	67
Introduction	68
Results	70
Data and design	70
ROI classifier accuracy	72
Psychobehavioral relevance	79
Whole-brain results	81
Discussion	83
Methods	88
Participants	89
Behavioral Task	89
iEEG data acquisition	90
Defining regions of interest (ROIs)	91
iEEG preprocessing	92
Cross-classification	93
Analyses of posterior probability	94
Paper IV: Probing blind spots in social neuroscience with source-space EEG and MEG	96

Abstract	96
Introduction	97
Signals and theory	100
Distributed hierarchical processing	100
Physiology	102
Spectral dynamics.....	103
Spatiotemporal dynamics.....	107
Questions for source-space M/EEG	108
Methods and physics.....	110
Equipment and data collection	111
Source localization.....	114
Implementation	119
Task design.....	120
Data collection.....	120
Early-stage preprocessing	121
Denoising	121
Source and spectral decomposition.....	122
Whole-brain univariate analysis	125
ROI univariate analyses	125
ROI effective connectivity.....	126
Multivariate analyses.....	127
Conclusion.....	129
General Discussion	130
One activation, three dimensions	130
A parallel distributed architecture.....	131
Two axes for self and other.....	132
Architecture and behavior	132
From apparent contradiction to reconciliation	133
A methodological argument.....	134
Beyond mentalizing	134
What the dimensions reveal.....	135
Supplemental Materials	136
Supplemental Figure 3.1.....	136
References	137

List of Figures

Figure 1.1. Data and design	7
Figure 1.2. Exemplar ROI sites	7
Figure 1.3. A neurocognitive sequence for mentalizing	8
Figure 1.4. Spatiotemporal gradients of functional and behavioral relevance for mentalizing	11
Figure 1.5. A common neurocognitive sequence for self- and other-mentalizing	16
Figure 1.6. Summary of evoked neuronal activity	20
Figure 2.1. Data and design	42
Figure 2.2. Event-related spectral perturbations (ERSPs).....	46
Figure 2.3. ERSP contrasts between task conditions	47
Figure 2.4. ERSPs for self- and other-mentalizing.....	49
Figure 2.5. Whole-brain maps from single-site analyses.....	52
Figure 3.1. Graphical abstract of the analytical design.....	71
Figure 3.2. ROI classification accuracy for autobiographical versus cognitive task conditions...	72
Figure 3.3. ROI posterior probabilities for each task condition.....	75
Figure 3.4. Psychobehavioral relevance of ROI PPs.....	79
Figure 3.5. Whole-brain results	81
Figure 4.1. Neuronal signals	101
Figure 4.2. EEG and MEG equipment	108
Figure 4.3. Source localization.....	112
Figure 4.4. Why/How task.....	118
Figure 4.5. Univariate analyses.....	125
Figure 4.6. Multivariate analyses.....	127

Acknowledgements

We are grateful for public financial support from the United States and Canada. This work was supported by National Science Foundation Graduate Research Fellowship DGE-1650604, Department of Defense Grant 13RSA281, National Institute of Neurological Disorders and Stroke Grant R01NS078396, National Institute of Mental Health Grant 1R01MH109954-01, National Science Foundation Grant BCS1358907, Department of Defense Grant 13RSA281, National Institute of Child Health and Human Development Postdoctoral Fellowship F32HD087028, Stanford University School of Medicine Medical Scholars Research Fellowship, Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship.

We thank the Laboratory of Behavioral and Cognitive Neuroscience for generously providing access and support for these data. We thank members of the Social Cognitive Neuroscience Lab and Computational Social Neuroscience Lab for their continued support.

Paper I in this dissertation has been published as a peer-reviewed article and is presented here in slightly edited format: Tan, K.M., Daitch, A.L., Pinheiro-Chagas, P. et al. Electrocorticographic evidence of a common neurocognitive sequence for mentalizing about the self and others. *Nat Commun* 13, 1919 (2022).

<https://doi.org/10.1038/s41467-022-29510-2>. For Paper I, J.P. and A.L.D. contributed to experimental design and data acquisition; J.P., A.L.D., P.P., and K.M.T. developed preprocessing and analysis tools; K.M.T. and M.D.L. conceptualized the analysis plan; K.M.T. performed formal data analysis; and all authors contributed to writing of the manuscript. Papers II, III, and IV are in preparation for publication with co-authors. Moreover, many figures in Paper IV are adapted from other publications, cited in each figure caption, and are used here with permission.

Vita

EDUCATION

University of California, Los Angeles

2018 – Present

Social and Affective Neuroscience

M.A., Computational Cognition

Advisors: Matthew D. Lieberman, Carolyn M. Parkinson

Carnegie Mellon University

2010 – 2014

B.S., Cognitive Neuroscience

Minors: Computational Neuroscience, Global Studies

Advisors: Michael J. Tarr, Greg J. Siegle

AWARDS & HONORS

Graduate Research Fellowship (DGE-1650604) • National Science Foundation

2019 – 2023

Harold H. Kelley Award • UCLA

2023

Editorial Spotlight • *Nature*

2022

Refereed Spotlight • *Trends in Cognitive Sciences*

2022

Graduate Summer Research Mentorship • UCLA

2019

Advanced Science Fellowship • São Paulo Research Foundation

2018

Sigma Xi Best Poster Award • Carnegie Mellon

2014

Dozzi Research Fellowship • Carnegie Mellon

2012

Undergraduate Research Grant • Carnegie Mellon

2010 – 2014

PEER-REVIEWED ARTICLES

See [Google Scholar](#) and [ORCID](#) for latest articles and metrics.

Tan KM, Daitch AL, Pinheiro-Chagas P, Fox KCR, Parvizi J, Lieberman MD. (2022). Electrocorticographic evidence of a common neurocognitive sequence for mentalizing about the self and others. *Nature Communications*. 13, 1919. doi.org/10.1038/s41467-022-29510-2

Straccia MA, Teed AR, Katzman PL, **Tan KM**, Parrish MH, Irwin MR, Eisenberger NI, Lieberman MD, Tabak BA. (2021). Null results of oxytocin and vasopressin administration on mentalizing in a large fMRI sample: Evidence from a randomized controlled trial. *Psychological Medicine*, 1-11. doi.org/10.1017/S0033291721004104

Tan KM, Burklund LJ, Craske MG, Lieberman MD. Posttraumatic stress disorder and the social brain: Affect-related disruption of the default and mirror networks. (2019). *Depression and anxiety*. 36(11):1058-71. doi.org/10.1002/da.22953

Moieni M, **Tan KM**, Inagaki TK, Muscatell KA, Dutcher JM, Jevtic I, Breen EC, Irwin MR, Eisenberger NI. Sex differences in the relationship between inflammation and reward sensitivity: a randomized controlled trial of endotoxin. (2019). *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2019 Jul; 4(7):619-26. doi.org/10.1016/j.bpsc.2019.03.010

Lieberman MD, Straccia MA, Meyer ML, Du M, **Tan KM**. Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. (2019). *Neuroscience & Biobehavioral Reviews*. 99, 311-28. doi.org/10.1016/j.neubiorev.2018.12.021

INVITED TALKS

Tan KM, Daitch AL, Pinheiro-Chagas P, Fox KCR, Parvizi J, Lieberman MD. (2021, Sep). *A common cortical hierarchy for mentalizing about the self and others*. Annual symposium, Brain Research Institute, UCLA.

Tan KM, Daitch AL, Fox KCR, Parvizi J, Lieberman MD. (2019, May). *Intracranial electrophysiology of mentalizing*. Annual meeting, Social and Affective Neuroscience Society, Miami Beach, USA. [\[YouTube\]](#)

CONFERENCE PRESENTATIONS

Tan KM, Daitch AL, Fox KCR, Parvizi J, Lieberman MD. (2019, May). *Intracranial electrophysiology of mentalizing*. Annual meeting, Social and Affective Neuroscience Society, Miami Beach, USA. [\[PDF\]](#)

Straccia MA, Teed AR, Katzman PL, **Tan KM**, Parrish MH, Irwin MR, Eisenberger NI, Lieberman MD, Tabak BA. (2019, May). *Null results of oxytocin and vasopressin administration on mentalizing in a large fMRI sample*. Annual meeting, Social and Affective Neuroscience Society, Miami Beach, USA.

Tan KM, Burklund LJ, Craske MG, Lieberman MD. (2018, May). *PTSD and the social brain: affect-related disruption of the default and mirror networks*. Annual meeting, Social and Affective Neuroscience Society, New York City, USA. [\[Poster PDF\]](#)

Du M, Straccia MA, **Tan KM**, Lieberman MD. (2018, May). *Meta-analysis of medial prefrontal cortex with NS+*. Annual meeting, Social and Affective Neuroscience Society, New York City, USA.

Ames DL, Gillespie GSR, Burns SM, **Tan KM**, Lieberman MD. (2018, May). *Predicting the effectiveness of health messaging through fMRI and fNIRS*. Annual meeting, Social and Affective Neuroscience Society, New York City, USA.

Cooper-White MA, Ames DL, Burns SM, **Tan KM**, Gillespie GSR, Lieberman MD. (2018, May). *Investigating the neural and cognitive mechanisms behind 'latitudes of acceptance' for the opinions of others*. Annual meeting, Social and Affective Neuroscience Society, New York City, USA.

Tan KM, Tarr MJ. (2016, April). *Valence-related visual encoding is dependent on valence strength prior to frontal affective processing*. Annual meeting, Social and Affective Neuroscience Society, New York City, USA. [\[Poster PDF\]](#)

Tan KM, Kraynak T, Siegle GJ. (2014, May). *Toward a pathophysiology of rumination: dorsal nexus resting-state functional connectivity in depression before and after serotonergic or behavioral interventions*. Annual symposium, Center for the Neural Basis of Cognition, University of Pittsburgh & Carnegie Mellon University. [\[Poster PDF\]](#)

Tan KM, Ferris L, Lee HS. (2013, May). *Effect of room temperature on interpersonal attraction*. Annual symposium, Dietrich College, Carnegie Mellon University

Introduction

A social neuroscientist could be forgiven for believing that cognitive abilities arise from static brain activations like those depicted in functional anatomy maps. Throughout its history, social neuroscience has overwhelmingly relied on functional magnetic resonance imaging (fMRI), a neuroimaging modality with high anatomical resolution but low temporal resolution (Lieberman, 2013). In fMRI, brain regions appear to respond simultaneously on a one-dimensional scale of ‘activation’. However, electrophysiology shows that brain regions respond through intricate spatiotemporal sequences with a multi-dimensional spectrum of activity (Felleman & Van Essen, 1991; Ward, 2003; Fox et al., 2018). As such, overreliance on fMRI leaves our field blind to critical neurocognitive dynamics instantiated by a spectrum of activity at millisecond scales, obscuring invaluable systems and computational mechanisms (Tan et al., 2022; Tan & Tarr, 2016).

We therefore introduce invasive and noninvasive electrophysiological modalities that provide millisecond temporal resolution and up to laminar-scale anatomical resolution (i.e., cortical layer discrimination). We aim to show that electrophysiology enables incisive neuromechanistic accounts of social cognitive functions – accounts that are lacking yet sorely needed in our field. To this end, our empirical analyses use invasive electrophysiology to maximize fidelity, serving as a foundation for a primer on noninvasive electrophysiology for social neuroscientists. Specifically, we began by establishing a common neurocognitive sequence for mentalizing about the self and others (published in [Tan et al., 2022](#)). We build upon this work by characterizing the spectrum of electrophysiological activity during mentalizing, which we then use to characterize the shared neurocomputation between mentalizing and autobiographical memory. To broaden the accessibility and scope of

electrophysiology in social neuroscience, we conclude with a practical and theoretical primer on noninvasive source-localized electrophysiology

Paper I: A common neurocognitive sequence for mentalizing about the self and others

Abstract

Neuroimaging studies of mentalizing (i.e., theory of mind) consistently implicate the default mode network (DMN). Nevertheless, the social cognitive functions of individual DMN regions remain unclear, perhaps due to limited spatiotemporal resolution in neuroimaging. Here we use electrocorticography (ECoG) to directly record neuronal population activity while 16 human participants judge the psychological traits of themselves and others. Self- and other-mentalizing recruit near-identical cortical sites in a common spatiotemporal sequence. Activations begin in the visual cortex, followed by temporoparietal DMN regions, then finally in medial prefrontal regions. Moreover, regions with later activations exhibit stronger functional specificity for mentalizing, stronger associations with behavioral responses, and stronger self/other differentiation. Specifically, other-mentalizing evokes slower and longer activations than self-mentalizing across successive DMN regions, implying lengthier processing at higher levels of representation. Our results suggest a common neurocognitive pathway for self- and other-mentalizing that follows a complex spatiotemporal gradient of functional specialization across DMN and beyond.

Introduction

Hundreds of fMRI studies have shown that mentalizing recruits default mode network (DMN) regions – including temporoparietal junction (TPJ), posteromedial cortex (PMC), and medial prefrontal cortex (mPFC) – with remarkable consistency across countless mentalizing paradigms instantiated in various sensory modalities (Frith & Frith, 2006; Van Overwalle & Baetens, 2009; Mars et al., 2012; Amft et al., 2015; Molenberghs et al., 2016; Schurz et al., 2021). Nevertheless, the specific social cognitive functions of individual DMN regions remain unclear. When seen through fMRI, DMN regions appear to respond concurrently, yet electrophysiological studies demonstrate that critical neurocognitive dynamics occur at millisecond timescales throughout DMN (Fox et al., 2018). Thus, the limited temporal resolution of fMRI may preclude more precise neurocognitive accounts of mentalizing and its component processes.

Several studies have investigated the fast spatiotemporal dynamics of mentalizing processing using source-space electroencephalography (EEG) and magnetoencephalography (MEG), neuroimaging modalities with millisecond temporal resolution but coarse spatial resolution (Schoffelen & Gross, 2009). These studies reveal a general spatiotemporal sequence of cortical recruitment, starting in visual cortex, followed by mirror neuron system regions (MNS; e.g., intraparietal sulcus and premotor cortex), then lastly in DMN regions (Vistoli et al., 2011; Van Duynslaeger et al., 2007; Mossad et al., 2016; H. Wang et al., 2016; Yuk et al., 2018; L. Zhang et al., 2018). These findings exemplify the emerging consensus that visual representations are used by MNS to identify observable actions (e.g., grasping for food), which are then used by DMN to infer unobservable mental states (e.g., hunger) (Van Overwalle & Baetens, 2009; Spunt & Lieberman, 2012; Barrett & Satpute, 2013; Catmur, 2015; Tan et al., 2019). Taken together, EEG/MEG studies of mentalizing suggest that visual cortex, MNS, and DMN act as a hierarchical neurocognitive pathway that transforms low-level visual inputs into high-level mentalistic inferences. However, despite broad agreement at the network level, these studies report

inconsistent recruitment across individual DMN regions. These inconsistencies may reflect limitations in EEG/MEG source localization, particularly in deeper regions such as mPFC and PMC (Akalin Acar & Makeig, 2013), which were not sampled in many of these studies. As such, the sequence of mentalizing processing across individual DMN regions remains unclear.

We sought a more spatiotemporally precise and mechanistic account of mentalizing by exploring neuronal population activity across individual DMN regions and beyond. Leveraging the benefits of human intracranial electrophysiology (Parvizi & Kastner, 2018), we recorded high-frequency broadband activity (HFB; 70-180 Hz), which reflects the rapid aggregate spiking of neuronal populations (Mukamel et al., 2005). In contrast, fMRI measures slow metabolic changes, although fMRI and HFB correspond in the anatomy and direction of measured effects (see Parvizi & Kastner, 2018). We show that self- and other-mentalizing share a complex spatiotemporal gradient of functional specialization at millisecond, millimeter, and cross-regional scales. Our findings demonstrate that high spatiotemporal resolution methods can provide critical insights into the neurocognitive mechanisms of human social cognition.

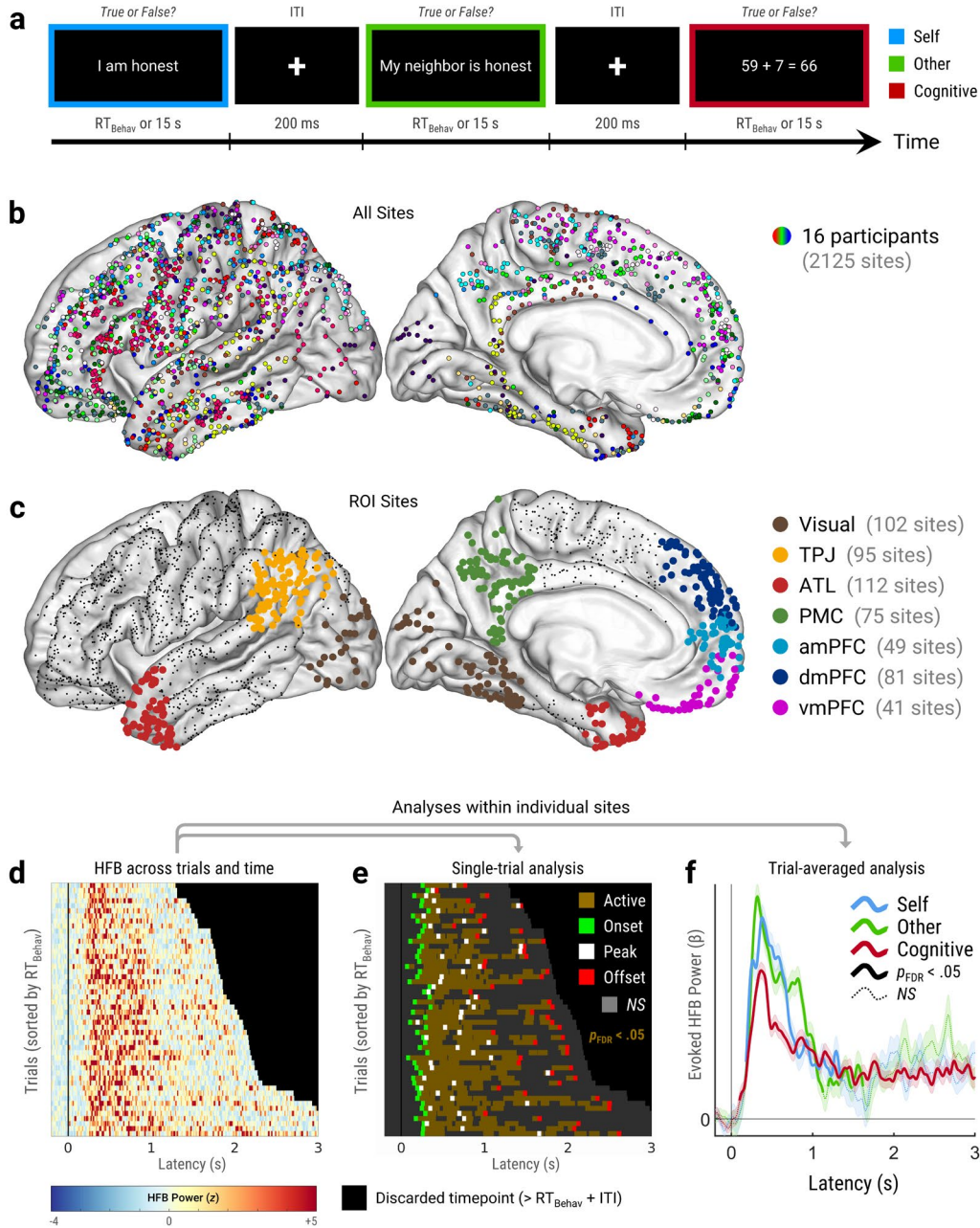


Figure 1.1: Data and design. All brain maps in main figures plot each site in left-hemisphere MNI space for display purposes. **a)** Behavioral task. **b)** Map of electrode sites colored by participant. **c)** Region-of-interest sites (ROI; colored circles) and non-ROI sites (black dots). ROIs were defined using each participant's native cortical surface. **d-f)** Within-site analytic pipeline using a mid-cingulate exemplar site. **d)** High-frequency broadband power (HFB; 70-180 Hz) across timepoints and trials. Black areas indicate discarded timepoints (post-trial). **e)** Single-trial analysis compared evoked HFB power versus pre-stimulus baseline (Welch's tests) to provide five key HFB metrics. Duration: total timepoints with significant activations (brown areas; $p_{FDR} < .05$, corrected across timepoints, trials, and sites). Onset Latency: earliest timepoint with significant activation (green squares). Peak Latency and Peak Power: timepoint and magnitude, respectively, of the strongest activation (white squares). Offset Latency: latest timepoint with significant activation (red squares). Gray areas indicate nonsignificant (NS) activations. **f)** Trial-averaged analysis identified sites that were active, inactive, or nonresponsive for each task condition. Sites were considered 'active' or 'inactive' if evoked HFB power was significantly higher or lower, respectively, than pre-stimulus baseline ($p_{FDR} < .05$, corrected across timepoints and sites). Colored lines show mean timecourses of evoked HFB power (β) estimated by linear mixed-effects modeling (nested within Trial). Thick solid lines indicate significant responses. Thin dashed lines indicate nonsignificant responses. Shaded areas indicate standard error of β .

Results

Data and design

We recruited sixteen human participants who had electrocorticography (ECoG) electrodes surgically implanted onto the cortical surface for epilepsy monitoring and treatment (Supplementary Table 1.2). Recordings were obtained from all 2125 electrode sites in our cohort (Fig. 1.1b). Our behavioral task (Fig. 1.1a) consisted of true/false text prompts for three conditions of interest: self-mentalizing (e.g., “I am honest”), other-mentalizing (e.g., “My neighbor is honest”), and a non-social ‘cognitive’ task involving simple arithmetic (e.g., “9+86=95”; Supplementary Table 1.4). Sites and trials underwent exclusion criteria for epileptic activity, noise, poor behavioral performance (trials only), and statistical outliers (see Methods).

We began by parcellating the brain into seven regions-of-interests (ROIs; Fig. 1.1c) using each participant’s native-space cortical surface (Supplementary Fig. 1.2). We included six DMN ROIs that are strongly implicated in mentalizing (Schurz et al., 2021): temporoparietal junction (TPJ), anterior temporal lobe (ATL), posteromedial cortex (PMC), anteromedial prefrontal cortex (amPFC), dorsomedial prefrontal cortex (dmPFC), and ventromedial prefrontal cortex (vmPFC). Visual cortex was included as a control ROI. Of 2125 electrode sites, 555 were included in our ROIs.

We examined HFB activity from each site in two ways: single-trial and trial-averaged analyses (Fig. 1.1def). Single-trial analysis compared trialwise HFB responses relative to the pre-stimulus baseline preceding each trial ($p_{\text{FDR}} < .05$, corrected across timepoints, trials, and sites). Single-trial analysis captured five key metrics of HFB activations: onset, peak, and offset latencies, duration, and peak power (defined in Fig. 1.1e). Trial-averaged analysis used linear mixed-effects models (LMEMs) to estimate mean timecourses of task-evoked HFB power relative to pre-stimulus baseline (Fig. 1.1f). Trial-averaged analysis identified sites with significant activations or deactivations for each task condition ($p_{\text{FDR}} < .05$, corrected across

timepoints and sites). See Fig. 1.2 for within-site analyses of exemplar ROI sites. All statistical tests herein were two-tailed.

Functional specificity for mentalizing strengthens from visual cortex to mPFC

To explore the fast spatiotemporal dynamics of mentalizing processing, we first examined its functional-anatomic foundations. We began by identifying sites with significant HFB responses to either self- or other-mentalizing, regardless of functional specificity or selectivity (Fig. 1.3ab). Using trial-averaged results, sites were considered ‘mentalizing-active’ (cyan) or ‘mentalizing-deactive’ (orange) if they produced higher or lower HFB power, respectively, relative to pre-stimulus baseline ($p_{\text{FDR}} < .05$). Sites were considered ‘mentalizing-nonresponsive’ if they produced nonsignificant HFB responses to mentalizing. We found mentalizing-active sites in nearly all parts of cortex, while mentalizing-deactive sites were rarer and generally located outside DMN (Fig. 1.3a). Overall, most cortical sites were mentalizing-nonresponsive (55% whole-brain; Fig. 1.3b).

Next, we examined the functional specificity of mentalizing-active sites (Fig. 1.3cd). Using trial-averaged results, we identified which mentalizing-active sites also produced significant HFB responses to arithmetic (cognitive task) relative to pre-stimulus baseline ($p_{\text{FDR}} < .05$). Additionally, we directly compared single-trial peak power across mentalizing and arithmetic using robust regression ($p_{\text{FDR}} < .05$, corrected across sites). Sites were considered ‘mentalizing-specific’ (light and dark turquoise) if they were 1) mentalizing-active but not arithmetic-active, and 2) produced significantly higher peak power for mentalizing over arithmetic. Sites were considered ‘non-specific’ (lime, pink, and red) if they coactivated for mentalizing and arithmetic, regardless of peak power differences.

Overall, most mentalizing-active sites were non-specific (68% whole-brain), while the remaining mentalizing-specific sites were unevenly distributed across cortex (Fig. 1.3cd). Within

ROIs, visual cortex (1%) showed the least mentalizing-specificity. Intermediate mentalizing-specificity was found in TPJ (38%), ATL (32%), and PMC (58%). Very high mentalizing-specificity was found in amPFC (87%), dmPFC (94%), and vmPFC (100%). Altogether, these results show a gradient of mentalizing-specificity from visual cortex to mPFC.

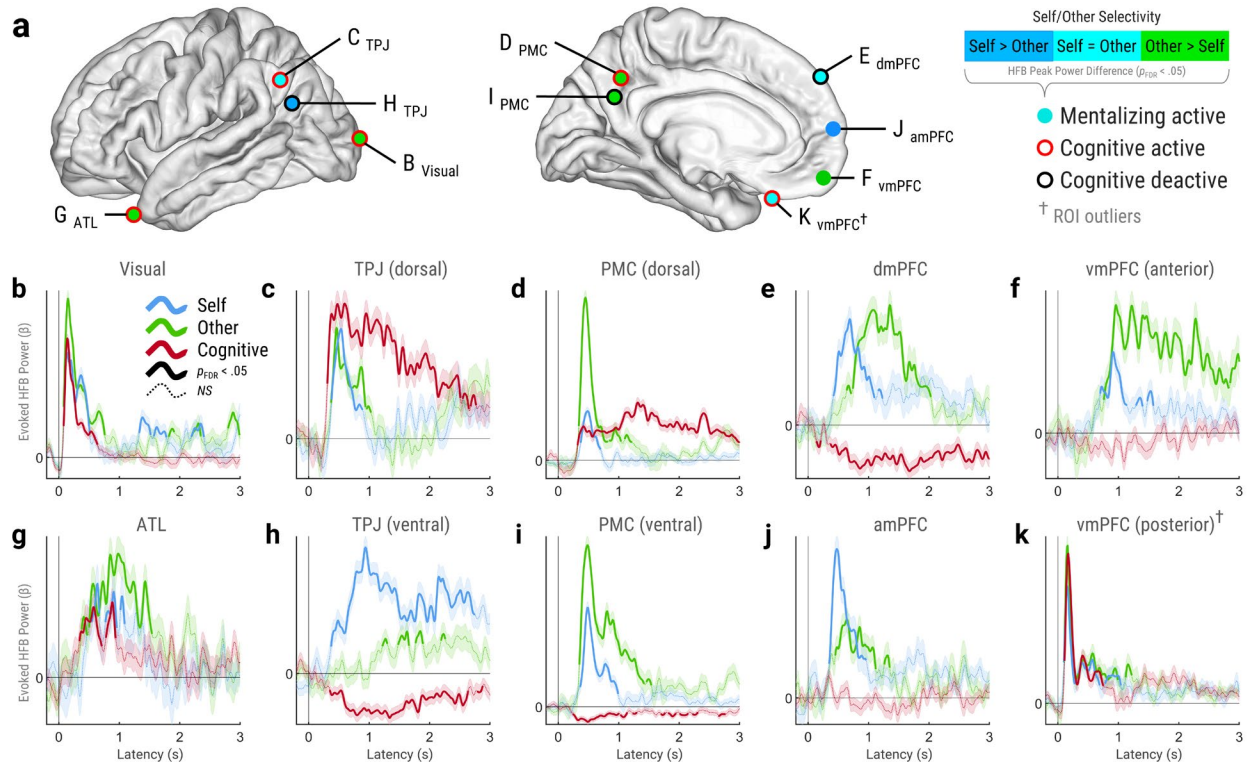


Figure 1.2: Exemplar ROI sites. **a)** Map of exemplar ROI sites, which were identified as mentalizing-active versus baseline in trial-averaged analysis (see Fig. 1f; $p_{FDR} < .05$, corrected across timepoints and sites). Circle fill color indicates self/other selectivity, which was determined by comparing single-trial HFB peak power (see Fig. 1e) across mentalizing type via robust regression ($p_{FDR} < .05$, corrected across sites). Circle outline indicates significant HFB responses to the cognitive (arithmetic) task versus baseline ($p_{FDR} < .05$, corrected across timepoints and sites). **b-k)** Trial-averaged timecourses of evoked HFB power (β) from exemplar sites in panel **a**. Thick solid lines indicate significant responses versus baseline ($p_{FDR} < .05$, corrected across timepoints and sites). Thin dashed lines indicate nonsignificant (NS) responses. Shaded areas indicate standard error of β . [†]Excluded from ROI analyses due to outlier thresholds (see Methods). *Abbreviations:* HFB = high-frequency broadband (70-180 Hz). ROI = region-of-interest, Visual = visual cortex, ATL = anterior temporal lobe, TPJ = temporoparietal junction, PMC = posteromedial cortex, amPFC = anteromedial prefrontal cortex, dmPFC = dorsomedial prefrontal cortex, vmPFC = ventromedial prefrontal cortex. Source data are provided as a Source Data file.

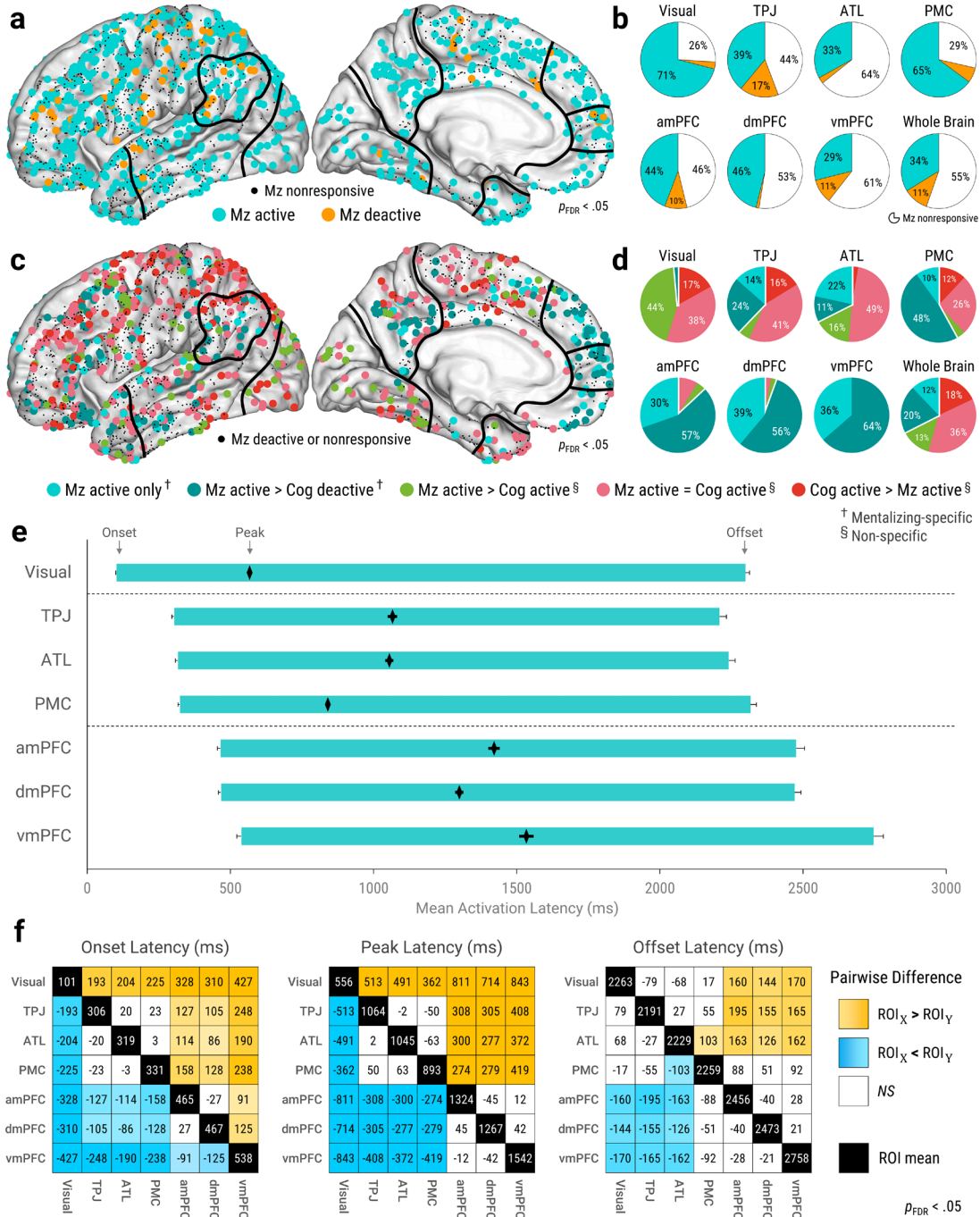


Figure 1.3: A neurocognitive sequence for mentalizing. Brain maps plot all sites on left hemisphere with approximate ROI outlines **ab)** Sites identified as active, deactive, or nonresponsive for mentalizing versus baseline (see Figs. 1f & 2; $p_{FDR} < .05$, corrected across timepoints and sites). **cd)** Functional specificity of mentalizing-active sites. [†]Mentalizing-specific sites, defined as mentalizing-active but not cognitive-active ($p_{FDR} < .05$, corrected across timepoints and sites) with significantly higher peak power for mentalizing ($p_{FDR} < .05$, corrected across sites). [§]Non-specific sites, defined as mentalizing-active and cognitive-active. **e)** Mean ROI activation latencies. Left and right floating bar edges depict onsets and offsets, respectively, while diamonds depict peaks (see Fig. 1e). Error bars depict standard error of the mean. **f)** Pairwise ROI comparisons of activation latencies. Black diagonal squares show ROI means. Off-diagonal squares show estimates for $ROI_X - ROI_Y$ (controlled for behavioral response times). Orange and blue squares indicate significant differences; color intensity indicates effect size ($p_{FDR} < .05$, corrected across unique ROI pairs). White squares indicate nonsignificant (NS) differences. Pairwise comparisons used linear mixed-effect models restricted to participants with mentalizing-active sites in both ROIs (nested within Trial within Participant).

Mentalizing activations propagate from visual cortex to mPFC

Next, we explored the timing of mentalizing-evoked activations across ROIs. To this end, we analyzed single-trial HFB latency metrics from mentalizing-active ROI sites (Fig. 1.3e). Pairwise ROI comparisons (Fig. 1.3f) measured trial-by-trial ROI differences while controlling for behavioral response times (RT_{Behav}) and participant-related heterogeneity using LMEMs (nested within Trial within Participant). Critically, pairwise comparisons only included participants with mentalizing-active sites in both ROIs.

We found that visual cortex produced the earliest activation onsets ($M=101\pm 1$ ms) of any ROI (Fig. 1.3f; $p_{\text{FDR}} < .05$, corrected across unique ROI pairs). Afterwards, mid-latency onsets were produced by TPJ ($M=303\pm 7$ ms), ATL ($M=316\pm 7$ ms), and PMC ($M=322\pm 5$ ms), with nonsignificant differences between them. Later onsets were produced by amPFC ($M=465\pm 10$ ms) and dmPFC ($M=466\pm 7$ ms), with nonsignificant differences between them, followed lastly by vmPFC ($M=537\pm 15$ ms). Critically, propagation of activations across ROIs had robust trial-by-trial consistency (Supplementary Fig. 1.3). Peak latencies showed a similar pattern of cross-ROI differences as onsets, although intra-mPFC differences were nonsignificant (Fig. 1.3f). Offset latencies showed the least cross-ROI differentiation. Earlier offsets were produced by visual cortex, TPJ, and ATL, with nonsignificant differences between them. Later offsets were produced by amPFC, dmPFC, and vmPFC, with nonsignificant differences between them. We also performed post-hoc analysis of offset latencies relative to RT_{Behav} (Fig. 1.4c), revealing that mPFC activations more closely preceded RT_{Behav} than other ROIs combined ($b_{20204}=138\pm 31$ ms, $p=4.28e-6$). Nevertheless, despite these robust cross-ROI latency differences, ROIs predominantly activated at overlapping times (Figs. 1.3e, 1.5c, 1.2 & 1.6; single-participant results in Supplementary Fig. 1.2).

In sum, mentalizing evoked largely concurrent activations across ROIs, as might be expected from neuroimaging literature. Nonetheless, fine-scale cross-ROI differences in onset,

peak, and offset latencies depict an overarching spatiotemporal sequence of activation from visual cortex to mPFC. This sequence extended across individual sites throughout cortex (Fig. 1.4a).

A spatiotemporal gradient of mentalizing-specificity from visual cortex to mPFC

Thus far, we have found spatial gradients in the timing (Figs. 1.3ef) and functional specificity (Fig. 1.3cd) of neuronal population responses to mentalizing. To examine the correspondence between these gradients, we used logistic mixed-effects classification to predict mentalizing-specificity from mean onset latencies across mentalizing-active ROI sites, nested within Participant (Fig. 1.4b). We found that the probability of mentalizing-specificity over non-specificity rose by 1.4% per one-millisecond increase in onsets (odds ratio=1.014, $b_{255}=0.014\pm.002$, $p=9.21e-11$). This robust effect was largely attributable to between-ROI differences, as within-ROI effects were weaker and significant only in TPJ ($p=.047$), ATL ($p=.039$), and PMC ($p=.005$). In sum, ROI sites with later activations were more likely to be mentalizing-specific.

These results portray a neurocognitive sequence (Felleman & Van Essen, 1991; Mormann et al., 2008; Diaconescu et al., 2017; Hilgetag & Goulas, 2020) that reflects both the timing and specificity of mentalizing processing across ROIs. Within this sequence, visual cortex appears to be a pre-mentalizing stage, as it produced the earliest activations and negligible mentalizing-specificity. Next, temporoparietal DMN regions (TPJ, ATL, and PMC) appear to be lower-level mentalizing stages, as they produced mid-latency activations and intermediate mentalizing-specificity. Finally, mPFC regions (amPFC, dmPFC, and vmPFC) appear to be higher-level mentalizing stages, as they featured the latest activations and overwhelming mentalizing-specificity.

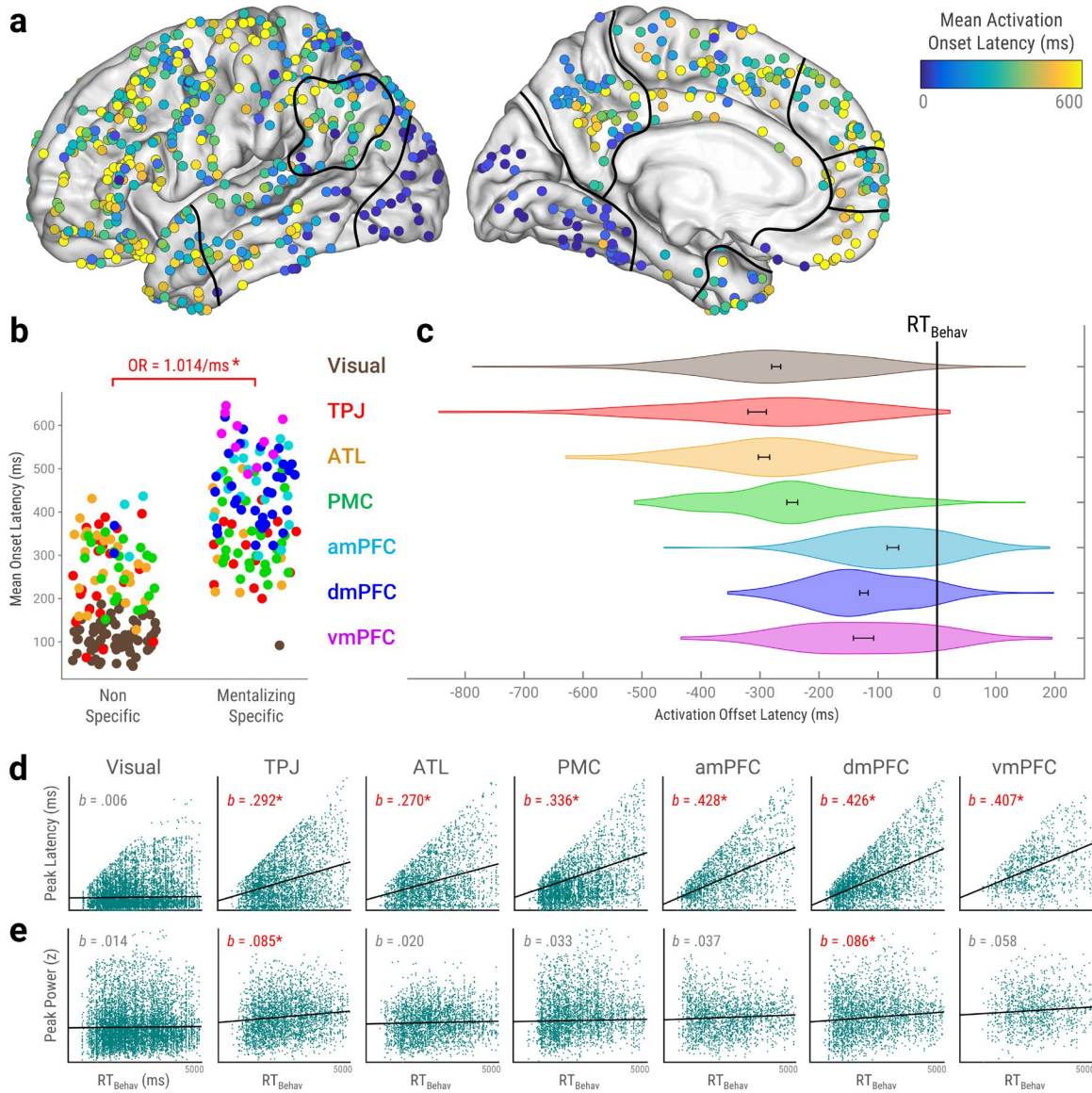


Figure 1.4: Spatiotemporal gradients of functional and behavioral relevance for mentalizing. **a)** Mean onset latencies of mentalizing-active sites ($p_{FDR} < .05$, corrected across timepoints and sites). All sites plotted on left hemisphere with approximate ROI outlines (full views in Supplementary Fig. 8). **b)** Mean activation onset latencies of ROI sites by functional specificity (see Fig. 3c). Color indicates ROI. Logistic mixed classification estimated the odds ratio (OR) of mentalizing-specificity over non-specificity per millisecond increase in onsets ($p = 9.21e-11$, $n = 257$ sites). **c)** RT_{Behav} versus ROI activation offset latencies. Error bar indicates standard error of the mean. Offsets in medial prefrontal cortex (mPFC) ROIs more closely preceded RT_{Behav} than other ROIs combined ($b_{20204} = 138 \pm 31$ ms, $p = 4.28e-6$, two-tailed), as per post-hoc LMEM. **d)** Associations between RT_{Behav} and peak activation metrics within ROIs. Slope (b) estimates were controlled for behavioral response times and stimulus visual dissimilarity via LMEMs. See Table 1 for exact statistics and n . **e)** RT_{Behav} and activation peak latency. Slopes (b) estimate change in peak latency per millisecond increase in RT_{Behav} . Horizontal slopes ($b = 0$) would indicate purely stimulus-locked activity, while 45° slopes ($b = 1$) would indicate purely RT_{Behav} -locked activity (see DiCarlo & Maunsell, 2005¹⁰⁶). **f)** RT_{Behav} and peak power (activation magnitude). Slopes (b) estimate change in peak power per 1000 ms increase in RT_{Behav} . *Significant OR and b estimates have red font ($p < .05$, uncorrected, two-tailed). *Abbreviations:* RT_{Behav} = behavioral response time. ROI = region-of-interest, Visual = visual cortex, ATL = anterior temporal lobe, TPJ = temporoparietal junction, PMC = posteromedial cortex, amPFC = anteromedial prefrontal cortex, dmPFC = dorsomedial prefrontal cortex, vmPFC = ventromedial prefrontal cortex, z = z-score, LMEM = linear mixed-effect model. Source data are provided as a Source Data file.

Behavioral responses are best predicted by TPJ and dmPFC activity

To explore the relationships between neuronal and behavioral responses, we determined which ROIs predicted RT_{Behav} and $\text{Choice}_{\text{Behav}}$ (choosing 'true' or 'false' for task prompts) by analyzing single-trial HFB metrics from mentalizing-active ROI sites. We used LMEMs to simultaneously estimate neuronal associations with RT_{Behav} and $\text{Choice}_{\text{Behav}}$ while controlling for stimulus visual dissimilarity (see Supplementary Methods), nested within Site and Participant. We found that onset latencies had significant ($p < .05$) positive associations with RT_{Behav} in visual cortex, TPJ, PMC, amPFC, and dmPFC (Table 1.1). Moreover, we found significant positive associations between RT_{Behav} and peak latencies (all DMN ROIs; Fig. 1.4d), offset latencies (all ROIs; Table 1.1), and activation duration (all ROIs). In contrast, activation magnitude (peak power) had significant RT_{Behav} associations in only two ROIs: TPJ and dmPFC (Fig. 1.4e). Intriguingly, dmPFC was the only ROI that significantly predicted $\text{Choice}_{\text{Behav}}$ (Table 1.1). Taken together, behavioral responses were best predicted by TPJ and dmPFC, which had more numerous significant neurobehavioral associations than other ROIs.

Next, we examined whether behavioral responses were better predicted by activity in successive ROI sites. Specifically, we compared mean onset latencies (Fig. 1.4a) and within-site RT_{Behav} random effect sizes (b from analyses in previous paragraph) using LMEMs nested within Participant. We found that sites with later onsets had stronger RT_{Behav} effects for peak latency ($b_{255} = 0.972 \pm .094$, $p = 1.51 \times 10^{-23}$) but not peak power ($b_{255} = 0.019 \pm .074$, $p = .796$). In sum, RT_{Behav} was better predicted by successive ROIs in terms of activation latency (Fig. 1.4d) but not activation magnitude (Fig. 1.4e).

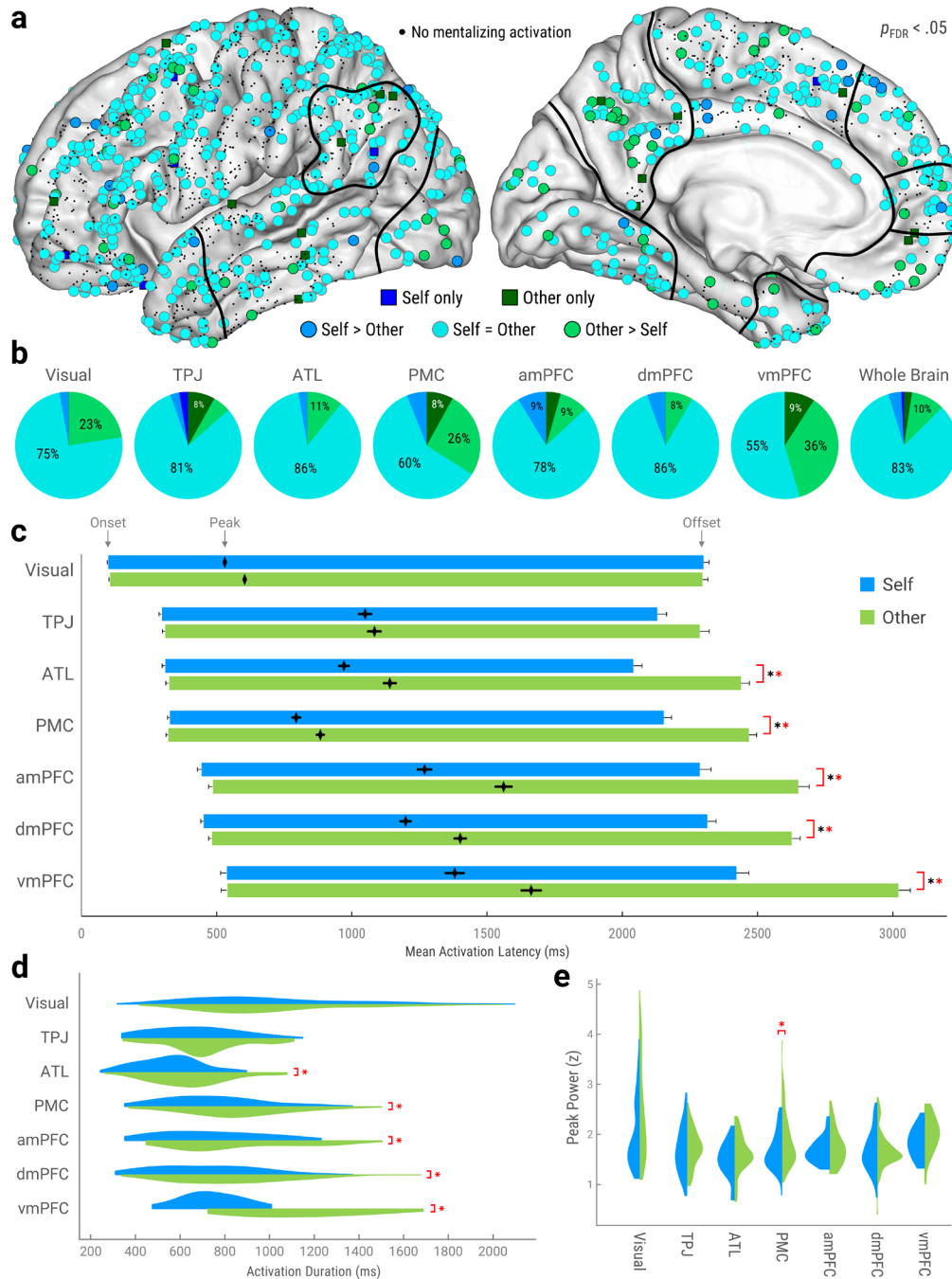


Figure 1.5: A common neurocognitive sequence for self- and other-mentalizing. **a**) Functional anatomy of self- and other-mentalizing. Circles represent sites identified as coactive for both mentalizing types versus baseline ($p_{FDR} < .05$, corrected across timepoints and sites), colored by self/other differences in peak power ($p_{FDR} < .05$, corrected across sites; see Fig. 2). Squares represent sites identified as active for only one mentalizing type with significantly higher peak power for that mentalizing type. Black dots represent sites with nonsignificant mentalizing activations. All sites plotted on left hemisphere with approximate ROI outlines (full views in Supplementary Fig. 9). **b**) Proportions of mentalizing-active ROI sites exhibiting the selectivity profiles in **a**. 'Whole Brain' refers to all mentalizing-active sites in cortex. **c**) Mean activation latencies by mentalizing type and ROI. Left and right floating bar edges depict onsets and offsets, respectively, while diamonds depict peaks (see Fig. 1e). Error bars depict standard error of the mean. *Significant self/other differences for peaks (black) and offsets (red); $p < .05$, uncorrected, controlled for RT_{Behav} and stimulus visual dissimilarity, two-tailed. **d**) Activation duration (d) and magnitude (e; peak power) by mentalizing type and ROI. *Red asterisks indicate significant self/other differences ($p < .05$, uncorrected, controlled for RT_{Behav} and stimulus visual dissimilarity, two-tailed).

Self- and other-mentalizing share a common neuroanatomical basis

To fractionate mentalizing's neurocognitive sequence across mentalizing type, we first explored the anatomical interrelations between self- and other-mentalizing (Fig. 1.5ab). We therefore identified sites that produced significant trial-averaged activations for each mentalizing type relative to baseline ($p_{FDR} < .05$, corrected across timepoints and sites). We also directly compared self/other differences in single-trial peak power using robust regression ($p_{FDR} < .05$, corrected across sites). Sites were considered 'self-only' or 'other-only' if they 1) produced significant trial-averaged activations for only one mentalizing type, and 2) produced significantly greater peak power for that mentalizing type over another. Sites that activated for both mentalizing types were labeled by peak power differences: 'self-greater' (Self>Other), 'other-greater' (Other>Self), or 'non-selective' (Self=Other).

We found that mentalizing-active sites were overwhelmingly coactive for both mentalizing types (non-selective + self-greater + other-greater = 97% whole-brain; Fig. 1.5ab). Moreover, non-selective sites formed the largest category in all ROIs and the whole-brain (range: 55-86%). We also compared amounts of 'self-selective' sites (self-only + self-greater) versus 'other-selective' sites (other-only + other-greater) using McNemar χ^2 (Yates-corrected; $df=1$). This revealed that other-selective sites significantly ($p < .05$) outnumbered self-selective sites in visual cortex (3% self/23% other; $\chi^2=9.39$), PMC (6% self/34% other; $\chi^2=8.45$), and the whole-brain (5% self/13% other; $\chi^2=24.40$). Nonsignificant self/other differences were found in TPJ (5% self/14% other; $\chi^2=0.57$), ATL (3% self/11% other; $\chi^2=0.37$), amPFC (9% self/13% other; $\chi^2=0$), dmPFC (6% self/8% other; $\chi^2=0$), and vmPFC (0% self/45% other; $\chi^2=3.20$).

In sum, self- and other-mentalizing recruited near-identical cortical sites in a largely non-selective manner, though selective sites were mostly other-selective in visual cortex and PMC. Unexpectedly, other-selective and self-selective sites did not significantly outnumber each other

in TPJ, amPFC, and dmPFC, which contradicts previous fMRI work from our group (Lieberman et al., 2019) and elsewhere (Bzdok et al., 2013; de la Vega et al., 2016; Denny et al., 2012).

Other-mentalizing evokes slower and lengthier activations across successive ROIs

Given the highly overlapping neuroanatomy of self- and other-mentalizing, we explored whether self/other differences might be better characterized by fast spatiotemporal functional dynamics. We therefore examined single-trial HFB metrics from mentalizing-active ROI sites. Mentalizing type was analyzed by LMEMs that controlled for RT_{Behav} and visual dissimilarity, nested within Site and Participant.

We found that self- and other-mentalizing evoked a common spatiotemporal sequence of HFB activations across ROIs (Fig. 1.5c). Within this sequence, onset latencies showed nonsignificant self/other differences (Table 1.1). However, other-mentalizing evoked significantly ($p < .05$) later peaks and offsets than self-mentalizing in all DMN ROIs except TPJ (Fig. 1.5c). Concordantly, activation duration was significantly longer for other- versus self-mentalizing in all DMN ROIs except TPJ (Fig. 1.5d). In contrast, self/other differences in activation magnitude were significant only in PMC, which produced greater peak power for other-mentalizing (Fig. 1.5e). Crucially, although HFB metrics often reflected RT_{Behav} and visual dissimilarity, significant self/other differences were ultimately dissociable from these covariates (Supplementary Table 1.1).

Next, we examined whether self/other differences became stronger in successive ROI sites. We therefore compared mean onset latencies (Fig. 1.4a) with single-site random effect sizes (b from analyses in previous paragraph) for peak latency (Fig. 1.5c) and peak power (e.g., self/other selectivity; Fig. 1.5e) using LMEMs nested within Participant. We found that later onsets predicted stronger self/other differences in peak latencies ($b_{255} = 0.404 \pm .131$, $p = .002$). However, onsets did not significantly predict peak power ($b_{255} = -0.303 \pm .0744$, $p = .796$).

In sum, we found that self- and other-mentalizing recruited near-identical sites in a common spatiotemporal sequence (Fig. 1.5a-c). Within this sequence, other-mentalizing evoked slower (Fig. 1.5c) and lengthier (Fig. 1.5d) activations in all DMN ROIs except TPJ. Intriguingly, successive ROI sites had greater self/other differentiation in the timing, rather than selectivity, of activations. Taken together, self/other functional differences were primarily characterized by the timing of activations throughout successive DMN regions.

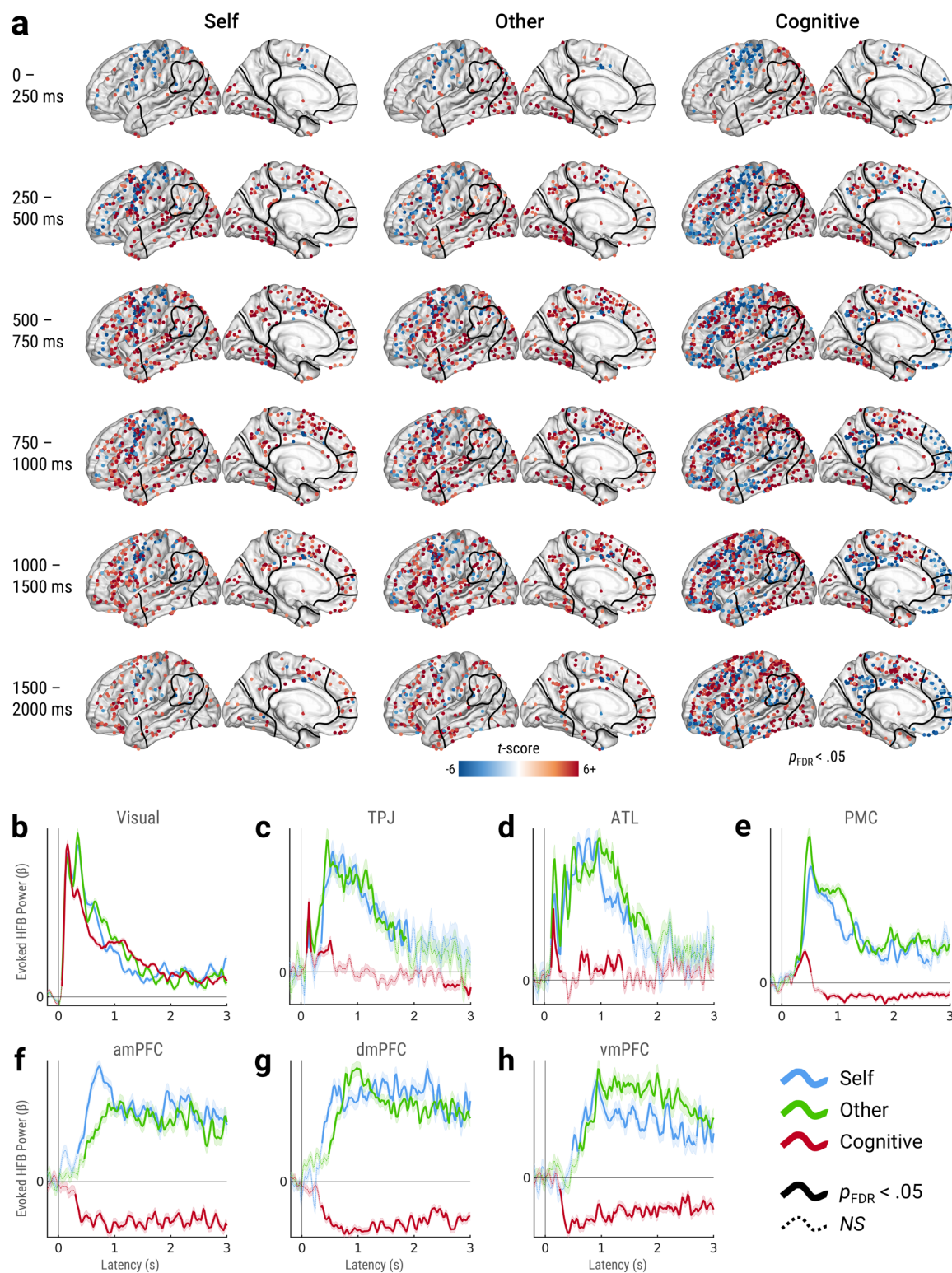


Figure 1.6: Summary of evoked neuronal activity. **a)** Sites with significant HFB responses during specific time windows versus pre-stimulus baseline, colored by LMEM t -values ($p_{FDR} < .05$, corrected across sites and time windows, two-tailed). All sites plotted on left hemisphere with approximate ROI outlines. **b-h)** Grand-average timecourses of evoked HFB power (β) within ROIs (LMEM-estimated). Thick solid lines indicate significant responses versus pre-stimulus baseline ($p_{FDR} < .05$, corrected across ROIs and timepoints, two-tailed). Thin dashed lines indicate nonsignificant responses. Shaded areas indicate standard error of β .

Summary of evoked neuronal activity

To summarize task-evoked neuronal population activity (Fig. 1.6a), we identified sites with significant HFB responses during specific time windows relative to pre-stimulus baseline ($p_{FDR} < .05$, corrected across sites and time windows). From 0-250 ms, activations were largely localized to visual cortex with negligible task condition differences. From 250-500ms, activations also encompassed temporoparietal and lateral frontal regions, where mentalizing and arithmetic (cognitive task) began to diverge. Specifically, in temporoparietal DMN regions, self- and other-mentalizing primarily evoked activations, while arithmetic evoked interdigitated activations and deactivations. From 500-750 ms, mentalizing activations began to encompass mPFC, especially for self-mentalizing. In contrast, arithmetic evoked mPFC deactivations, which continued for all successive time windows. From 750-1000 ms, both mentalizing types evoked similar mPFC activations. From 1000-2000 ms, other-mentalizing evoked more sustained activations than self-mentalizing, particularly in mPFC.

Controlling for stimulus visual dissimilarity (VD)

To distinguish mentalizing-related neuronal effects (mentalizing type, RT_{Behav} , and $Choice_{Behav}$) from stimulus VD (e.g., prompt length), we used computer vision (Jarrett et al., 2009; Leeds et al., 2013) (see Supplementary Methods). After controlling for VD, mentalizing-related effects were still significant in all DMN ROIs (Figs. 1.4de, 1.5c-e & Table 1.1). However, in visual cortex, marked self/other differences in peak power became nonsignificant (Fig. 1.5e). In sum, mentalizing-related effects in DMN ROIs were not explained by prompt length and other visual features.

Behavioral results

To confirm that participants performed mentalizing for task prompts, we analyzed $\text{Choice}_{\text{Behav}}$ for self/other biases towards positive or negative affective traits (Supplementary Fig. 1.4c) using logistic mixed-effects classification (nested within Trait within Participant). Overall, we found a greater probability of 'true' choices for positive versus negative traits (odds ratio=2.53, $b_{1471}=0.931\pm.069$, $p=1.04e-39$). This bias was stronger during self-mentalizing versus other-mentalizing (odds ratio=1.38, $b_{1471}=0.326\pm.064$, $p=4.56e-7$). These results indicate that participants engaged the mentalistic content in task prompts with a self-positivity bias, which is a canonical feature of mentalizing (Igou, 2008; Pahl & Eiser, 2005; Silani et al., 2013). Moreover, high accuracy in arithmetic trials (median=92.5%) suggests effortful attention to task prompts (Supplementary Fig. 1.4b).

To determine whether RT_{Behav} varied by Mentalizing Type, $\text{Choice}_{\text{Behav}}$, and stimulus visual features (VD1 and VD2; see Supplementary Results), we used a LMEM nested within Participant (Supplementary Fig. 1.4a). We found faster RT_{Behav} for self-mentalizing ($M=2406\pm 49$ ms) versus other-mentalizing ($M=2797\pm 52$ ms; $t_{1646}=2.85$, $p=.004$). Additionally, RT_{Behav} was faster for 'true' choices ($M=2482\pm 59$ ms) versus 'false' choices ($M=2807\pm 45$ ms; $t_{1646}=2.16$, $p=.031$). However, the Mentalizing Type x $\text{Choice}_{\text{Behav}}$ interaction was nonsignificant ($t_{1646}=1.47$, $p=.140$). Unsurprisingly, longer prompt lengths (VD1) elicited slower RT_{Behav} ($t_{1646}=4.85$, $p=1.36e-6$). Meanwhile, remaining visual features (VD2) evoked nonsignificant RT_{Behav} effects ($t_{1646}=-0.25$, $p=.803$). In sum, mentalizing type, $\text{Choice}_{\text{Behav}}$, and VD1 had significant and dissociable effects on RT_{Behav} .

Discussion

Using electrocorticography (ECoG), we probed the neurocognitive substrates of mentalizing at the level of neuronal populations. We found that mentalizing about the self and

others recruited near-identical cortical sites (Fig. 1.5ab) in a common spatiotemporal sequence (Figs. 1.5c & 1.6). Within our ROIs, activations began in visual cortex, followed by temporoparietal DMN regions (TPJ, ATL, and PMC), and lastly in mPFC regions (amPFC, dmPFC, and vmPFC; Fig. 1.3ef). Critically, regions with later activations exhibited greater functional specialization for mentalizing as measured by three metrics: functional specificity for mentalizing versus arithmetic (Fig. 1.3cd & 1.4b), self/other differentiation in activation timing (Fig. 1.5cd), and temporal associations with behavioral responses (Fig. 1.4d & Table 1.1). Taken together, these results reveal a common neurocognitive sequence (Felleman & Van Essen, 1991; Mormann et al., 2008; Diaconescu et al., 2017; Hilgetag & Goulas, 2020) for self- and other-mentalizing, beginning in visual cortex (low specialization), ascending through temporoparietal DMN areas (intermediate specialization), then reaching its apex in mPFC regions (high specialization).

Our results are consistent with gradient-based models of brain function, which posit that concrete sensorimotor processing in unimodal regions (e.g., visual cortex) gradually yields to increasingly abstract and inferential processing in ‘high-level’ transmodal regions like mPFC (Huntenburg et al., 2018; Margulies et al., 2016). We found that the strength of self/other differences in activation timing increased along a gradient from visual cortex to mPFC. Specifically, other-mentalizing evoked slower (Fig. 1.5c) and lengthier (Fig. 1.5d) activations than self-mentalizing throughout successive DMN ROIs. These self/other functional differences corresponded with self/other differences in RT_{Behav} (Supplementary Fig. 1.4), although the two were often dissociable (Table 1.1). Thus, perhaps because we know ourselves better than others, other-mentalizing may require lengthier processing at more abstract and inferential levels of representation, ultimately resulting in slower behavioral responses.

What might our results imply about extant fMRI findings? Hundreds of fMRI studies consistently suggest that: (1) TPJ and dmPFC are most crucial for mentalizing (Apperly et al.,

2004; Lieberman, 2010; Molenberghs et al., 2016; Samson et al., 2004; Saxe, 2010; Saxe & Powell, 2006; Schurz et al., 2021; Van Overwalle & Baetens, 2009), and (2) dmPFC is selective for thinking about others over oneself (Bzdok et al., 2013; de la Vega et al., 2016; Denny et al., 2012; Lieberman et al., 2019). However, when examined with ECoG, we found that both pieces of received wisdom are not what they seem. Below, we discuss both issues before moving onto our peculiar vmPFC results, and then conclude with systems-level discussion.

Unsurprisingly, we found that DMN regions such as TPJ and dmPFC contained higher proportions of ‘mentalizing-specific’ sites (i.e., mentalizing-active but not cognitive-active) relative to the whole-brain average (Fig. 1.3cd). The spatial distribution of mentalizing-specific sites roughly resembles the ‘mentalizing network’ reported in countless fMRI studies (Amft et al., 2015; Frith & Frith, 2006; M. D. Lieberman, 2010; Mars et al., 2012; Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle & Baetens, 2009). However, our DMN ROIs were by no means functionally homogenous. Relative to other ROIs, TPJ and dmPFC activity best predicted RT_{Behav} (Fig. 1.4de & Table 1.1), supporting the notion that both regions are most crucial for mentalizing performance (M. D. Lieberman, 2010; Schurz et al., 2021; Apperly et al., 2004; Saxe & Powell, 2006; Saxe, 2010; Cook, 2014; Spunt et al., 2015).

We also found numerous functional distinctions between TPJ and dmPFC, which is surprising given their remarkably similar functional profiles in fMRI literature (Amft et al., 2015; Andrews-Hanna et al., 2014; Frith & Frith, 2006; M. D. Lieberman, 2010; Mars et al., 2012; Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle & Baetens, 2009). Specifically, we found that TPJ produced earlier activations (Fig. 1.3ef) that were notably coactive for mentalizing and arithmetic (cognitive task; Fig. 1.3cd). Indeed, the onsets of TPJ activations were the earliest of any DMN ROI. In contrast, dmPFC produced significantly later activations (Fig. 1.3ef) that were overwhelmingly mentalizing-specific (Fig. 1.3cd), indicating that dmPFC sits at a higher level of mentalizing’s neurocognitive sequence than TPJ. Moreover, aggregate

ROI analyses revealed no significant self/other differences in TPJ (Table 1.1), while dmPFC featured robust self/other timing differences (Fig. 1.5cd), suggesting that dmPFC is more sensitive to variation in mentalistic content. In terms of behavior, although TPJ and dmPFC best predicted RT_{Behav} , dmPFC was the only ROI that predicted $\text{Choice}_{\text{Behav}}$ (choosing ‘true’ or ‘false’ for task prompts; Table 1.1), implying that dmPFC instantiates more aspects of mentalistic behavior than TPJ. Strikingly, unlike TPJ, dmPFC activation offsets closely preceded behavioral responses (within 124 ± 7 ms; Fig. 1.4c), suggesting that dmPFC is more deeply involved in the final stages of mentalistic reasoning. Taken together, while TPJ and dmPFC are both clearly crucial for mentalizing performance, dmPFC appears more specialized for mentalizing itself.

Given the marked functional differentiation between TPJ and dmPFC, what specific neurocognitive roles might they play in mentalizing? In social neuroscience, TPJ is often considered a functionally-specific locus for explicit belief reasoning (Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005; Saxe, 2010). Yet here, TPJ was less functionally specialized relative to dmPFC (Figs. 1.3c-f, 4 & 5cd). To explain this discrepancy, we suggest that TPJ provides crucial antecedents for explicit belief reasoning in dmPFC. Given TPJ’s central role in automatic evaluations of thematic semantics (Sass et al., 2009; Binder & Desai, 2011; Geng & Vossel, 2013; Davey et al., 2015; Lewis et al., 2015; Lambon Ralph et al., 2017; Murphy et al., 2017; Teige et al., 2018; Patel et al., 2019), we propose that TPJ automatically represents integrative psycho-semantic models of exemplar contexts for a given inference. In simpler terms, TPJ may help us ‘see’ the psycho-semantic gestalt of a given situation (M. Lieberman, 2021). Accordingly, tasks that ‘show’ concrete mentalistic content (e.g., social animations) often recruit TPJ but not dmPFC, while tasks that require mentalistic logical inferences (e.g., false belief) often recruit dmPFC in addition to TPJ (Schurz et al., 2021; Amft et al., 2015; Molenberghs et al., 2016). Thus, when mentalistic content feels ‘seeable’ from perceptual processing, TPJ could generate mentalistic inferences without explicit belief reasoning. Indeed,

work on implicit and spontaneous mentalizing show that TPJ can encode an actor's beliefs without any explicit reasoning (Van Duynslaeger et al., 2007; Ma et al., 2011; Kestemont et al., 2013; Kovács et al., 2014; Hyde et al., 2015; Schuwerk et al., 2015; Naughtin et al., 2017; Schneider et al., 2017; Bardi et al., 2017; Boccadoro et al., 2019). Taken together, TPJ may implicitly set the psycho-semantic stage for explicit belief reasoning that occurs later in dmPFC when necessary (e.g., our trait judgment task; Fig. 1.1a).

The dmPFC may be well-suited for explicit belief reasoning (J. P. Mitchell, 2006; Krol & Krol, 2010; Tamir, Thornton, et al., 2016; Tamir & Thornton, 2018). We found substantial concurrent activation between dmPFC and all other ROIs (Fig. 1.3e, 1.6 & 1.2), suggesting that dmPFC could work iteratively with lower-level regions to refine what is 'seen', thus providing dmPFC with increasingly-useful inputs from which to draw better inferences (Cunningham & Zelazo, 2007; C. J. Mitchell et al., 2009). Moreover, studies on strategic reasoning show that dmPFC can arbitrate between multiple mental models (S. W. Lee et al., 2014) and prospective choices (Doll et al., 2015) by simultaneously evaluating multiple possibilities (Barron et al., 2013) through 'fuzzy' propositional reasoning (C. J. Mitchell et al., 2009; J. P. Mitchell, 2009). As such, dmPFC may arbitrate between multiple exemplar contexts to help extract the most relevant and enduring semantic features for a given psychological inference. Taken together, dmPFC may integrate and refine representations throughout mentalizing's neurocognitive pathway to strategically reason about minds.

As for mentalizing about the self or others, fMRI studies routinely suggest that dmPFC is 'other-selective' (Bzdok et al., 2013; de la Vega et al., 2016; Denny et al., 2012; M. D. Lieberman et al., 2019). What underlying neuronal population dynamics could result in stronger hemodynamic responses for one mentalizing type over another? The standard assumption would be that the magnitude (i.e., intensity) of neuronal activations differs across mentalizing type. This might be seen in aggregate ROI activity, or perhaps across individual ROI sites. We tested both

possibilities by examining self/other differences in activation magnitude (HFB peak power). Unexpectedly, aggregate ROI analysis revealed that dmPFC produced nonsignificant self/other differences in peak power (Fig. 1.5e & Table 1.1). Similarly, single-site analysis showed that most dmPFC sites had nonsignificant self/other differences in peak power (86%), while the remaining 'self-greater' and 'other-greater' sites did not significantly outnumber each other (Fig. 1.5ab). Strikingly, dmPFC did not contain sites that only activated for one mentalizing type. In sum, both mentalizing types recruited identical dmPFC sites (100% overlap) at equivalent intensities, which appears inconsistent with numerous fMRI reports of 'other-selective' dmPFC responses.

We instead found robust self/other differences in the timing of dmPFC activations. Specifically, other-mentalizing evoked slower and lengthier activations compared to self-mentalizing (Fig. 1.5cd). In other words, dmPFC activity remained significantly above baseline for longer during other-mentalizing (see Figs. 1.2e & 1.6a). This suggests a different account of why dmPFC produces stronger hemodynamic responses for mentalizing about others over oneself. The typical story is that dmPFC is highly specialized for thinking about other people's minds (Bzdok et al., 2013; de la Vega et al., 2016; Denny et al., 2012; M. D. Lieberman et al., 2019). Alternatively, dmPFC could be sensitive to the inherently greater difficulty of other-mentalizing, which may necessitate additional processing cycles before completion. This additional processing may involve 'anchoring-and-adjustment'; the use of self-representations as an anchor from which to adjust other-mentalizing inferences – a function strongly linked with dmPFC and its role in strategic reasoning (Tamir & Mitchell, 2010, 2013; Tamir & Thornton, 2018). Self-mentalizing may be simplified by the rich compendium of accessible information we have about ourselves, thus resulting in brief but equally intense processing. Given that standard fMRI analysis does not distinguish activation intensity from activation duration, it appears that

the latter has been mistaken for the former – though we cannot exclude the possibility of confounds related to differences between ECoG and fMRI (Parvizi & Kastner, 2018).

Our peculiar vmPFC results may shed light on schematic contributions to mentalizing. Although there is growing consensus that vmPFC is important for schema processing (M. D. Lieberman et al., 2004; Aminoff, 2014; Spalding et al., 2015; Bein et al., 2018; Giuliano et al., 2021; M. D. Lieberman et al., 2019), our results appear contrary to this at first blush. We found that vmPFC produced the latest activations across ROIs (Fig. 1.3ef), which contradicts numerous reports of early (<200 ms) schema-based ‘gist’ predictions in vmPFC (Cardinal et al., 2002; Bar et al., 2006; Kveraga et al., 2007; Olofsson et al., 2008; Barrett & Bar, 2009; Kveraga et al., 2011; Guillory & Bujarski, 2014). Early vmPFC activation has even been observed during mentalizing (Mossad et al., 2016). Of note, many studies have observed ‘double waves’ of early and late vmPFC activity: early activity may reflect vision-based gist construal, while late activity may reflect elaborative situational construal (Barrett & Bar, 2009; Rudrauf et al., 2008; Grabenhorst & Rolls, 2011; Wyczesany et al., 2015; Panichello et al., 2017; Tan & Tarr, in prep). There is some evidence of a ‘double wave’ in the present study, as three posterior vmPFC sites produced very early activations (<200 ms; Figs. 1.2k & 1.4a) that met outlier exclusion criteria for ROI analyses. These early vmPFC sites produced equivalent activations for all task conditions (Figs. 1.2k, 1.3c & 1.5a), aligning with characterizations of rapid magnocellular gist processing, which likely cannot discern our alphanumeric stimuli (Chaumon et al., 2014; Kveraga et al., 2007). In contrast, late-onset vmPFC sites produced longer activations for other- versus self-mentalizing (Figs. 1.5cd & 1.2f), perhaps reflecting other-mentalizing’s greater reliance on schematic feedback, especially in our trait judgement task (FeldmanHall & Shenhav, 2019; M. D. Lieberman et al., 2004; Raykov et al., 2020; Shenhav et al., 2013). Nevertheless, self- and other-mentalizing recruited near-identical vmPFC sites (91% overlap; Fig. 1.5ab), suggesting common schematic underpinnings (Jenkins et al., 2008; M. D. Lieberman et al., 2004, 2019).

Taken together, we propose that vmPFC provides schematic contributions to mentalizing in two ways: rapid predictive processing from coarse visual afferents, followed by slow situational processing involving schematic feedback and integration across DMN.

At the systems level, we found complex and hierarchical processing dynamics across mentalizing's putative neurocognitive pathway. We observed consistent trial-by-trial propagation of activation onsets across ROIs (Supplementary Fig. 1.3), portraying an initial 'feedforward sweep' of coarser processing (Felleman & Van Essen, 1991; DiCarlo & Maunsell, 2005; Alilović et al., 2019) along the pathway. Indeed, onset latencies were insensitive to self/other differences (Table 1.1). Onsets were followed by considerable concurrent activations across all ROIs (Figs. 1.3e, 1.5c, 1.6, 1.2 & S2), suggesting that distinct pathway regions largely work together (i.e., recurrent processing (Felleman & Van Essen, 1991; Hebb, 1949; Lamme & Roelfsema, 2000)) within an overarching processing sequence. Recurrent processing may be crucial for self/other differentiation, as self/other differences in aggregate ROI activity did not reach significance until concurrent activation was achieved across all ROIs (e.g., activation peaks and offsets; Fig. 1.5). Taken together, mentalizing may be supported by a brief initial 'feedforward sweep' of coarser processing along the pathway, followed by substantial recurrent processing that may integrate and refine representations across pathway regions. These dynamics could obscure cross-regional functional distinctions in fMRI studies. However, further research involving connectivity analyses are needed for more conclusive claims.

Temporoparietal DMN regions (tpDMN; TPJ, ATL, and PMC) may help integrate representations throughout mentalizing's neurocognitive pathway. Classification analysis revealed two distinct functional types (Fig. 1.4b) that were anatomically interdigitated in tpDMN ROIs: sites with earlier non-specific activations, and sites with later mentalizing-specific activations (Fig. 1.3c). Intriguingly, non-specific tpDMN sites often coactivated with lower-level regions like visual cortex (e.g., Fig. 1.2b-d), suggesting attunement to lower-level afferents.

Meanwhile, mentalizing-specific tpDMN sites often coactivated with higher-level regions like mPFC (e.g., Fig. 1.2efhi), suggesting attunement to higher-level feedback. Critically, we also found lengthy concurrent activations across site types and ROIs (Figs. 1.2 & 1.6), which could recurrently integrate low- and high-level representations (Cunningham & Zelazo, 2007; Fischmeister et al., 2017; Krol & Krol, 2010; C. J. Mitchell et al., 2009). Indeed, we found that visual and mentalistic representations were simultaneously encoded in all tpDMN ROIs (Supplementary Table 1.1). Taken together, we propose that interactions between distinct interdigitated neuronal populations in tpDMN (Parvizi et al., 2006; Foster et al., 2015; Daitch & Parvizi, 2018) may help integrate and refine distant representations across mentalizing's neurocognitive pathway.

This study is not without confounds and limitations. Some of these limitations are inherent to ECoG: the use of participants with epilepsy, inconsistent brain coverage across participants, and sampling bias for cortical gyri (Parvizi & Kastner, 2018). Although these limitations were mitigated to the best of our ability (see Methods), they cannot be completely ameliorated. Thus, our ECoG findings could be corroborated by examining healthy participants with recent advances in source-space EEG/MEG, such as ultra-high density EEG (Robinson et al., 2017), optically-pumped MEG (Boto et al., 2016), and laminar source localization (Bonaiuto, Rossiter, et al., 2018). Another important confound was the sparse right-hemisphere coverage of our cohort, which may limit the interpretability of our ATL and TPJ results (Igelstrom et al., 2016). Nonetheless, our few right-hemisphere sites appear functionally analogous to their left-hemisphere homologues (Supplementary Figs. 1.6–1.9). A task-related limitation was the short pre-stimulus baseline (Fig. 1.1a), which sometimes contained residual activity from prior trials, likely resulting in artifactual 'deactivations' in somatomotor sites (Fig. 1.6a). Another task-related confound was greater prompt length for other-mentalizing (e.g., "My neighbor is...") versus self-mentalizing (e.g., "I am..."; Supplementary Table 1.4), which we controlled for using

computer vision (Jarrett et al., 2009; Leeds et al., 2013) (see Supplementary Methods).

Relatedly, mentalizing-specificity (Fig. 1.3cd) could arise from differences between sentences and arithmetic equations. However, mentalizing-specific sites were not concentrated in reading-related regions (Llorens et al., 2011; Long et al., 2021; Miller et al., 2011; Wu et al., 2011), but rather in the ‘mentalizing network’ reported by countless fMRI studies (Schurz et al., 2021).

Distributed hierarchical processing is a central organizing principle of neurocognitive systems (Hebb, 1949; Felleman & Van Essen, 1991; Huntenburg et al., 2018; Hilgetag & Goulas, 2020). Characterizing such hierarchies has enabled incisive neuromechanistic accounts of many psychological functions (Diaconescu et al., 2017; Demirtaş et al., 2019). Here we provide a comprehensive electrophysiological exploration of the human social brain, revealing that mentalizing is characterized by complex and hierarchical neurocognitive dynamics at millisecond, millimeter, and cross-regional scales. While many questions remain, our findings contribute to a solid foundation upon which more conclusive neurocognitive accounts of mentalizing can be built.

Methods

All research activities herein were conducted in accordance with protocol approved by the Stanford Institutional Review Board for human experimentation. All computational procedures and analyses herein were implemented in MATLAB (Mathworks, 2019) unless otherwise specified.

Participants

We employed a cohort of sixteen human participants while they underwent neurosurgical treatment for drug-resistant epilepsy (demographics in Supplementary Table 1.2). Each participant provided written informed consent in accordance with the Stanford

Institutional Review Board. Participants were not compensated as per Stanford Institutional Review Board guidelines for inpatient clinical research.

As part of their presurgical evaluation, participants were implanted with ECoG at Stanford University Medical Center. The anatomical placement of electrode sites was determined according to each participant's clinical needs. Participants were included in this study's cohort if they had electrode coverage in key DMN regions: mPFC, PMC, TPJ, and ATL. Each participant was monitored in hospital for six to ten days prior to surgery, during which the study was conducted.

Behavioral Task

ECoG data was recorded while participants performed an event-related behavioral task with six conditions (trial types; Fig.1.1a & Supplementary Table 1.4). Five of these conditions required true/false responses to written prompts, while one condition consisted of cued rest. Two conditions featured mentalizing prompts, either about oneself (e.g., "I am honest") or others (e.g., "My neighbor is honest"). Participants were instructed to select a single neighbor (current or past) as the target for other-mentalizing. Cognitive task trials consisted of basic arithmetic (e.g., " $9 + 86 = 95$ "). Two conditions featured memory-related prompts: episodic (e.g., "I ate candy yesterday") and self-semantic (e.g., "I eat a lot of candy"). The cued rest condition required no response and displayed a fixation crosshair for 5-10 seconds. The memory and rest conditions were not relevant to the current analyses have been reported elsewhere (Daitch & Parvizi, 2018; Foster et al., 2015; Raccach et al., 2018). Stimuli were presented in random order and were self-paced, advancing to the next trial after the participants' response, or up to 15 seconds if no response. The inter-trial interval (ITI) occurred -200-0 ms before each trial. The experiment was broken into two separate runs (mean run duration= 12.50 ± 1.64 minutes). Participants were allowed a short break between behavioral runs. On average, each run featured

25 trials of each sentence condition, 40 cognitive trials, and 36 rest trials. Each non-rest trial contained unique prompts; prompts were not repeated within participant. Responses were made via a handheld keypad using either the '1' (true) or '2' (false) key. Participants were instructed to perform the task as accurately and as quickly as possible. All stimuli were presented in white font on a black background (1200x800 pixels) using Psychophysics Toolbox 3 (Kleiner et al., 2007).

For all analyses, trials were excluded if meeting any of the following criteria: high-frequency epileptic oscillations, no behavioral responses, irrelevant button presses, or RT_{Behav} under 400 ms.

Electrocorticography data acquisition

ECoG recordings were obtained via 2125 subdural electrodes (Fig. 1.1b). Electrodes (platinum plates with diameter of 1.2-2.3 mm) were implanted subdurally onto the cortical surface in grids or strips with center-to-center interelectrode spacing of 4-10 mm (Adtech Medical Instruments). Electrodes were connected to a multichannel recording system (Nihon Kohden; Tucker Davis Technologies) with sampling rate of 1,000 Hz or above. Anatomical data was acquired using a GE 3-Tesla SIGNA Magnetic Resonance Imaging (MRI) scanner at Stanford University. A T1-weighted anterior-posterior commissure-aligned pulse sequence was used. T1 data was resampled to 1 mm isotropic voxels, then segmented to distinguish gray and white matter using FreeSurfer (Fischl, 2012). To facilitate electrode localization, postimplant computerized tomography (CT) scans were coregistered to the preoperative MRI anatomical brain volume (Hermes et al., 2010). For each participant, electrodes sites were localized in BiImage Suite (Papademetris et al., 2006) and displayed on the participants' own reconstructed 3D cortical surface using the iELVis toolbox (Groppe et al., 2017). Electrode

positions were corrected for postimplantation brain shift , allowing for the accurate anatomical localization of electrodes sites (Dykstra et al., 2012).

Defining regions of interest (ROIs) and brain networks

Each participant's native-space cortical surface reconstruction (e.g., Supplementary Fig. 1.2) was used to classify electrode sites into a priori ROIs that are strongly implicated in mentalizing, with visual cortex included as a control ROI (Fig. 1.1c). Standardized brain-based parcellation was avoided due to known transformation inconsistencies in ECoG (Groppe et al., 2017). ROIs were defined through FreeSurfer cortical parcellation combined with visual inspection of anatomical landmarks. The ROI for 'visual cortex' consisted of occipital cortex, lingual gyrus, posterior fusiform gyrus, and posterior inferotemporal cortex. The 'ATL' ROI consisted of a bilateral anterior subregion of temporal cortex with precentral sulcus as the posterior bound, comprising the temporal poles and adjacent sections of entorhinal cortex and superior, middle, and inferior temporal sulci/gyri. The 'TPJ' ROI was a bilateral posterior subregion of inferior parietal lobule with lateral sulcus as the anterior bound, comprising angular gyrus and adjacent sections of supramarginal gyrus and superior temporal sulcus/gyrus. The 'PMC' ROI consisted of precuneus, posterior cingulate, and retrosplenial cortex. The 'amPFC' ROI was an mPFC subregion bounded between the ventral and dorsal reaches of corpus callosum. The 'dmPFC' ROI was a mPFC subregion ventrally bounded by the amPFC ROI and posteriorly bounded by the callosal rostrum. The 'vmPFC' ROI was an mPFC subregion dorsally bounded by the amPFC ROI and posteriorly bounded by the callosal rostrum, including the medial orbitofrontal surface.

ECoG preprocessing

Preprocessing was performed on a single-site/single-participant basis using custom routines (see Code Availability). First, data were notch filtered for power-line noise (57-63 Hz) and harmonics (117-123 Hz, 177-183 Hz). Electrode sites were discarded from further analyses if they were marked as pathological or 'noisy' by postclinical evaluation. The data was then rereferenced by subtracting the mean signal of the remaining electrodes from each electrode's signal. The rereferenced data underwent time-frequency decomposition into 4-200 Hz spectra in 1-10 Hz bands using 5-cycle Morlet wavelet transforms. The power of the signal in each frequency band was z-transformed across time; this helps correct the 1/frequency decay of neurophysiological signals and improves interpretability. Data was then epoched into trials that were time-locked to stimulus presentation, ranging from 200 ms pre-stimulus to 2000 ms peri-stimulus. For each trial and frequency band, baseline correction was performed by subtracting the mean power across the pre-stimulus baseline period (-200–0 ms) from all timepoints within a trial. To reconstruct the high-frequency broadband (HFB) signal, the primary signal of interest, z-transformed power of frequency bands within 70-180 Hz were averaged to produce a single HFB timecourse per electrode. Lastly, HFB timecourses from each electrode were low-pass filtered with a gaussian window (width=50 ms) for further analysis. Trials were rejected from further analyses if epileptic high-frequency oscillations were observed. Within epochs, timepoints were discarded after presentation of the next trial's stimulus (i.e., timepoints after $RT_{\text{Behav}} + 200 \text{ ms ITI}$).

Statistics

Statistical analyses described below were implemented in MATLAB Statistics and Machine Learning Toolbox (Mathworks, 2019). Data was visualized using the GRAMM toolbox (Morel, 2018). All statistical tests were two-tailed. All multiple comparisons corrections

maintained the False Discovery Rate (FDR) under 0.05 through the Benjamini-Yekutieli procedure for data with any dependence structure (Yekutieli & Benjamini, 1999), with p -values adjusted accordingly (p_{FDR}).

We primarily used mixed models due to their design flexibility and robustness to sampling heterogeneity, multicollinearity, and statistical confounds (Hox et al., 2017). Specifically, we employed linear mixed-effects modeling (LME) for continuous outcomes and logistic mixed-effects classification (LMEC) for binary outcomes. To minimize estimation bias, mixed models used restricted maximum likelihood-based estimators (Hox et al., 2017). To account for full data dependence structure with reduced bias and assumptions, mixed models used unconstrained variance-covariance matrices with log-Cholesky parametrization (Barr et al., 2013). To account for heterogeneous variances across mixed model terms, Satterthwaite approximation for degrees of freedom (effective DF) was used (Gaylor & Hopper, 1969). To rectify violations of assumptions and overparameterization in mixed models, we evaluated objective function Hessian matrices and ensured positive definiteness (Schielzeth et al., 2020). Additionally, non-mixed models were used when specified below.

Within-site analyses

Within-site analyses (Fig. 1.1d-f) were performed to provide the bases for the primary multi-site analyses. The dependent variable for within-site analyses was z-scored HFB power across timepoints and trials (Fig. 1.1d). Trials were excluded if containing high-frequency epileptic oscillations, no behavioral responses, irrelevant button presses, or RT_{Behav} under 400 ms. Timepoint-by-timepoint observations met outlier exclusion criteria if HFB power exceeded three median absolute deviations from other observations of the same timepoint and task condition.

Trial-averaged analysis (within-site)

To identify sites with significant evoked HFB responses for each task condition ($p_{\text{FDR}} < .05$, corrected across timepoints and sites), we used LMEMs (Fig. 1.1f). The intercept (null distribution) consisted of timepoints within the pre-stimulus baseline (-200–0 ms). Each peri-stimulus timepoint (0–5000 ms) was represented as a separate dummy variable. The intercept was nested within trial to account for trial-specific variance. This LMEM specification estimates mean timecourses of z-scored HFB power (β) for each task condition. To dampen spikes and other noise, timepoints were not considered significant unless $p_{\text{FDR}} < .05$ was maintained for 50 ms consecutively. For each task condition, sites were considered ‘active’ or ‘deactive’ if evoked HFB power was significantly higher or lower, respectively, than pre-stimulus baseline; if sites produced both, the polarity of the greatest deflection was used. Sites with nonsignificant differences from baseline were considered ‘nonresponsive’.

Single-trial analysis (within-site)

To identify timepoints with significant evoked HFB activations for each trial (Fig. 1.1e), timepoints between stimulus onset and the forthcoming trial ($\text{RT}_{\text{Behav}} + 200$ ms) were run through a sliding window test (width = ± 10 ms). Observations (z-scored HFB power) in each sliding window were tested against observations from the pre-stimulus baseline via two-sample Welch’s t-tests, which accounted for unequal variances and sample sizes. This analysis identified timepoints within individual trials that featured significant stimulus-evoked responses ($p_{\text{FDR}} < .05$, corrected across timepoints, trials, and sites) relative to the pre-stimulus baseline preceding each trial. To dampen spikes and other noise, timepoints were not considered significant unless $p_{\text{FDR}} < .05$ was maintained for 50 ms consecutively.

Single-trial analysis provided five key metrics of HFB activity (Fig. 1.1e). Onset latency is the earliest timepoint with significant activation (green squares). Peak latency and peak power

are the timepoint and magnitude, respectively, of the strongest activation (white squares). Offset latency is the latest timepoint with significant activation (red squares). Duration is the total number of timepoints with significant activations (brown areas).

Multi-site analyses

Multi-site analyses used results from within-site analyses as response measures. Outliers were identified with respect to the two-dimensional distance between stimulus onset and RT_{Behav} using bisquare robust regression (DiCarlo & Maunsell, 2005). For each site and single-trial HFB metric, observations were considered outliers and discarded if residuals were greater than three median absolute deviations (MAD). For ROI sites, this method was applied a second time using observations from all mentalizing-active sites within each ROI. Sites were excluded from ROI analyses if over 50% of its observations exceeded three MAD. Of all ROI sites, only three sites in vmPFC were excluded.

Functional specificity and selectivity (multi-site)

Functional specificity (Fig. 1.3cd) of each site was determined by comparing mentalizing (collapsed across self/other) and arithmetic (cognitive task) results from trial-averaged analysis, along with direct comparisons of single-trial peak power (including trials with nonsignificant activations) for mentalizing versus arithmetic using bisquare robust regression. Specifically, sites were considered 'mentalizing-specific' if they produced 1) significant trial-averaged activations for mentalizing but not arithmetic ($p_{\text{FDR}} < .05$, corrected across sites and timepoints), and 2) produced greater peak power for mentalizing over arithmetic ($p_{\text{FDR}} < .05$, corrected across sites). Sites with significant trial-averaged coactivations for mentalizing and arithmetic were considered 'non-specific', regardless of peak power differences. All mentalizing-

active sites with nonsignificant mentalizing/arithmetic peak power differences were labeled as 'non-specific' (Mentalizing active = Cognitive active).

Self/other selectivity (Fig. 1.5ab) of each site was determined by trial-averaged results and direct comparisons of single-trial peak power (including trials with nonsignificant activations) for self- versus other-mentalizing using bisquare robust regression. Sites were considered 'self-only' or 'other-only' (not considering cognitive task) if they only produced 1) significant trial-averaged activations for only one mentalizing type ($p_{FDR} < .05$, corrected across sites and timepoints) and 2) produced greater peak power for that mentalizing type over the other ($p_{FDR} < .05$, corrected across sites). Sites that activated for both mentalizing types were labeled by self/other differences in peak power: 'self-greater' (Self>Other), 'other-greater' (Other>Self), and 'non-selective' (Self=Other). Sites that activated for only one mentalizing type but had nonsignificant self/other differences in peak power were considered 'non-selective' (Self=Other). In sum, self-only and self-greater sites were considered 'self-selective', while other-only and other-greater sites were considered 'other-selective'.

Aggregate ROI analyses (multi-site)

To reveal the spatiotemporal dynamics of mentalizing processing, we analyzed single-trial HFB metrics from mentalizing-active ROI sites. Each ROI and HFB metric was analyzed using separate LMEMs. All LMEMs were nested within Participant, which accounted for within- and between-participant heterogeneity through unconstrained variance/covariance matrices across within-participant random effects. LMEMs that included Site or Trial as nesting factor likewise accounted for site- and trial-related heterogeneity, respectively. Stimulus visual dissimilarity was controlled for by including two visual dimensions (VDs) as random effects specified below; VDs were derived from a popular computer vision model based on the ventral visual stream (Jarrett et al., 2009; Leeds et al., 2013) (see Supplementary Methods). In addition to outlier exclusion

criteria in the *multi-site analyses* section above, trials with RT_{Behav} over 5000 ms were excluded given our 5000 ms epoch lengths.

LMEMs for self/other differences (Fig. 1.5c-e & Table 1.1) represented mentalizing type as a fixed and random effect. We also included random effects for RT_{Behav} and VDs, which controlled for RT_{Behav} and stimulus visual dissimilarity. All random effects were nested within Site and Participant to account for site- and participant-related heterogeneity.

LMEMs for neurobehavioral associations (Fig. 1.4de & Table 1.1) simultaneously estimated RT_{Behav} and $\text{Choice}_{\text{Behav}}$ (controlling for one another) as fixed and random effects, while VDs were specified as random effects. All random effects were nested within Site and Participant to account for site- and participant-related heterogeneity.

Pairwise ROI comparisons (Fig. 1.3f) were performed using LMEMs with ROI as a fixed and random effect nested within Trial within Participant, which estimated within-trial ROI differences and accounted for trial- and participant-related heterogeneity. Additionally, a random effect for RT_{Behav} was nested within Trial within Participant to control for RT_{Behav} . Critically, each pairwise comparison was restricted to participants with mentalizing-active sites in both ROIs.

Whole-brain HFB responses within time-windows (multi-site)

To provide a broad overview of the neuronal spatiotemporal dynamics evoked by each task condition, we performed whole-brain analysis of HFB responses within specific time-windows (Fig. 1.6a). For each site and task condition, separate LMEMs were used to analyze trial-by-trial HFB power (z-scored). The intercept (null distribution) consisted of observations within the pre-stimulus baseline (-200–0 ms). Dummy variables consisted of observations within each time window. All model terms were specified as fixed and random effects nested within Trial, which accounted for trial-related heterogeneity. This specification estimates HFB responses evoked by each task condition within time windows ($p_{\text{FDR}} < .05$, corrected across sites and time windows).

Grand-average ROI timecourses (multi-site)

To summarize task-evoked neuronal dynamics within ROIs, we performed grand-averaged analysis (Fig. 1.6b-h) of trial-by-trial HFB timecourses (z-scored) from ROI sites that were identified as active or deactive for a given task condition by trial-averaged analysis (see Figs. 1.1f & 1.2). Separate LMEMs were used for each ROI and task condition. The intercept (null distribution) consisted of timepoints within the pre-stimulus baseline (-200–0 ms). Peri-stimulus timepoints (0–3000 ms) were represented as dummy variables. Model terms were nested within Site and Participant to account for site- and participant-related heterogeneity. This specification estimates grand-average HFB timecourses evoked by each task condition within ROIs ($p_{\text{FDR}} < .05$, corrected across ROIs and timepoints).

Paper II: A spectrum of neurocognitive processing for mentalizing

Abstract

Mentalizing reliably engages the default mode network (DMN), but the high-frequency broadband (HFB) activity emphasized by intracranial studies – and the hemodynamics emphasized by fMRI – capture only a fraction of the electrophysiological spectrum through which neuronal populations coordinate. Here we use intracranial EEG (iEEG) to characterize the full spatiotemporospectral architecture of mentalizing across 34 human participants performing judgments about themselves and others, alongside autobiographical memory and an arithmetic cognitive control task. We find that mentalizing and autobiographical memory engage a dual-phase architecture: an early, broadly synchronous theta phase, followed by a posterior-to-anterior high-frequency broadband cascade with late low-frequency deactivation. This architecture was shared near-identically between mentalizing and autobiographical memory and jointly distinguished them from arithmetic, which instead evoked sustained, anterior-weighted low-frequency activity. Self- and other-mentalizing engaged equivalent spectral magnitudes, differing only in small, sustained low-frequency effects consistent with the lengthier processing previously documented for other-mentalizing. Our results situate the HFB activation sequence established for mentalizing within a fuller spectral context, revealing an early synchronous phase that precedes and helps scaffold the content-specific cascade – organizational structure that HFB alone, and fMRI still less, cannot capture.

Introduction

In the previous study, we examined HFB activity, the closest electrophysiological correlate of hemodynamic responses. However, HFB is but one part of vast electrophysiological spectrum, in which frequency – the temporal rhythm of neuronal activity – distinguishes both physiology and function (Engel et al., 2001; Ward, 2003; Jensen & Colgin, 2007; Mendoza-Halliday et al., 2024). This is akin to the radio spectrum, where frequency differentiates equipment (e.g., walkie-talkie or Wi-Fi router) and function (e.g., voice or internet). Electrophysiological techniques can decompose the activity of frequencies across the spectrum, providing numerous distinct measures of neuronal processing for a given anatomical location. Here, we aim to build upon the neurocognitive and systems-level insights of our previous study (Tan et al., 2022; Paper I) by exploring the rich spectrum of neuronal processing in mentalizing.

The electrophysiological spectrum is divided into major bands (Fig. 2.1 & 4.1): delta (<4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-30 Hz), gamma (30-60 Hz), and high-frequency broadband (HFB; >60 Hz), the latter also called high gamma (Buzsáki, 2006). These bands play diverse neurocognitive and systems-level roles, which can vary by anatomical location and environmental context. In this paragraph, we summarize domain-general functions of these bands in DMN. Delta is thought to be the primary carrier of within-DMN synchronization, particularly for information integration at long timescales (Das et al., 2022; Das & Menon, 2024). Theta appears to regulate the timing of DMN activity by gating local excitability and intra-DMN signaling (Fox et al., 2018; Cona et al., 2020; Roehri et al., 2022). Alpha may instantiate attentional control of DMN processing by mediating interactions between DMN and frontoparietal control network, though effects are inconsistent (Marzetti et al., 2014; Samogin et al., 2019; Bowman et al., 2017; Cona et al., 2020). Beta may be the primary carrier of interactions

between DMN and other cortical networks, perhaps driving cross-network integration and segregation dynamics (Das et al., 2022). Gamma may carry feedforward signals into DMN while facilitating synchronization at intraregional and subnetwork-level scales. However, gamma functions remain nebulous as HFB activity often exhibits broadband extension into gamma frequencies in DMN. Lastly, HFB is unique from other bands in that it reflects local neuronal population spiking rather than coordinated signaling across spatially-distributed neuronal ensembles, both within and outside of DMN (Fox et al., 2018; Parvizi & Kastner, 2018).

A handful of studies have examined anatomically-resolved spectral activity in social cognition. A study on pragmatic social inferences (sarcasm/metaphor) found that TPJ produced robust HFB responses, but that delta and gamma best differentiated non-literal versus literal processing, highlighting concurrent low- and high-frequency signatures in a core DMN node (Soni et al., 2024). Similarly, theta-gamma coupling in dmPFC has been shown to index the efficiency of DMN coordination in social tasks (Friedrich et al., 2026). Another study reports that perspective-taking recruits theta activity in right TPJ, anterior cingulate, and lateral prefrontal cortex, with the latter two regions exerting top-down control over TPJ, indicating executive control over self and other perspectives (Seymour et al., 2018). This aligns with causal evidence for the importance of theta in right TPJ in embodied perspective-taking (H. Wang et al., 2016). Moreover, theta-enhancing neurofeedback and brain stimulation can result in greater mentalization and DMN connectivity (Imperator et al., 2017). These studies have tended to focus on specific brain regions, spectral bands, and timepoints. In the present study, we provide holistic spatiotemporal characterization of mentalizing-evoked activity across the electrophysiological spectrum. Moreover, we compare mentalizing-evoked spatiotemporal spectral activity patterns with those of autobiographical memory and arithmetic, cognitive functions with convergent and divergent functional anatomy relative to mentalizing.

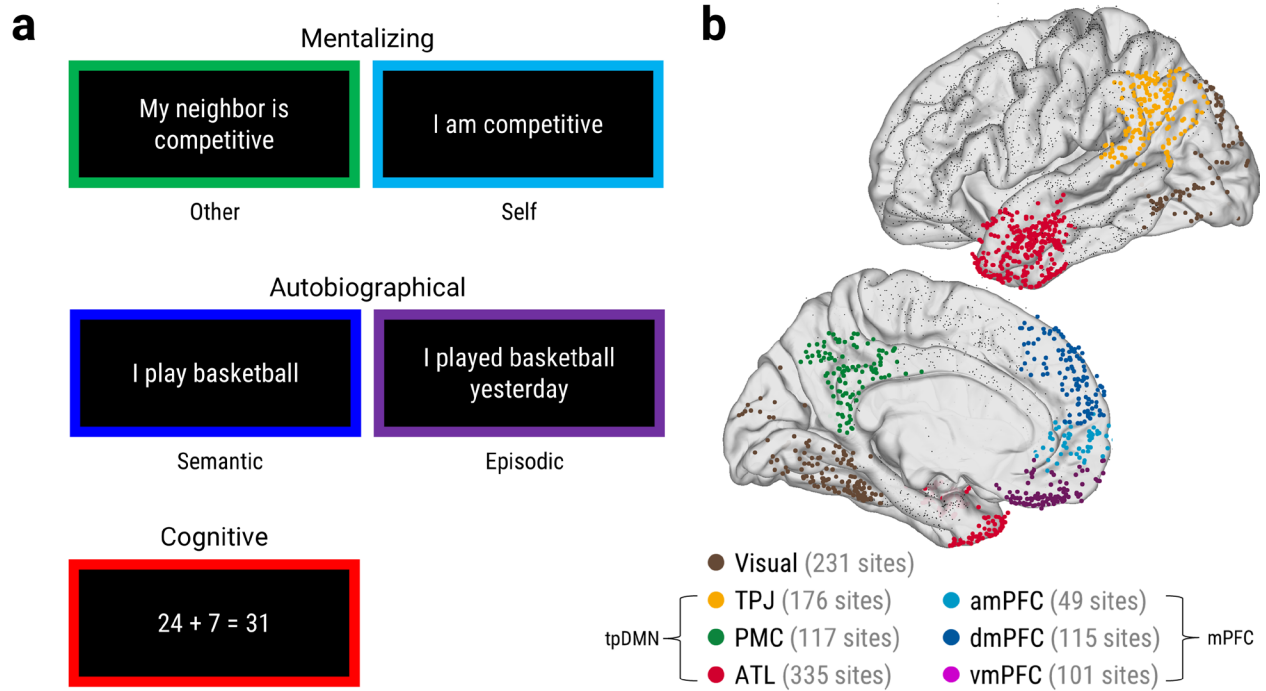


Figure 2.1. Data and design. **a)** Task stimuli. The task is the same as in Paper I with additional conditions. **b)** Regions-of-interest (ROIs) are the same as Paper I but with additional subjects and sites. ROIs were defined by the Glasser HCP atlas. ROIs are grouped as temporoparietal DMN (tpDMN: ATL, TPJ, PMC) and medial prefrontal cortex (mPFC: amPFC, dmPFC, vmPFC).

Results

Data and Design

We analyzed intracranial EEG (iEEG) recordings from 34 human participants who underwent invasive monitoring for drug-resistant epilepsy at Stanford University Medical Center (see Methods; Fig. 2.1). As in our prior work (Paper I), participants made true/false judgments about written prompts spanning multiple conditions; here we examined self- and other-mentalizing, semantic and episodic autobiographical memory, and an arithmetic cognitive control, alongside a cued rest condition. For analyses contrasting broad cognitive domains, semantic and episodic trials were pooled as autobiographical memory and self- and other-mentalizing as mentalizing.

To characterize neuronal activity across the electrophysiological spectrum, we decomposed each site's single-trial recordings into spectral power spanning delta through high-frequency broadband (HFB), and computed event-related spectral perturbations (ERSPs) relative to pre-stimulus baseline. We analyzed these dynamics at two spatial scales: within each of seven regions of interest – visual cortex (control), TPJ, ATL, PMC, amPFC, dmPFC, and vmPFC, defined through the Glasser Human Connectome Atlas – and across the whole brain at the single-site level. All analyses were corrected for multiple comparisons across timepoints, frequencies, and sites or ROIs (see Methods).

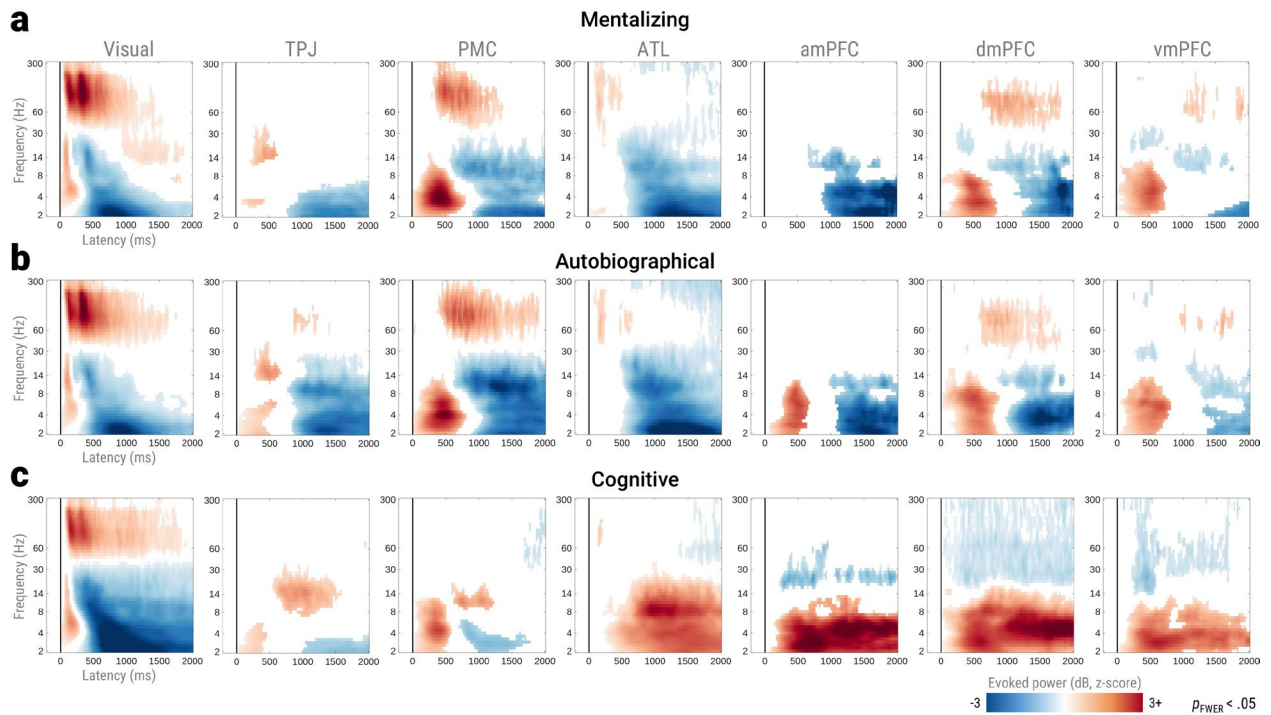


Figure 2.2: Event-related spectral perturbations (ERSPs) relative to prestimulus baseline, corrected for familywise error rate ($p_{\text{FWER}} < .05$) by threshold-free cluster enhancement across timepoints and frequencies. The canonical spectral band frequencies are demarcated. **a)** Mentalizing-evoked ERSPs (Other + Self). **b)** Autobiographical memory-evoked ERSPs (Semantic + Episodic). **c)** Cognitive-evoked ERSPs (arithmetic).

Mentalizing and autobiographical memory share a dual-phase architecture

We first characterized event-related spectral perturbations (ERSPs) relative to pre-stimulus baseline for mentalizing, autobiographical memory, and the cognitive control

(arithmetic) conditions across all ROIs (Fig. 2.2, $p_{\text{FWE}} < .05$, TFCE-corrected). Two temporally distinct phases were apparent across the DMN.

In the first phase, mentalizing and autobiographical memory evoked early low-frequency activation, concentrated in theta ($\sim 2\text{--}8$ Hz) and emerging at roughly 200–800 ms. This early theta activation was most pronounced in the medial regions – dmPFC, vmPFC, amPFC – and PMC, briefer in ATL, and weakest in TPJ. Because it appeared at similar latencies across these regions rather than propagating, this early phase had a broadly synchronous character.

In the second phase, mentalizing and autobiographical memory evoked high-frequency broadband activation (HFB; >60 Hz) whose onset followed a posterior-to-anterior gradient: immediate in visual cortex ($\sim 0\text{--}100$ ms), early in PMC ($\sim 100\text{--}300$ ms), and progressively later in anterior regions. This HFB cascade was accompanied by sustained low-frequency ($\sim 2\text{--}14$ Hz) deactivation developing from $\sim 700\text{--}1000$ ms onward across most of the DMN. The spatial progression of this second phase recapitulates, in the spectral domain, the posterior-to-anterior HFB activation sequence we previously established for mentalizing (Paper I).

Mentalizing and autobiographical memory produced strikingly similar ERSPs in every ROI – sharing both the early theta phase and the later HFB cascade – consistent with closely overlapping spectral processing. The cognitive condition, by contrast, was the spectral outlier. Rather than the phasic early-theta-then-HFB pattern of the mentalizing and autobiographical conditions, it evoked sustained low-frequency ($\sim 2\text{--}8$ Hz) activation throughout the trial together with sustained beta/gamma deactivation. The cognitive condition also showed an inverted spatial profile: it engaged anterior regions strongly (amPFC, dmPFC, vmPFC, ATL) but barely engaged PMC beyond the initial visual response, whereas mentalizing and autobiographical memory engaged the full posterior-to-anterior extent of the DMN.

Taken together, these results indicate that mentalizing and autobiographical memory share a dual-phase spatiotemporospectral architecture. This architecture is characterized by an

early, broadly spatiotemporally-synchronous theta phase followed by a posterior-to-anterior HFB cascade with late low-frequency deactivation. This dual-phase architecture is highly divergent from arithmetic's sustained, anterior-weighted low-frequency evoked activity.

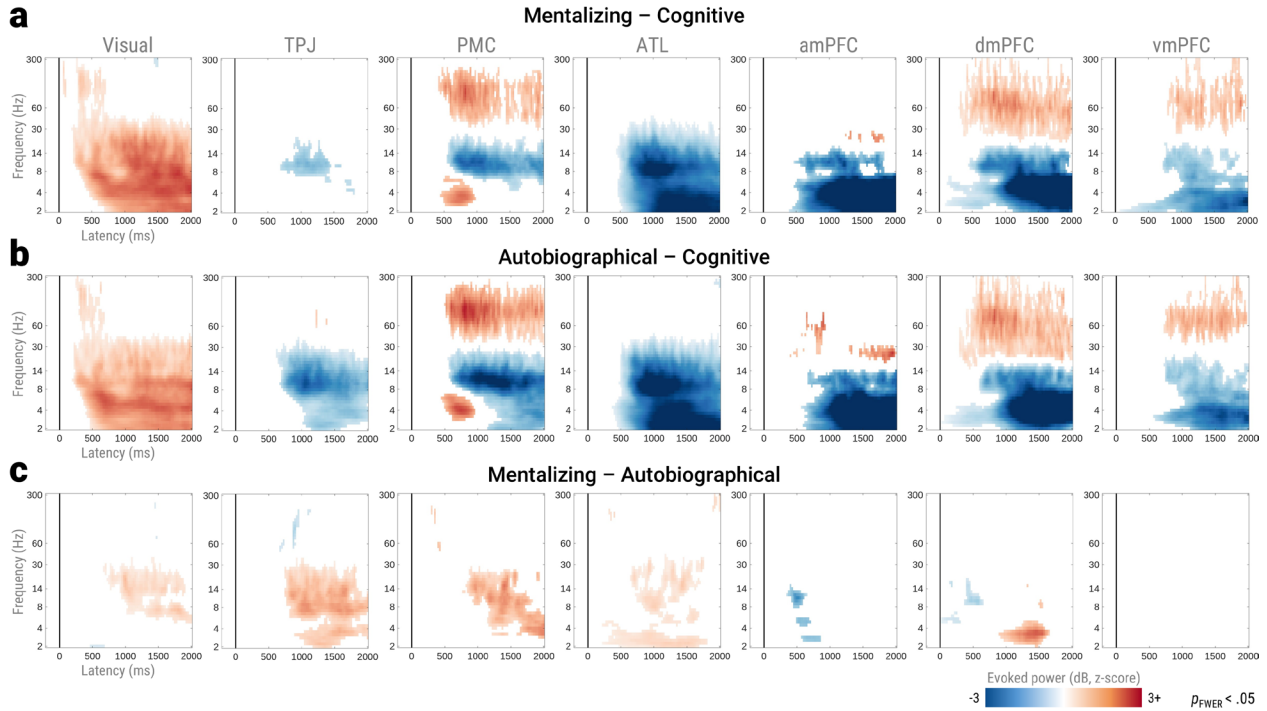


Figure 2.3: ERSP contrasts between task conditions, corrected for familywise error rate ($p_{FWER} < .05$) by threshold-free cluster enhancement across timepoints and frequencies. The canonical spectral band frequencies are demarcated. **a)** Mentalizing (Self + Other) versus Cognitive (arithmetic). **b)** Autobiographical memory (Semantic + Episodic) versus Cognitive. **c)** Mentalizing versus Autobiographical memory.

Mentalizing and autobiographical memory diverge from arithmetic but not each other

We next contrasted the task conditions directly (Fig. 2.3; $p_{FWER} < .05$, TFCE-corrected across timepoints and frequencies). Mentalizing and autobiographical memory each differed from the cognitive control (arithmetic) in a consistent, spatiotemporospectrally-structured manner. Critically, mentalizing and autobiographical memory differed from arithmetic in the same way: the Mentalizing–Cognitive and Autobio–Cognitive maps were nearly identical in every ROI.

This shared divergence from arithmetic had two components that varied in prominence along the posterior-to-anterior axis. In posterior and dorsomedial regions (PMC, dmPFC, vmPFC), mentalizing and autobiographical memory produced greater high-frequency broadband and gamma activity than arithmetic (warm high-frequency clusters), reflecting the HFB cascade absent in the cognitive task. In anterior and temporal regions (amPFC, ATL), mentalizing and autobiographical memory instead showed pronounced low-frequency deactivation relative to arithmetic (cool low-frequency clusters), reflecting the sustained low-frequency activation that arithmetic – but not the mentalizing or autobiographical conditions – evoked there. Across the DMN, then, the two phases of the mentalizing/autobiographical architecture (HFB cascade and low-frequency dynamics) are precisely what distinguish these conditions from arithmetic.

In direct contrast, mentalizing and autobiographical memory were spectrally near-indistinguishable. The Mentalizing–Autobio map was empty or near-empty in every ROI – entirely empty in vmPFC – with only sparse, low-magnitude clusters elsewhere. Thus, while mentalizing and autobiographical memory each diverged robustly from arithmetic, they did not reliably differ from one another at any frequency, latency, or region.

Taken together, these contrasts show that mentalizing and autobiographical memory evoke convergent spatiotemporospectral neuronal dynamics across DMN, jointly distinguished from arithmetic by the dual-phase architecture rather than by any feature separating them from each other.

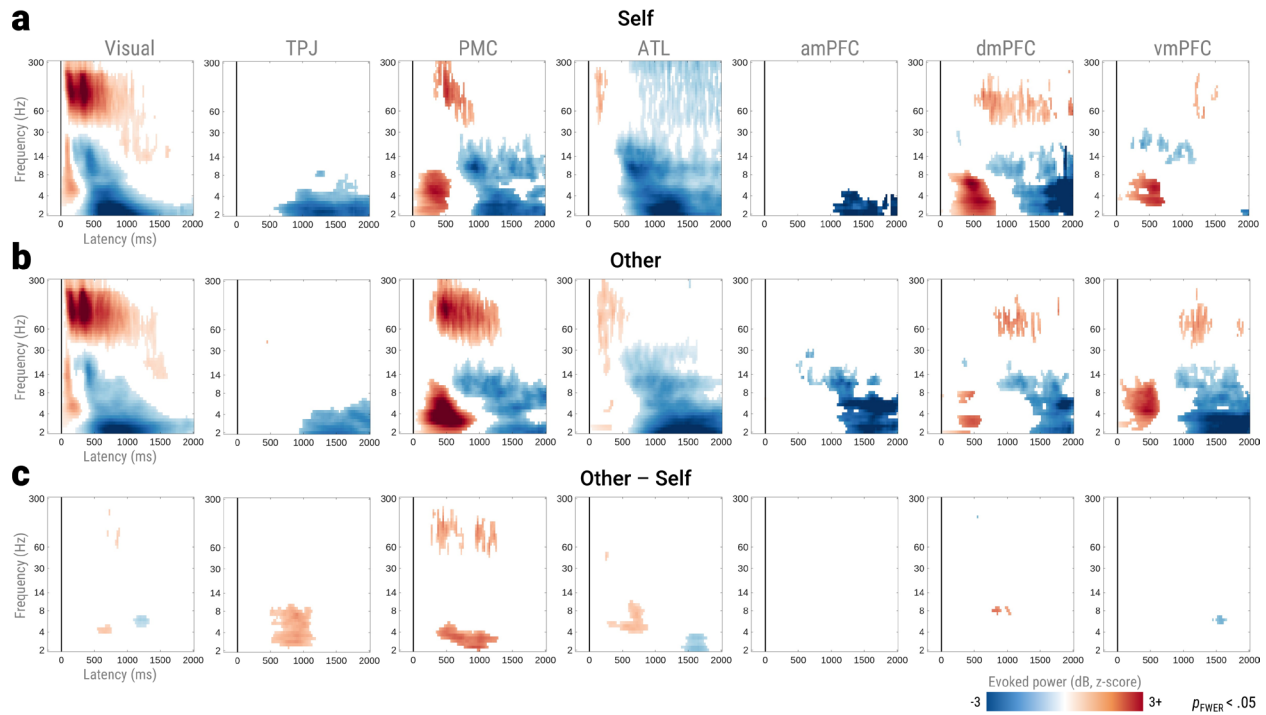


Figure 2.4: ERSPs for self- and other-mentalizing, corrected for familywise error rate ($p_{\text{FWER}} < .05$) by threshold-free cluster enhancement across timepoints and frequencies. The canonical spectral band frequencies are demarcated. **a)** Other-mentalizing versus prestimulus baseline. **b)** Self-mentalizing versus prestimulus baseline. **c)** Other-mentalizing versus self-mentalizing contrast.

Self- and other-mentalizing evoke near-identical spatiotemporospectral dynamics

Finally, we compared self- and other-mentalizing (Fig. 2.4; $p_{\text{FWER}} < .05$, TFCE-corrected across timepoints and frequencies). Self- and other-mentalizing produced closely matched ERSPs across all ROIs, each showing the same dual-phase architecture – early theta activation and a posterior-to-anterior HFB cascade with late low-frequency deactivation – described above.

The direct Other–Self contrast was largely null. Where significant differences did emerge, they were small, largely confined to low frequencies at mid-to-late latencies, and predominantly reflected greater power for other- than self-mentalizing: low-frequency theta in TPJ and ATL, along with low theta concurrent with HFB in PMC. The dmPFC showed only a minuscule high-theta/low-alpha cluster, and amPFC showed no significant differences. Late ATL and vmPFC delta and theta responses were the exceptions to the dominant direction, with small low-

frequency clusters of the opposite sign. No region showed a broad or sustained magnitude difference between self- and other-mentalizing.

This pattern bears directly on our prior account of self/other differentiation. In Paper I, we found that self- and other-mentalizing engage near-identical neuronal populations distinguished by activation timing – other-mentalizing evoking slower, lengthier activations – rather than by activation magnitude. The present contrast specifically tests magnitude and finds the two conditions largely equivalent. The few residual differences are consistent with the timing account rather than a counterpoint to it: modestly greater sustained low-frequency power for other-mentalizing at mid-to-late latencies is the expected spectral signature of the lengthier processing we previously documented, not evidence of a distinct magnitude-based representation.

Taken together, self- and other-mentalizing evoke near-identical spatiotemporospectral neuronal dynamics across the DMN, reinforcing our prior conclusion that they are distinguished primarily by timing rather than by the intensity or spectral content of neuronal activity.

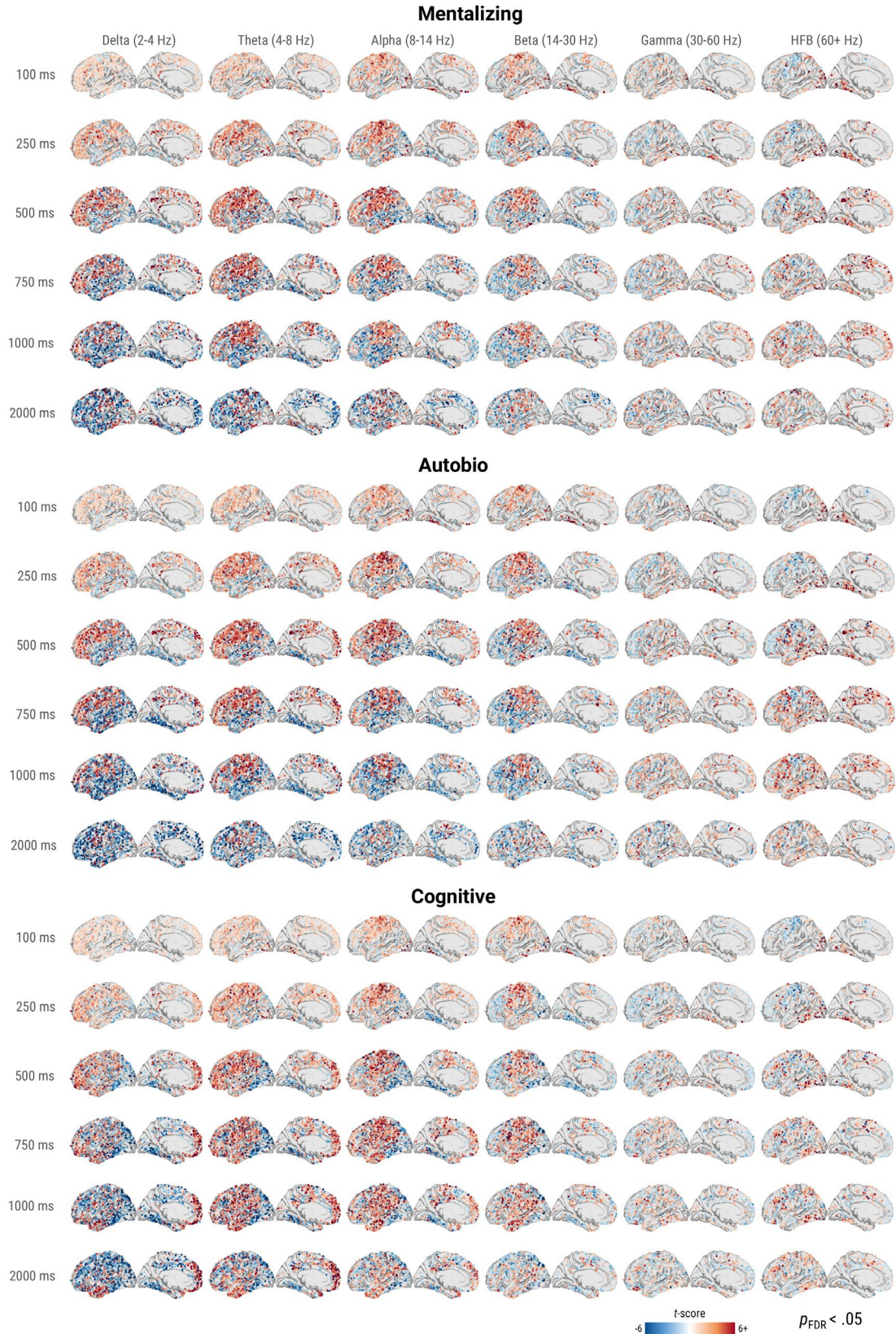


Figure 2.5: Whole-brain maps from single-site analyses for task conditions relative to prestimulus baseline. Sites only shown if $p_{FDR} < .05$, corrected across sites and timepoints.

Whole-brain spectral dynamics recapitulate the dual-phase architecture

To chart these dynamics across the whole brain, we mapped single-site evoked power within each spectral band at successive latencies (Fig. 2.5; $p_{\text{FDR}} < .05$, corrected across timepoints, bands, and sites). The whole-brain maps reproduced the dual-phase architecture observed in our ROI analyses. Early activation appeared in posterior cortex across higher-frequency bands and in theta, then spread anteriorly over the trial, while low-frequency (delta/theta) activation gave way to widespread low-frequency deactivation at later latencies – the spatial and spectral signature of the posterior-to-anterior cascade and late deactivation seen in the ROI ERSPs.

Consistent with the ROI results, mentalizing and autobiographical memory produced near-identical whole-brain maps across all bands and latencies. The arithmetic condition again diverged: rather than transitioning to late low-frequency deactivation, it sustained low-frequency (delta/theta) activation across the trial, extending into anterior cortex – mirroring the sustained, anterior-weighted low-frequency engagement seen in the ROI analyses.

Together, the whole-brain maps confirm that the dual-phase architecture – and the divergence of arithmetic from the mentalizing and autobiographical conditions – holds across the cortex and across the spectral bands.

Discussion

By decomposing mentalizing-evoked activity across the electrophysiological spectrum, we found that mentalizing and autobiographical memory engage a dual-phase spatiotemporospectral architecture: an early, broadly synchronous theta phase, followed by a posterior-to-anterior high-frequency broadband (HFB) cascade with late low-frequency deactivation. This architecture was shared near-identically between mentalizing and autobiographical memory, jointly distinguished them from arithmetic, and held across both ROI

and whole-brain analyses. By contrast, self- and other-mentalizing engaged equivalent spectral magnitudes, and the cognitive control task (arithmetic) diverged from the mentalizing and autobiographical conditions in both its spectral content and its spatial distribution.

The two phases appear to reflect different aspects of mentalizing-related neural activity. The early low-frequency phase appeared at similar latencies across DMN regions rather than propagating, giving it a synchronous character that contrasts with the sequential HFB cascade. Its spectral profile was broad rather than narrowband: an early peak in upper delta (~ 3 Hz) with a secondary peak in theta (~ 7 Hz), transient and shared near-identically across mentalizing, autobiographical memory, and the cognitive task. This breadth is consistent with the temporal-resolution prioritization of our wavelet decomposition (TimeBandwidth=20), which trades spectral for temporal precision, and may also reflect contributions from phase-locked slow evoked components that ERSP power alone cannot fully separate from oscillatory activity (Tallon-Baudry & Bertrand, 1999). A large literature ties low-frequency DMN activity to coordination and timing across regions (Fox et al., 2018; Cona et al., 2020; Roehri et al., 2022); the natural reading is that this early phase reflects a fast, shared engagement of the DMN before any region performs its specific computation.

We flag this functional interpretation as a framework we are applying rather than directly testing, since our data establish the timing, the spectral profile, and the cross-condition consistency of the early phase, but not its mechanistic role with respect to the subsequent cascade. The later HFB cascade, propagating from posterior to anterior cortex, carries the content-specific processing, consistent with HFB's role as a marker of local population spiking (Fox et al., 2018; Parvizi & Kastner, 2018) and its tight correspondence with the spatiotemporal sequence we documented for mentalizing using HFB alone (Paper I). The dual-phase architecture thus situates our prior HFB findings within a fuller spectral context: the posterior-to-

anterior activation sequence we previously described is preceded by, and embedded within, a synchronous low-frequency phase that an HFB-only analysis could not reveal.

Read as cognitive operations rather than signal, the dual-phase architecture invites a tentative correspondence with the two-process distinctions familiar from social cognition (Satpute & Lieberman, 2006; Frith & Frith, 2008). Speculatively, the early, shared, broad low-frequency phase may correspond to an automatic, content-general appraisal – the default mode network engaging the *mode* of social or autobiographical thought before computing the specific *content* – while the later, posterior-to-anterior HFB cascade carries the deliberate, content-specific inference. A more grounded instance of this temporal structure appears in our prior findings: TPJ produces earlier activations that are coactive across mentalizing and arithmetic and are tied to automatic psycho-semantic appraisal (Paper I), while dmPFC produces later activations that are mentalizing-specific and tied to explicit belief reasoning. The dual-phase architecture provides a spectral counterpart to that latency-based dissociation: automatic, content-general engagement first, deliberate, content-specific computation second, layered onto the same network. We offer this correspondence as an interpretive frame rather than a tested claim; demonstrating it directly would require analyses that link the early-phase activity to behavioral signatures of automatic processing and the late-phase activity to deliberate inference. We note that one such link is present in Paper III, where early-phase recruitment predicts faster behavioral responses, indicating that the early phase is functional rather than inert.

That mentalizing and autobiographical memory shared this architecture so closely – producing near-indistinguishable ERSPs across every region, band, and latency – provides spectral evidence for their overlapping neurocognitive basis (Andrews-Hanna et al., 2010; Mars et al., 2012; Buckner & DiNicola, 2019). This overlap motivates the central question of the next chapter: whether the shared spatiotemporospectral architecture reflects shared computation.

We note, however, that the near-identity of mentalizing and autobiographical memory in the present univariate analyses does not preclude finer differences between them; indeed, in Paper III we show that multivariate cross-classification resolves a reliable, graded distinction between mentalizing and autobiographical memory that the univariate contrasts here do not detect, illustrating the complementary sensitivity of the two approaches.

Arithmetic was the spectral outlier on two dimensions. Temporally, arithmetic evoked sustained low-frequency activation throughout the trial rather than the phasic early-then-cascade pattern of the mentalizing and autobiographical conditions. Spatially, it engaged anterior DMN regions strongly while barely engaging posteromedial cortex – an inverted profile relative to the mentalizing and autobiographical conditions, which engaged the full posterior-to-anterior extent of the DMN. This double divergence indicates that the dual-phase architecture is not a generic consequence of effortful cognition, but is specific to the mentalizing and autobiographical domains.

The sustained low-frequency activity in arithmetic was not a continuation of the early shared low-frequency phase. In the anterior DMN regions where it was most prominent (amPFC and dmPFC), the early broad low-frequency component (peaks at ~3 and ~7 Hz) was temporally separated from a later, sustained component centered in mid-theta (~5 Hz) by a window of reduced delta/theta activation (Figs. 2.2 & 2.3). The two components thus differ on three measured dimensions: timing (early transient versus late sustained), spectral profile (broad/bimodal versus narrower mid-theta), and condition-specificity (shared across conditions versus arithmetic-specific). The convergence across all three dimensions supports treating them as separable components rather than one continuous low-frequency process.

The late, sustained arithmetic theta is consistent with frontal-midline theta (FM θ), the canonical EEG/MEG signature of effortful cognitive control localized to dorsomedial frontal cortex (Cavanagh & Frank, 2014; Hsieh & Ranganath, 2014; M. X. Cohen, 2014). Its spatial profile

(concentrated in amPFC and dmPFC), spectral profile (sustained mid-theta), temporal profile (built and held across the presumed calculation period rather than near stimulus onset), and condition-specificity (driven by the effortful arithmetic task) jointly match the FM θ characterization in the cognitive control literature. Mental arithmetic is itself one of the historically established FM θ -evoking paradigms (Gevins et al., 1997; Ishii et al., 1999; Mizuhara et al., 2004). We identify our sustained arithmetic theta with FM θ at the level of its measured features and the established literature; we do not claim mechanistic dissociation from other theta phenomena, like the early shared theta described in the previous paragraph, as this would require formal component decomposition and cross-frequency analyses beyond the present work. Nonetheless, the temporal and functional separation between early shared theta and the late FM θ is suggestive of two separable components rather than a single continuous low-frequency process (Figs. 2.2, 2.3 & 2.5).

This identification has a direct consequence for how the standard fMRI account of DMN deactivation should be read. In dmPFC, the arithmetic condition produced a striking dissociation between bands: sustained HFB suppression in the same sites that produced sustained low-frequency activation, with the two effects overlapping in time across the late portion of the trial. Mentalizing produced the opposite pattern at the same sites – sustained HFB elevation accompanied by late low-frequency deactivation. The same dmPFC sites thus showed opposite signs in two bands depending on the task. Because BOLD couples most tightly to HFB and is largely blind to low-frequency power (Mukamel et al., 2005; Parvizi & Kastner, 2018), fMRI reads the HFB axis faithfully and reads the low-frequency axis not at all. The consequence is that the DMN deactivation widely reported during demanding cognitive tasks (Andrews-Hanna et al., 2010, 2014; Buckner & DiNicola, 2019) corresponds in our data to the HFB-suppression half of a dissociation in which the low-frequency half is moving in the opposite direction. The deactivation is real, in the sense that the HFB suppression is real and BOLD reads it faithfully;

what it is not is regional disengagement. The dmPFC during arithmetic is not quiet; it is producing a sustained FM θ -consistent control signal that hemodynamics cannot resolve.

The framework for this reconciliation has been in place for some time, but the spatial precision required to apply it to the DMN-deactivation literature arrived only recently. FM θ was characterized for decades in scalp EEG, where its midline-frontal topography lacked the localization needed to cross-reference against the specific DMN regions that fMRI was independently characterizing. Source-space EEG and MEG eventually localized FM θ to medial frontal cortex (Ishii et al., 1999; Asada et al., 1999; Onton et al., 2005; Meltzer et al., 2007), and subsequent intracranial recordings have corroborated theta-band generators in dACC and dmPFC during cognitive control (Raghavachari et al., 2001; C. Wang et al., 2005; M. X. Cohen et al., 2008; Pourtois et al., 2010; M. X. Cohen, 2014). The arithmetic-evoked late theta in the present intracranial work is consistent with the neuronal generators of FM θ in the source-space and intracranial work cited above. We accordingly view the present results not as overturning the canonical fMRI finding that the DMN deactivates during cognitive tasks, but as resolving an apparent contradiction between the DMN/fMRI and cognitive-control/electrophysiology literatures: the deactivation BOLD reports and the FM θ control signal the electrophysiology literature describes are, at least in dmPFC, two aspects of the same underlying activity, reported through instruments with complementary spectral sensitivity.

Our self/other comparison reinforces the account of mentalizing developed in Paper I. The direct contrast tested whether self- and other-mentalizing differ in the magnitude or spectral content of evoked activity, and found them largely equivalent across the DMN. The few residual differences were small, confined to low frequencies at mid-to-late latencies, and predominantly reflected greater sustained power for other-mentalizing – the expected spectral signature of the slower, lengthier processing we previously documented for other-mentalizing (Paper I), rather than a distinct magnitude-based representation. Thus, across both HFB (Paper

l) and the full spectrum (here), self- and other-mentalizing are distinguished by the timing rather than the intensity or spectral content of neuronal activity.

This study shares the limitations inherent to intracranial research, including the use of participants with epilepsy, variable electrode coverage across participants, and sampling bias toward cortical gyri (Parvizi & Kastner, 2018). Our characterization was also primarily qualitative: we described the spatiotemporal and spectral structure of evoked activity rather than formally modeling the interactions between phases or regions. Several specific limitations follow. First, our wavelet decomposition prioritizes temporal over spectral resolution (TimeBandwidth=20), which sharpens the timing of low-frequency events at the cost of fine frequency localization; the spectral peaks we report are approximate, and the broader low-frequency components in particular should be read as such. Second, ERSP power cannot distinguish induced oscillatory activity from phase-locked slow evoked components, an issue that especially constrains interpretation of the early, transient low-frequency phase; intertrial phase coherence and induced/evoked decomposition would be needed to adjudicate the oscillatory contribution (Canolty & Knight, 2010; Das & Menon, 2024). Third, the framework we apply to the low-frequency channel – coordination and control, with FM θ as the specific identification for the late arithmetic theta – is interpretive: we apply established characterizations from the oscillation and cognitive-control literatures rather than testing them in these data. Direct mechanistic claims about how the early phase relates to the subsequent cascade, or how the FM θ -consistent activity in arithmetic relates causally to behavior, would require cross-frequency coupling and directed connectivity analyses that we did not undertake here.

Neuronal oscillations are increasingly understood not as a single signal but as a set of partially distinct channels, each carrying different aspects of cortical computation and coordination across the frequency spectrum (Buzsáki & Draguhn, 2004; Siegel et al., 2012). Here we recover those channels in the social brain, revealing that mentalizing unfolds as a dual-phase

spatiotemporospectral architecture – shared with autobiographical memory, distinct from arithmetic – and that reading the full spectrum resolves an apparent contradiction between the fMRI and cognitive-control literatures, recasting medial-prefrontal "deactivation" as one half of a band-by-band dissociation. While the mechanisms linking these phases remain to be characterized, our findings establish that the frequency spectrum carries organizational structure in mentalizing that high-frequency activity alone, and fMRI still less, cannot reveal.

Methods

All research activities herein were conducted in accordance with protocol approved by the Stanford Institutional Review Board for human experimentation. All computational procedures and analyses herein were implemented in MATLAB unless otherwise specified. Preprocessing, machine learning and statistical analyses of electrophysiological data were performed using electroCUDA (<https://github.com/kevmtn/electroCUDA>) running on Nvidia GPUs with CUDA 13.1.

Participants

We employed a cohort of thirty-four human participants while they underwent neurosurgical treatment for drug-resistant epilepsy. Each participant provided written informed consent in accordance with the Stanford Institutional Review Board. Participants were not compensated as per Stanford Institutional Review Board guidelines for inpatient clinical research. As part of their presurgical evaluation, participants were implanted with iEEG at Stanford University Medical Center. The anatomical placement of electrode sites was determined according to each participant's clinical needs. Each participant was monitored in hospital for six to ten days prior to surgery, during which the study was conducted.

Behavioral Task

iEEG data was recorded while participants performed an event-related behavioral task with six conditions (trial types; Fig. 2.1a & Supplementary Table 1.4). Five of these conditions required true/false responses to written prompts, while one condition consisted of cued rest. Two conditions featured mentalizing prompts, either about oneself (e.g., “I am honest”) or others (e.g., “My neighbor is honest”). Participants were instructed to select a single neighbor (current or past) as the target for other-mentalizing. Cognitive task trials consisted of basic arithmetic (e.g., “9 + 86 = 95”). Two conditions featured memory-related prompts: episodic (e.g., “I ate candy yesterday”) and self-semantic (e.g., “I eat a lot of candy”). The cued rest condition required no response and displayed a fixation crosshair for 5-10 seconds. Stimuli were presented in random order and were self-paced, advancing to the next trial after the participants’ response, or up to 15 seconds if no response. The inter-trial interval (ITI) occurred -200-0 ms before each trial. The experiment was broken into two separate runs (mean run duration=12.50±1.64 minutes). Participants were allowed a short break between behavioral runs. On average, each run featured 25 trials of each sentence condition, 40 cognitive trials, and 36 rest trials. Each non-rest trial contained unique prompts; prompts were not repeated within participant. Responses were made via a handheld keypad using either the ‘1’ (true) or ‘2’ (false) key. Participants were instructed to perform the task as accurately and as quickly as possible. All stimuli were presented in white font on a black background (1200x800 pixels) using Psychophysics Toolbox 3 (Kleiner et al., 2007). For all analyses, trials were excluded if meeting any of the following criteria: high-frequency epileptic oscillations, no behavioral responses, irrelevant button presses, or RT_{Behav} under 400 ms.

iEEG data acquisition

Whereas Paper I used electrocorticography (ECoG) alone, the expanded cohort here additionally includes stereo-EEG (sEEG) depth electrodes, which sample sulcal, deep, and subcortical-adjacent structures (e.g., hippocampus) inaccessible to subdural grids and can record across cortical depth rather than the surface alone. We use “iEEG” to refer to both modalities throughout. iEEG recordings were obtained via 4128 subdural and depth electrodes. Electrodes (platinum plates with diameter of 1.2-2.3 mm) were implanted subdurally onto the cortical surface in grids or strips with center-to-center interelectrode spacing of 4-10 mm (Adtech Medical Instruments). Electrodes were connected to a multichannel recording system (Nihon Kohden; Tucker Davis Technologies) with sampling rate of 1,000 Hz or above. Anatomical data was acquired using a GE 3-Tesla SIGNA Magnetic Resonance Imaging (MRI) scanner at Stanford University. A T1-weighted anterior-posterior commissure-aligned pulse sequence was used. T1 data was resampled to 1 mm isotropic voxels, then segmented to distinguish gray and white matter using FreeSurfer (Fischl, 2012). To facilitate electrode localization, postimplant computerized tomography (CT) scans were coregistered to the preoperative MRI anatomical brain volume (Hermes et al., 2010). For each participant, electrodes sites were localized in BioImage Suite (Papademetris et al., 2006) and displayed on the participants’ own reconstructed 3D cortical surface using the iELVis toolbox (Groppe et al., 2017). Electrode positions were corrected for postimplantation brain shift, allowing for the accurate anatomical localization of electrodes sites (Dykstra et al., 2012).

Defining regions of interest (ROIs)

Because the expanded cohort included stereo-EEG (sEEG) depth electrodes sampling sulcal and deep cortex that surface-based native-space parcellation localizes poorly, electrode sites were assigned to ROIs using the Glasser Human Connectome Project multimodal

parcellation (HCP-MMP1.0; Glasser et al., 2016), projected from fsaverage onto each participant's native FreeSurfer cortical surface via folding-based registration, using a custom electrode-localization pipeline built on iELVis (Groppe et al., 2017). MNI coordinates were used as a secondary criterion to disambiguate parcels spanning ROI boundaries (specified below). Each site's HCP parcel was assigned to one of seven a priori ROIs – visual cortex (control), ATL, TPJ, PMC, amPFC, dmPFC, and vmPFC – which were organized into three ROI groups: Visual, temporoparietal DMN (tpDMN: ATL, TPJ, PMC), and medial prefrontal cortex (mPFC: dmPFC, amPFC, vmPFC). Parcel assignments were applied bilaterally and are given below without hemispheric prefixes.

Visual cortex: early visual parcels (V1, V2, V3, V3A, V3B, V4, V6, V6A, V7, V8, and the ventral VVC and VMV1), the MT+ complex (MT, MST, FST, V4t), and higher-order occipitotemporal parcels LO1–3, PIT, PH, and FFC. **ATL:** anterior temporal parcels STGa, TGd, TGv, STSda, STSva, TE1a, and TE2a (TE2a restricted to anterior sites, MNI $y > -19$). **TPJ:** inferior-parietal and temporoparietal-junction parcels PGs, PGi, PFm, STV, and TPOJ1–3, together with the posterior portion of PSL (MNI $y < -44$). **PMC:** precuneal, posterior-cingulate, and retrosplenial parcels POS1, POS2, 7Pm, 7m, PCV, v23ab, d23ab, 31pd, 31pv, 31a, and RSC. **amPFC:** a24, p32, and 10r, together with the ventral, near-midline portion of 10d (MNI $z \leq 10$ and $|x| < 12$). **dmPFC:** 9m, 9a, 9p, d32, a32pr, 8BM, and 8BL, together with the dorsal portion of 10d (MNI $z > 10$); all dmPFC parcels were restricted to anterior, near-midline sites (MNI $y > 28$ and $|x| < 12$). **vmPFC:** OFC, 25, s32, and 10v.

Sites were assigned to ROIs by this atlas; they were not excluded on the basis of anatomical gray/white tissue classification, which was unreliable for some depth contacts. Instead, site inclusion for each analysis followed a functional response criterion (≥ 100 ms of contiguous activation, $p_{FDR} < .05$, relative to the pre-stimulus baseline in at least one of the conditions entering that analysis; see Statistics), with a small number of sites reassigned after

manual inspection. This atlas-based approach differs from the native-space FreeSurfer (Desikan-Killiany-style) parcellation with manual inspection used in Paper I, which suited that ECoG-only cohort but does not provide the fine, deep-coverage parcels needed for the sEEG sites sampled here.

iEEG preprocessing

Preprocessing was performed using the electroCUDA package: robust electrophysiology tools with GPU acceleration (<https://github.com/kevmtn/electroCUDA>). Preprocessing was performed on a single-subject basis. First, electrodes marked as unusable in postclinical evaluation (pathological, loose contact, etc.) were discarded. Remaining problematic electrodes were identified using the ImaGIN classifier and other heuristics (Tuyisenge et al., 2018). This left 3827 electrodes for use in analyses.

Recordings from remaining electrodes underwent a computationally-driven robust denoising routine to maximize data quality. First, robust detrending was used to remove slow drifts from recordings (de Cheveigné & Arzounian, 2018). Afterwards, robust average referencing was applied with rank correction. Electrical line noise was then removed using Zapline Plus (Klug & Kloosterman, 2022). Next, non-stationary noise was removed using artifact subspace reconstruction (Chang et al., 2020). Stationary noise was then identified by independent components analysis and the ImaGIN classifier; artifactual components were subtracted from recordings.

Denoised recordings then underwent time-frequency decomposition via continuous wavelet transform with a Morse mother wavelet (TimeBandwidth = 20), applied to each continuous run to avoid epoching-related edge artifacts. We used a higher time-bandwidth product than the default (TimeBandwidth = 60) to prioritize temporal over spectral specificity, yielding 41 log-spaced frequencies spanning 5–300 Hz. Spectral magnitude was log-

transformed into decibel-scaled power. Spectral timecourses were epoched by trial and baseline-corrected by z-scoring each epoch to its pre-stimulus baseline (−200–0 ms). Region-of-interest event-related spectral perturbations (ERSPs) were computed from the full-resolution spectral timecourses. For whole-brain analyses, spectral timecourses were additionally averaged into six canonical bands – delta (2–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), gamma (30–60 Hz), and high-frequency broadband (HFB; 60–200 Hz) – to summarize spatial dynamics within each band.

Statistics

Statistical analyses described below were performed using the electroCUDA package (<https://github.com/kevmtn/electroCUDA>). Data was visualized using the GRAMM toolbox (Morel, 2018). All statistical tests were two-tailed. To correct for multiple comparisons across the time-frequency plane while preserving sensitivity to spatially and spectrally contiguous effects, we used threshold-free cluster enhancement (TFCE) with family-wise error rate control ($p_{\text{FWER}} < .05$) for analyses of event-related spectral perturbations. For band-based single-site analyses (used for whole-brain analysis), all multiple comparisons corrections maintained the False Discovery Rate (FDR) under 0.05 through the Benjamini-Yekutieli procedure for data with any dependence structure (Yekutieli & Benjamini, 1999), with p -values adjusted accordingly (p_{FDR}).

For single-site band-based analyses, we used linear mixed-effects modeling (LMEMs) due to their design flexibility and robustness to sampling heterogeneity, multicollinearity, and statistical confounds (Hox et al., 2017). To minimize estimation bias, mixed models used restricted maximum likelihood-based estimators (Hox et al., 2017). To account for full data dependence structure with reduced bias and assumptions, mixed models used unconstrained variance-covariance matrices with Cholesky parametrization (Barr et al., 2013). To account for

heterogenous variances across mixed model terms, Satterthwaite approximation for degrees of freedom (effective DF) was used (Gaylor & Hopper, 1969). To rectify violations of assumptions and overparameterization in mixed models, we evaluated objective function Hessian matrices and ensured positive definiteness (Schielzeth et al., 2020). Additionally, non-mixed models were used when specified below.

Within-site analyses

Within-site analyses were performed to provide the bases for whole-brain analysis, which plots all single-site results across the entire brain. The dependent variable for within-site analyses was spectral band power across timepoints and trials. Trials were excluded if containing high-frequency epileptic oscillations, no behavioral responses, irrelevant button presses, or RT_{Behav} under 400 ms. Timepoint-by-timepoint observations met outlier exclusion criteria if HFB power exceeded five median absolute deviations from other observations of the same timepoint and task condition.

ROI condition contrasts

To identify evoked spectral responses relative to prestimulus baseline for task conditions, along with differences between conditions, we contrasted ERSPs at each frequency and timepoint (TFCE-corrected, $p_{\text{FWER}} < .05$) across our ROIs. We performed three domain contrasts – mentalizing versus arithmetic, autobiographical memory versus arithmetic, and mentalizing versus autobiographical memory – and one within-mentalizing contrast, other- versus self-mentalizing.

Whole-brain analysis

To chart spectral dynamics across the cortex, we mapped single-site evoked power within each of the six canonical bands at successive peri-stimulus latencies (100, 250, 500, 750, 1000, and 2000 ms), for pooled mentalizing, autobiographical memory, and arithmetic. Each site's band power was tested against pre-stimulus baseline ($p_{\text{FDR}} < .05$, corrected across timepoints, spectral bands and sites) using LMEMs nested within trial, site, and subject.

Paper III: Autobiographical computations in mentalizing

Abstract

Mentalizing is widely assumed to recruit autobiographical computations such as self-projection and episodic recombination, given the near-indistinguishable functional anatomy of mentalizing and autobiographical memory within the default mode network (DMN). However, shared anatomy and overlapping activity cannot establish shared computation. Here we use intracranial EEG (iEEG) and spectral cross-classification to test this assumption directly in 34 human participants. We trained classifiers to distinguish autobiographical memory from an arithmetic cognitive control task, then applied them to held-out mentalizing trials, yielding a continuous index of autobiographical versus cognitive recruitment. Mentalizing engaged autobiographical-like spectral computation across the DMN, following the same posterior-to-anterior spatiotemporal gradient established for mentalizing-evoked activity. This recruitment was substantial but reliably distinct from full autobiographical engagement and was behaviorally consequential: autobiographical-like recruitment predicted faster responses at later latencies, and choice-predictive activity escalated along the hierarchy to converge on medial prefrontal cortex. Self- and other-mentalizing recruited autobiographical-like computation to an equivalent degree once stimulus visual dissimilarity was controlled, converging with prior evidence that they share a common neural code distinguished by timing rather than content. Together, these results provide evidence that mentalizing engages autobiographical-like computation and demonstrate that spectral cross-classification can resolve the computational content of mentalizing with anatomical and temporal precision unavailable to fMRI.

Introduction

Many forms of mentalizing are not computed solely by social cognitive processes (Schaafsma et al., 2015). Instead, mentalizing is thought to arise in part from autobiographical machinery, such as self-projection (Buckner & Carroll, 2007), scene construction (Mullally et al., 2012), and episodic recombination (Spreng et al., 2008). Mentalizing and these autobiographical processes feature near-indistinguishable functional anatomy, namely within DMN (Andrews-Hanna et al., 2010; Mars et al., 2012; Buckner & DiNicola, 2019). Indeed, in our previous study, we found closely overlapping spatiotemporal patterns of spectral activity between mentalizing and autobiographical memory (Paper II). Thus, it is commonly assumed that mentalizing recruits autobiographical computations. However, to our knowledge, this assumption has yet to be directly tested.

In extant literature, the strongest evidence for shared computations between mentalizing and autobiographical cognition comes from two fMRI studies that examine multivariate covariance (Spreng & Grady, 2010; Hughes et al., 2024). Both studies found that mentalizing and autobiographical memory produce common multivariate patterns of activation within DMN. Moreover, Hughes et al., 2024 also revealed common multivariate connectivity patterns across DMN subnetworks during mentalizing and autobiographical processing. Nevertheless, overlapping activity patterns alone cannot establish that mentalizing recruits autobiographical computations, as such overlap could equally reflect distinct but anatomically co-localized processes.

In the present study, we use spectral cross-classification to characterize the shared neurocomputation between mentalizing and autobiographical memory. Because the electrophysiological spectrum operates at the level of neuronal ensembles, activity at specific frequencies carries signatures of the specific computations those ensembles perform (Engel et al., 2001; Buzsáki & Draguhn, 2004; Buzsáki, 2006; Buzsáki et al., 2012). We exploit this

specificity by training classifiers to distinguish autobiographical memory from a cognitive control task (arithmetic) – two domains with convergent and divergent functional anatomy relative to mentalizing – and applying these classifiers to held-out mentalizing trials. The resulting posterior probabilities index the degree to which mentalizing engages autobiographical-like versus cognitive-like spectral computations, providing a continuous, anatomically- and temporally-resolved measure of autobiographical recruitment during mentalizing. Critically, we chart the spatiotemporal trajectory of this recruitment across the DMN, and ask whether self- and other-mentalizing recruit autobiographical computations differently.

This approach also connects to our prior account of self- and other-mentalizing (Paper I), in which both engaged near-identical neuronal populations distinguished by the timing rather than the magnitude or content of their activity. If self- and other-mentalizing share a common neural code, they should recruit autobiographical computations to a similar degree, differing – if at all – in when rather than how much. We test this prediction directly. Finally, although we focus on autobiographical memory as a whole, we note that finer distinctions within autobiographical cognition – such as between semantic and episodic retrieval – may require greater statistical power than the present design affords, and we return to this in the Discussion.

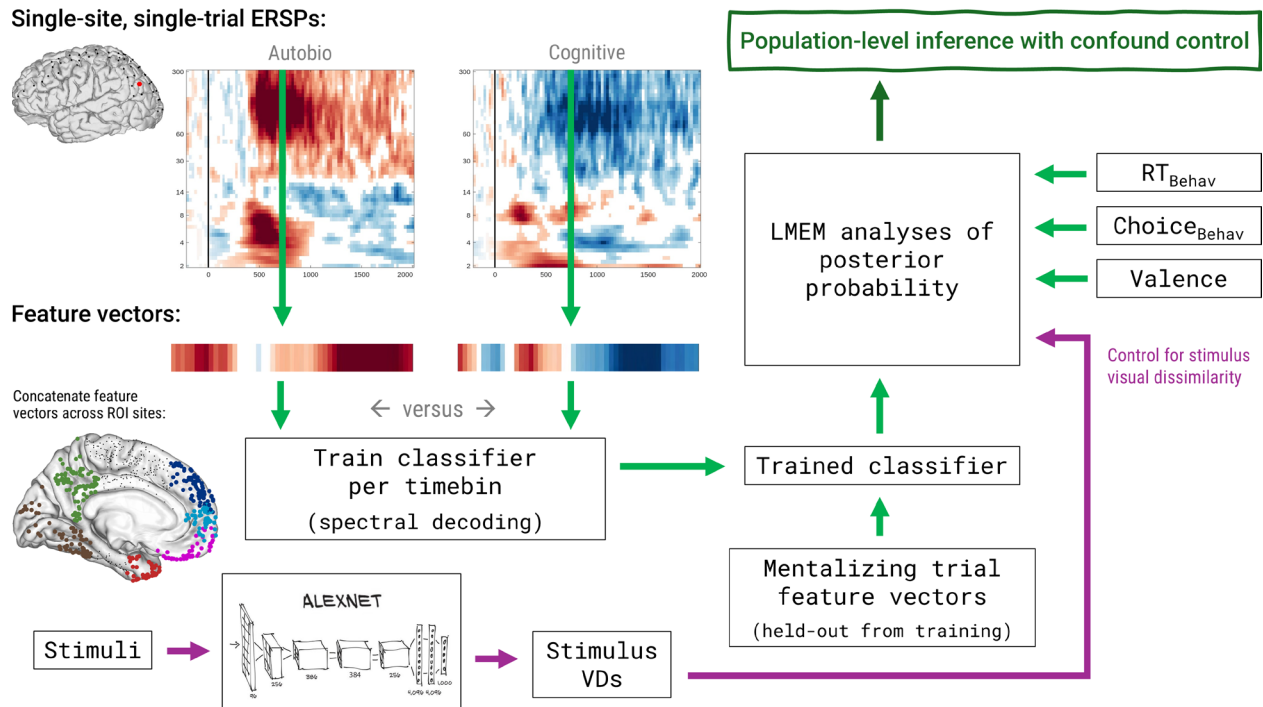


Figure 3.1: Graphical abstract of the analytical design. Features are evoked spectral magnitudes in 41 frequencies (5-300 Hz) from each ROI site, which were concatenated across sites within an ROI per observation. Classification was performed independently per ROI and timebin. Classification used LDA with nested hyperparameter tuning of shrinkage and sparsity. AlexNet was used to control for stimulus visual dissimilarity in LMEM analyses of classifier posterior probabilities.

Results

Data and design

We analyzed iEEG (intracranial electroencephalography) recordings from 34 human participants who underwent invasive monitoring for drug-resistant epilepsy at Stanford University Medical Center (see Methods). As in our prior work (Papers I–II), participants judged true/false text prompts across multiple conditions; here we focus on four: self-mentalizing, other-mentalizing, semantic-autobiographical, and episodic-autobiographical, along with an arithmetic (cognitive) condition that served as an externally-oriented executive function contrast.

Our central analysis asked whether mentalizing recruits autobiographical computations by testing whether mentalizing-evoked spectral activity resembles that of autobiographical

memory more than that of a cognitive control task (arithmetic). To this end, we trained classifiers to discriminate autobiographical from cognitive trials and applied them to held-out mentalizing trials (Fig. 3.1). The autobiographical class pooled semantic and episodic trials, capturing autobiographical retrieval broadly rather than either subtype alone. Classification was performed at each timebin independently, yielding timecourses of classification accuracy (for the trained autobiographical-versus-cognitive contrast) and posterior probability (for the trained and held-out mentalizing trials), at both single-site and ROI spatial scales.

Features consisted of single-trial spectral magnitude across 41 frequencies (5–300 Hz) at each timepoint (Fig. 3.1). ROI classifiers concatenated these spectral features across all sites within an ROI. We analyzed the same seven regions of interest as in our prior work – visual cortex (control), TPJ, ATL, PMC, amPFC, dmPFC, and vmPFC – assigned to Glasser HCP atlas parcels on each participant's native cortical surface (see Methods).

To ensure that classification reflected cognitive rather than low-level visual differences between conditions, all posterior-probability analyses included two visual dimensions (VD1 and VD2) as nuisance covariates. These dimensions were derived from AlexNet, a deep learning image classifier (see Methods). AlexNet captured low-level visual properties such as stimulus extent – VD1 was strongly correlated with prompt length, and in the two-dimensional visual-feature space the arithmetic condition separated from the sentence conditions while self- and other-mentalizing, semantic, and episodic prompts intermingled (see Supplementary Fig. 3.1) – together indicating that the visual dimensions primarily captured low-level stimulus properties rather than semantic content.

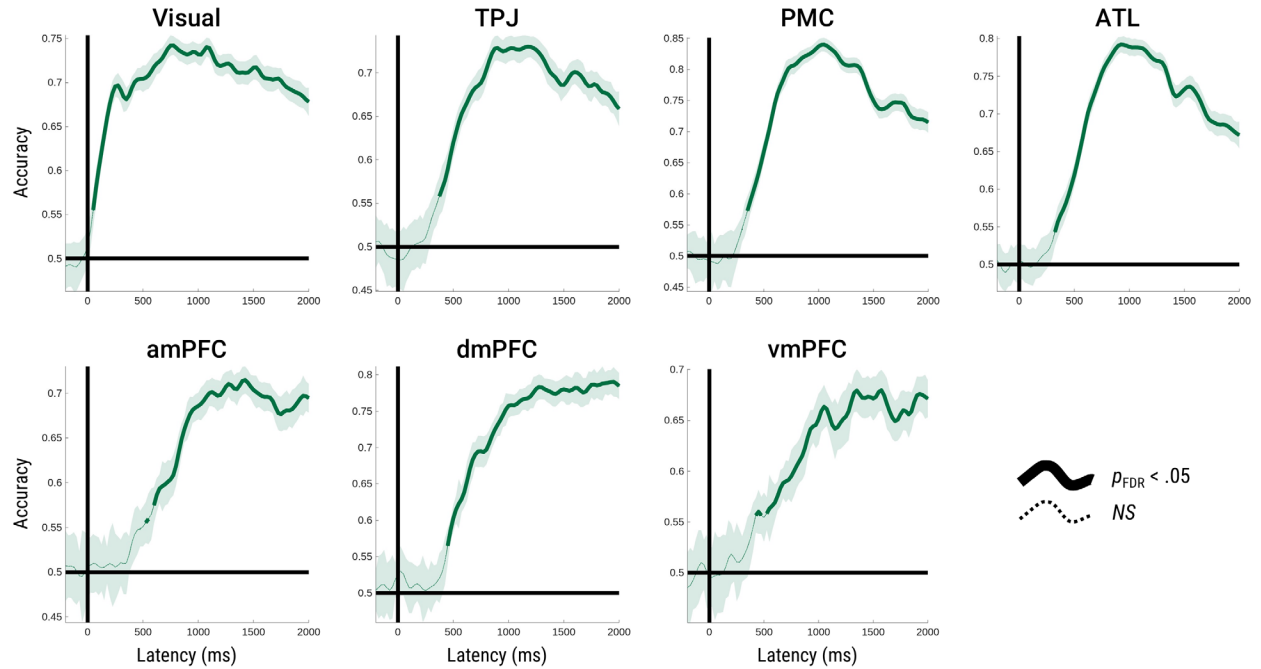


Figure 3.2: ROI classification accuracy for autobiographical versus cognitive task conditions. Accuracy was measured by logistic mixed-effects models nested within subject within fold within class within trial. This provides population-level estimates of accuracy while accounting for variance across subjects, folds, classes, and trials. Thick solid lines indicate significant timepoints ($p_{FDR} < .05$, corrected across timebins and ROIs); thin dotted lines indicate nonsignificant (NS) timepoints. Shaded areas indicate standard error.

ROI classifier accuracy

All our ROIs achieved significantly above-chance accuracy ($p_{FDR} < 0.05$, corrected across timebins and ROIs) when we decoded the autobiographical versus cognitive task conditions from evoked spectral responses. However, the temporal dynamics of classification performance differed markedly between them (Fig. 3.2). Unsurprisingly, visual cortex was the first ROI to produce significantly above-chance accuracy, with an onset latency of 50 ms and peak latency of 775 ms when it achieved 74.2% accuracy. Like the rest of our ROIs, visual cortex sustained significantly above-chance accuracy until the end of the analysis timeframe (2000 ms), therefore offset latencies could not be assessed. Visual cortex was succeeded by the tpDMN ROIs, the first of which was PMC with an onset latency of 300 ms and peak latency of 1050 ms, at which time PMC achieved the highest accuracy (85.1%) of any ROI. PMC was quickly succeeded by ATL, which had an onset latency of 325 ms and peak latency of 925 ms at 79.3%

accuracy. TPJ achieved significance soon thereafter with an onset latency of 375 ms and peak latency of 1175 ms at 73.0% accuracy. The tpDMN ROIs were succeeded by the mPFC ROIs, the first of which was vmPFC with an onset latency of 425 ms and peak latency of 1575 ms at 67.9% accuracy, which somewhat surprisingly was the lowest peak accuracy among our ROIs. Next came dmPFC with an onset latency of 450 ms and peak latency of 1950 ms (79.1% accuracy), which was the latest peak latency among our ROIs. The last ROI to achieve significantly above-chance accuracy was amPFC, with an onset latency of 475 ms and peak latency of 1425 ms at 71.5% accuracy.

Taken together, we found a spatiotemporal gradient of spectral classification performance for autobiographical memory versus arithmetic from visual cortex to amPFC. When considering both onset and peak latencies, our ROIs can be subdivided into three tiers: visual cortex with the earliest significantly above-chance accuracies, then the tpDMN ROIs, and lastly the mPFC ROIs. These tiers are further supported by the shape of the accuracy timecourses (Fig. 3.2). Visual cortex produced a nearly-vertical rise in classification accuracy almost immediately after stimulus onset. The tpDMN ROIs produced flat, near-chance accuracies until around 250 ms, roughly plateaued around their peaks between 800 and 1200 ms, then steadily declined afterwards. In contrast, the mPFC ROIs stayed at near-chance accuracy for the longest, had the most gradual ascent in accuracy, and sustained near-peak accuracy until the end of the analysis period (2000 ms).

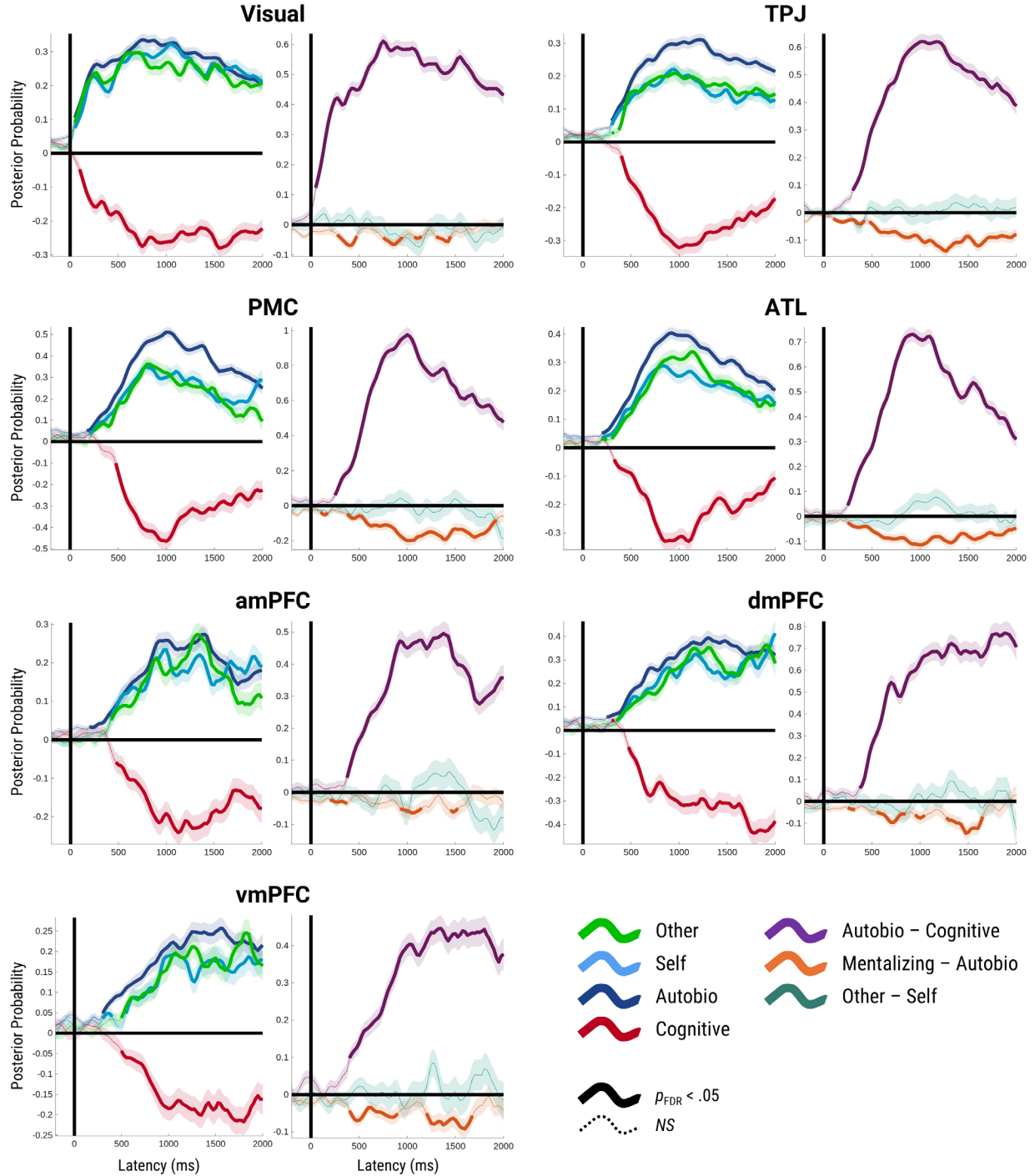


Figure 3.3: ROI posterior probabilities for each task condition. Posterior probabilities (PPs) were derived from the autobiographical-versus-cognitive classifiers (Fig. 3.1) and applied to held-out self- and other-mentalizing trials, providing a continuous proxy for the recruitment of autobiographical- versus cognitive-like (arithmetic) computations during mentalizing. Positive PPs indicate autobiographical-like spectral patterns, negative PPs indicate cognitive-like patterns. For each ROI, the **left panel** shows PP timecourses for the four task conditions and the **right panel** shows PP condition contrasts. PPs were estimated with unified linear mixed-effects models (LMEMs) jointly representing all four conditions, nested within subject within trial, with two visual dimensions (VD1, VD2) included as nuisance covariates to control for stimulus visual dissimilarity (see Fig. 3.1 and Methods). Cognitive and Autobiographical PPs are derived from cross-validated classifiers. PPs for self- and other-mentalizing are derived from classifiers trained on the full training set. Thick solid lines indicate significant timebins ($p_{FDR} < .05$, corrected across timebins and ROIs); thin dotted lines indicate nonsignificant (NS) timebins. Shaded areas indicate standard error.

Posterior probabilities (PPs) were derived from the autobiographical-versus-cognitive classifiers and then applied to held-out self- and other-mentalizing trials, providing a continuous proxy for the recruitment of autobiographical versus cognitive computations during mentalizing (Fig. 3.3). To obtain valid population-level estimates, PP timecourses were modeled with unified linear mixed-effects models (LMEMs) that jointly represented all four conditions (autobiographical, cognitive, self-, and other-mentalizing), nested within subject within trial to account for between-subject and trial-by-trial-variance. Two visual dimensions (VD1 and VD2; see Fig. 3.1, Data and Design, and Methods) were included as nuisance covariates to control for stimulus visual dissimilarity (e.g., prompt length). Positive PPs indicate autobiographical-like spectral patterns, negative PPs indicate cognitive-like (arithmetic) patterns, and PPs near zero indicate patterns equidistant from the two trained conditions; PPs range from -1 to $+1$, such that the maximum possible condition difference is 2.

As a manipulation check, we first confirmed that the trained conditions separated as expected: autobiographical trials produced robustly positive PPs and cognitive trials produced robustly negative PPs across all ROIs (Autobio–Cognitive contrast; $p_{\text{FDR}} < .05$, corrected across timebins and ROIs). This separation was largest in PMC, which produced a peak contrast estimate of ~ 1.0 (half the maximum possible difference of 2). Slightly less separation was produced by dmPFC and ATL (~ 0.75), with the rest of the DMN ROIs still producing substantial separation (e.g., TPJ at ~ 0.6).

Critically, both self- and other-mentalizing produced positive PPs across all ROIs, indicating that mentalizing recruits autobiographical-like spectral patterns in all ROIs. The magnitude of these PPs followed a consistent gradient: autobiographical trials produced the highest PPs, followed by self- and other-mentalizing at intermediate values, with cognitive trials at the negative extreme (Autobio > Self $\approx$ Other > Cognitive). This gradient was evident across all ROIs with the strongest differences in PMC and dmPFC.

Although mentalizing produced clearly autobiographical-like PPs, these patterns remained reliably distinct from those of autobiographical memory itself. The Mentalizing–Autobio contrast was significantly negative across all DMN ROIs ($p_{\text{FDR}} < .05$), indicating that mentalizing recruited autobiographical-like patterns but at a significant distance from full autobiographical engagement. This distinction was robust and sustained throughout the DMN, particularly in the high-accuracy ROIs (PMC, ATL, and dmPFC). The temporal emergence of these effects was also informative: autobiographical and cognitive PPs separated early (~250–500 ms), whereas mentalizing PPs remained near zero before climbing toward autobiographical-like values later in the trial. This sequence is consistent with autobiographical computations being recruited into mentalizing rather than driving it from stimulus onset.

In contrast to the robust Mentalizing–Autobio distinction, self- and other-mentalizing were statistically indistinguishable along the autobiographical-cognitive axis once stimulus visual dissimilarity was accounted for. Before visual control, self/other PP differences were robust in visual cortex, beginning at ~150 ms and sustained throughout the trial. PMC and ATL also showed robust, sustained self/other differences at mid-range latencies, though smaller in magnitude than visual cortex; the remaining DMN ROIs showed only brief, intermittent differences. The early onset and persistence of the visual-cortex effect is characteristic of low-level stimulus differences (e.g., the greater prompt length of other-mentalizing trials) rather than differential cognitive processing. Accordingly, after including the two visual dimensions (VD1 and VD2) as covariates, the Other–Self contrast was non-significant across all ROIs, indicating that residual self/other differences in PP were attributable to visual confounds rather than to differential recruitment of autobiographical computations. This convergence echoes our prior finding that self- and other-mentalizing engage near-identical neuronal populations distinguished by activation timing rather than magnitude or computational content (Paper I).

Taken together, these results indicate that mentalizing recruits autobiographical-like computations across DMN ROIs. The PP gradient (Autobio > Self $\approx$ Other > Cognitive) quantifies this recruitment: mentalizing engages autobiographical-like spectral patterns at substantial magnitude but reliably short of full autobiographical representation. The absence of self/other differences along this axis – once visual dissimilarity is controlled – indicates that both forms of mentalizing draw upon a common autobiographical substrate.

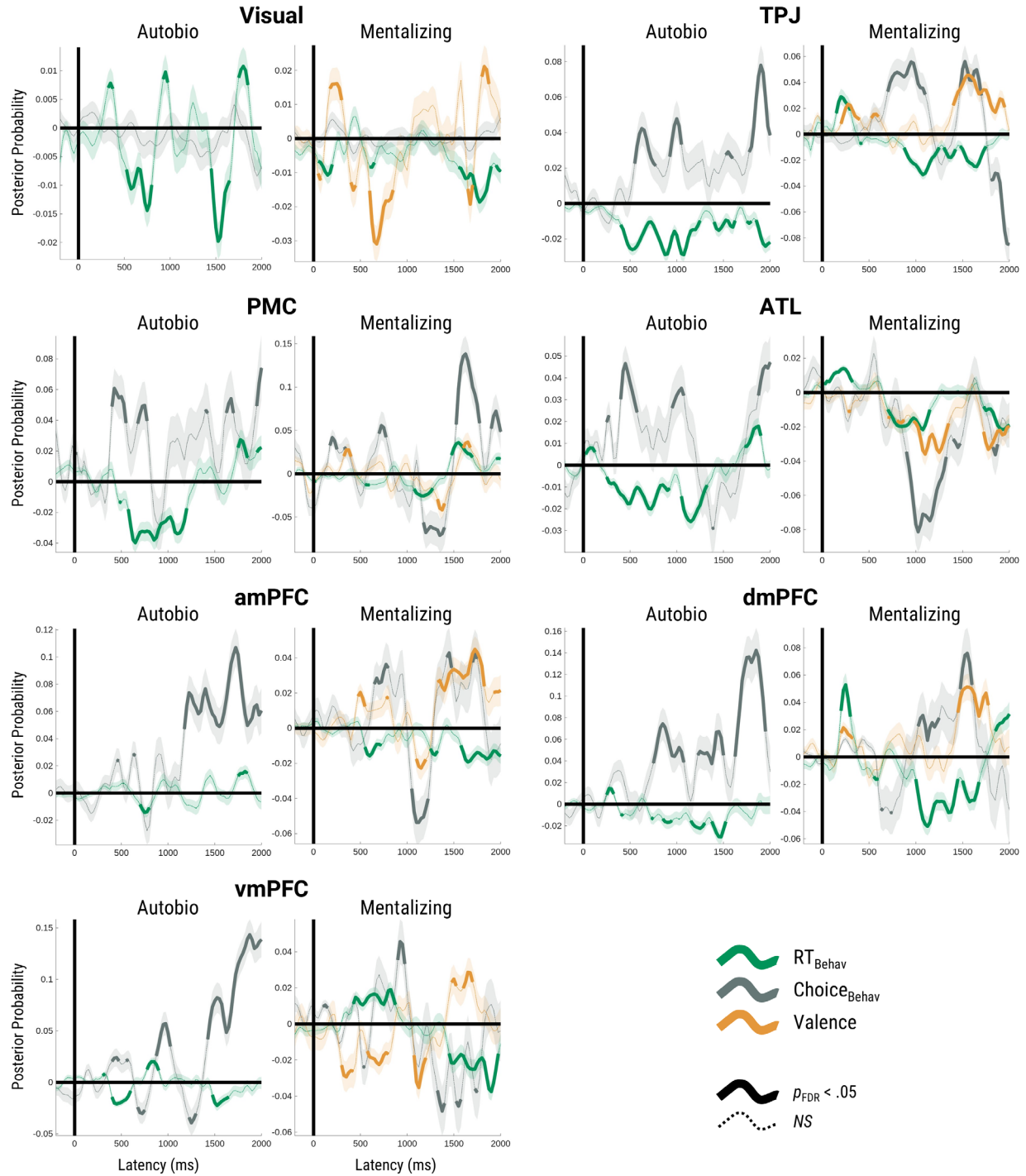


Figure 3.4: Psychobehavioral relevance of ROI PPs. Associations between PPs and behavioral response time (RT_{Behav} , green), behavioral response choice ($\text{Choice}_{\text{Behav}}$, true/false; gray), and affective valence (negative/neutral/positive; orange) were assessed. Since positive and negative PPs reflect autobiographical-like and cognitive-like spectral patterns, respectively (Fig. 3.3), the sign of each trace indicates whether higher values of a psychobehavioral metric tracked more autobiographical-like (positive) or more cognitive-like (negative) recruitment (e.g., positive RT_{Behav} associations denote slower responses for autobiographical-like computations). For each ROI, the **left panel** shows associations for autobiographical trials and the **right panel** for mentalizing trials. Valence associations were estimated for mentalizing trials only, as autobiographical stimuli were affectively neutral. Associations were tested via LMEMs, controlling for original task condition (Cond; Fig. 2.1) and stimulus visual dissimilarity, nested within subject within Cond within trial. Thick solid lines indicate significant timebins ($p_{\text{FDR}} < .05$, corrected across timebins and ROIs); thin dotted lines indicate nonsignificant (NS) timebins. Shaded areas indicate standard error.

Psychobehavioral relevance

To assess the psychobehavioral relevance of autobiographical recruitment, we analyzed behavioral response time (RT_{Behav}), behavioral response choice ($\text{Choice}_{\text{Behav}}$; true vs. false), and affective valence (negative/neutral/positive) as predictors of PP timecourses using independent LMEMs, each controlling for the original task condition (Cond: Semantic, Episodic, Self, and Other) and the visual dimensions (VD1, VD2), nested within subject within Cond within trial. RT_{Behav} and $\text{Choice}_{\text{Behav}}$ were estimated separately for autobiographical (Semantic+Episodic) and mentalizing (Self+Other) trials. Valence was estimated for mentalizing only, as autobiographical stimuli were affectively neutral.

RT_{Behav} showed a striking temporal reversal during mentalizing (Fig. 3.4). At early latencies, cognitive-like recruitment predicted faster behavioral responses across most DMN ROIs (TPJ, ATL, dmPFC, and vmPFC): trials whose spectral patterns leaned toward the cognitive pole early in the trial were answered more quickly. At mid-to-late latencies, this relationship reversed to the expected direction, with autobiographical-like recruitment predicting faster behavioral responses. This early cognitive-like advantage is consistent with a fast, domain-general engagement of the DMN that scaffolds subsequent content-specific processing (see Paper II). For autobiographical trials, RT_{Behav} effects were more uniformly in the expected direction, with autobiographical-like recruitment predicting faster behavioral responses across the DMN.

$\text{Choice}_{\text{Behav}}$ was noisier but revealed a clear spatial gradient that recapitulated the recruitment hierarchy seen in classification accuracy and the PP-by-condition gradients. Visual cortex showed essentially no $\text{Choice}_{\text{Behav}}$ effects, the tpDMN ROIs showed intermediate effects, and the mPFC ROIs showed the most robust, consistent, and late-onset effects. This gradient was clearest for autobiographical trials. During mentalizing, dmPFC showed the most robust $\text{Choice}_{\text{Behav}}$ effects of any ROI. The late onset of mPFC choice effects – building through the

latter half of the trial – is consistent with these regions carrying decision-related information that converges as a behavioral response is formed.

Affective valence modulated autobiographical recruitment during mentalizing in a largely consistent direction: across most DMN ROIs, positive valence was associated with more autobiographical-like spectral patterns. Two regions departed from this pattern. ATL showed the opposite relationship (positive valence associated with less autobiographical-like recruitment), and vmPFC showed an early reversal that flipped to the dominant direction (positive = more autobiographical-like) at later latencies. Notably, visual cortex also tracked valence despite carrying no mentalizing-specific computation, a pattern we consider further in the Discussion.

Taken together, these psychobehavioral analyses indicate that autobiographical recruitment during mentalizing is behaviorally consequential. The temporal reversal in RT_{Behav} effects suggests an early domain-general contribution followed by a later autobiographical one. The posterior-to-anterior gradient in $\text{Choice}_{\text{Behav}}$ effects recapitulates the recruitment hierarchy and converges on mPFC near the behavioral response, with valence modulating autobiographical recruitment across most of the DMN.

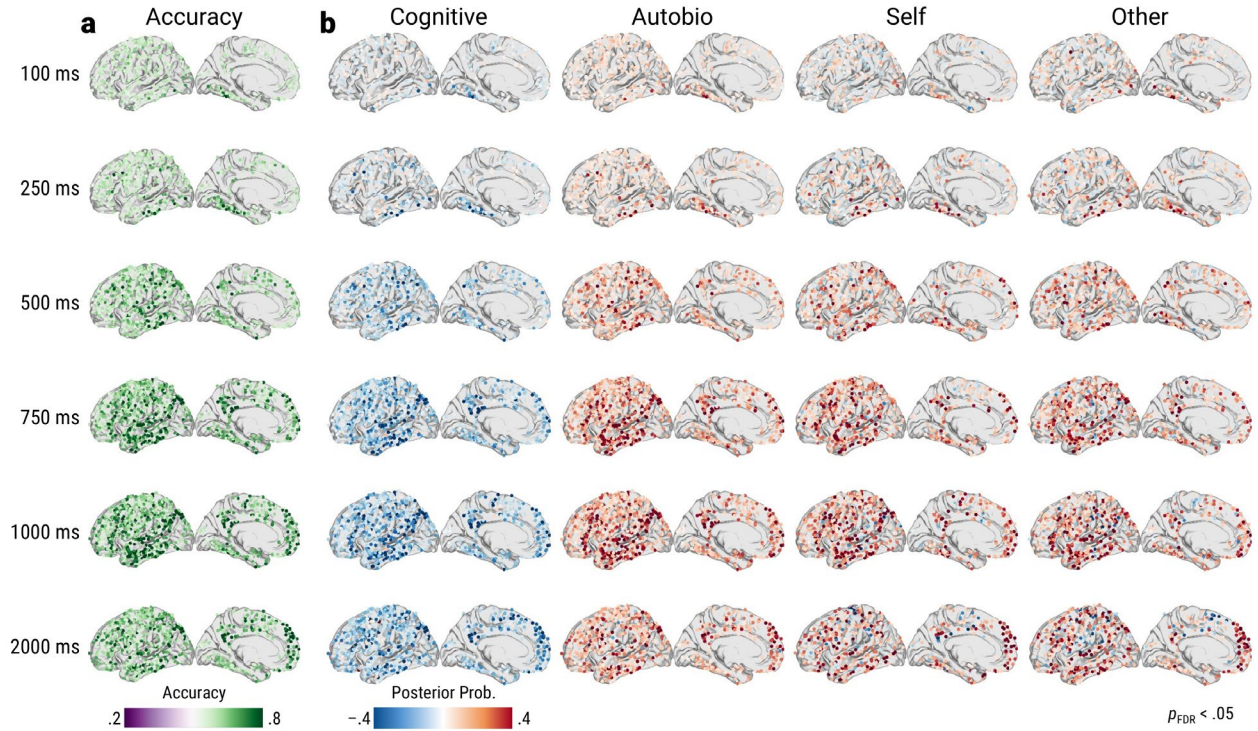


Figure 3.5: Whole-brain results. These results are obtained from single-site cross-classification analyses. **a)** Cross-validated accuracy for the autobiographical vs. cognitive training set. **b)** Posterior probabilities (PPs) per condition. Cognitive and Autobiographical PPs are derived from cross-validated classifiers. PPs for self- and other-mentalizing are derived from classifiers trained on the full training set. For each brain map, sites are only shown if $p_{FDR} < .05$, corrected across sites and timebins.

Whole-brain results

Finally, we mapped cross-classification across the whole brain to chart its spatiotemporal progression at the single-site level (Fig. 3.5). Whole-brain classifiers used the same procedure as the ROI classifiers but were trained on individual sites without nested cross-validation (see Methods); ROI classifiers, by contrast, concatenated features across sites within each ROI. All effects were corrected across both timebins and sites ($p_{FDR} < .05$) – a stringent threshold given the thousands of site-by-timebin comparisons.

Despite this conservative correction, significant cross-classification was distributed across the cortex and, critically, recapitulated the spatiotemporal hierarchy observed in our ROI analyses. Above-chance accuracy (Fig. 3.5a) and condition-specific posterior probabilities (Fig. 3.5b) emerged early in posterior cortex and spread anteriorly over the trial: effects were sparse

and largely posterior at early latencies (~100–250 ms), became dense across temporoparietal cortex by mid-latencies (~500–750 ms), and extended into prefrontal cortex at later latencies (~1000–2000 ms). That this gradient is visible at the level of individual sites – each surviving correction across the full space-by-time grid – indicates that the autobiographical-versus-cognitive signal is robust at the single-electrode level and not an artifact of pooling sites within ROIs.

The condition maps (Fig. 3.5b) further corroborated our ROI findings. Cognitive and autobiographical trials produced spatially opponent patterns: cognitive trials yielded predominantly negative (cognitive-like) posteriors while autobiographical trials yielded predominantly positive (autobiographical-like) posteriors across overlapping cortical territory. Self- and other-mentalizing maps both resembled the autobiographical map, recruiting autobiographical-like patterns across the DMN, consistent with the autobiographical-like recruitment and absent self/other magnitude difference observed in the ROI posterior-probability analyses.

Notably, however, the single-site maps revealed a self/other timing difference that is obscured when sites are pooled within ROIs. Self-mentalizing began recruiting autobiographical-like patterns earlier – emerging around 500 ms and approaching peak recruitment by ~1000 ms – whereas other-mentalizing recruited autobiographical-like patterns later, not reaching mPFC until ~750 ms and continuing to build through the end of the trial (~2000 ms). Thus, self- and other-mentalizing converge on the same autobiographical-like endpoint but differ in the latency with which they reach it. This temporal dissociation, undetectable in the pooled magnitude contrast, directly echoes our prior finding that self- and other-mentalizing are distinguished by activation timing rather than magnitude, with other-mentalizing evoking slower, lengthier processing across successive DMN regions (Paper I).

Taken together, the whole-brain analysis reproduces, at single-site resolution, the posterior-to-anterior spatiotemporal gradient that characterizes autobiographical recruitment during mentalizing – now established across four complementary analyses (classification accuracy, the posterior-probability gradient, psychobehavioral associations, and whole-brain mapping) – while additionally revealing a self/other timing dissociation consistent with our prior work.

Discussion

Using spectral cross-classification, we tested whether mentalizing recruits autobiographical computations. Classifiers trained to distinguish autobiographical memory from arithmetic, then applied to held-out mentalizing trials, revealed that mentalizing engages autobiographical-like spectral patterns across the DMN. This recruitment followed a posterior-to-anterior spatiotemporal gradient – established independently across four analyses (classification accuracy, the posterior-probability gradient, psychobehavioral associations, and whole-brain mapping) – and was quantified by a consistent ordering of conditions along the autobiographical-cognitive axis: autobiographical memory engaged autobiographical-like computation most strongly, followed by self- and other-mentalizing at intermediate values, with arithmetic at the cognitive extreme (Autobio > Self $\approx$ Other > Cognitive). Mentalizing thus recruits autobiographical-like computation substantially, but reliably short of full autobiographical engagement.

This last point bears emphasis, because it distinguishes our account from a weaker one. A skeptic might argue that the autobiographical-cognitive axis is too coarse to be informative – that any DMN-engaging task would score as autobiographical-like simply by virtue of engaging the DMN. Several features of our results argue against this interpretation. First, the Mentalizing–Autobio contrast was significantly negative across the DMN: mentalizing was

reliably distinguishable from autobiographical memory itself, indicating that the axis resolves graded differences rather than collapsing all DMN activity into a single category. Second, recruitment was regionally heterogeneous and behaviorally consequential (see below), not a uniform offset. Third, the spatiotemporal trajectory of autobiographical recruitment recapitulated the posterior-to-anterior sequence we previously documented for mentalizing (Paper I), indicating that the posterior probabilities preserve fine-grained spatiotemporal structure rather than reflecting generic engagement. Together, these features indicate that the autobiographical-cognitive axis captures meaningful computational structure, while we acknowledge that our claim is appropriately bounded: cross-classification establishes shared population-level spectral signatures, which are consistent with – but do not by themselves prove – shared underlying computation.

Our self/other findings converge with our prior account of mentalizing (Paper I) through an independent method. Self- and other-mentalizing were statistically indistinguishable along the autobiographical-cognitive axis once stimulus visual dissimilarity was controlled, indicating that both forms of mentalizing recruit autobiographical computation to an equivalent degree. This is precisely what a shared neural code predicts: in Paper I, we found that self- and other-mentalizing engage near-identical neuronal populations distinguished by the timing rather than the magnitude or content of their activity. Here, using a method sensitive to computational content rather than activation magnitude, we again find no content-based distinction between self- and other-mentalizing. Notably, the single-site whole-brain maps did suggest a residual self/other timing difference – other-mentalizing reaching peak autobiographical-like recruitment later than self-mentalizing – qualitatively consistent with the slower, lengthier processing we previously documented for other-mentalizing. We did not test this timing difference formally here, and we return to it below. Taken together, mentalizing appears to be organized along two largely orthogonal axes: a code-content axis, along which mentalizing is distinguished from

arithmetic by its recruitment of autobiographical computation, and a temporal-dynamics axis, along which self- and other-mentalizing are distinguished by timing rather than computational content.

Autobiographical recruitment was behaviorally consequential, and in a temporally structured way. Most strikingly, the relationship between recruitment and behavioral response time reversed over the trial: early cognitive-like recruitment predicted faster responses, whereas later autobiographical-like recruitment predicted faster responses. The early effect is consistent with a fast, domain-general engagement of the DMN that scaffolds subsequent content-specific processing – a dynamic we characterize more fully in Paper II, where mentalizing and autobiographical memory share an early, broadly synchronous theta phase preceding a content-specific high-frequency cascade. The later effect, in which autobiographical-like recruitment facilitates responding, is consistent with the autobiographical computations themselves contributing to mentalistic judgments.

Behavioral response choice was predicted by autobiographical recruitment in a manner that varied across mentalizing's neurocognitive hierarchy. In visual cortex, recruitment carried no information about whether participants judged a prompt true or false; in temporoparietal DMN regions, choice-predictive effects were intermediate; and in mPFC, they were most robust and latest in onset, building through the latter half of the trial. This posterior-to-anterior progression, culminating in mPFC near the time of response, is consistent with evidence accumulating toward a decision along the same hierarchy that carries autobiographical recruitment, with mPFC as the apex at which accumulation may resolve (Gold & Shadlen, 2007; Hunt et al., 2012; Tan et al., 2022; Paper I). We frame this as a pattern consistent with an accumulation account rather than a demonstration of one: we did not fit accumulator or attractor models, and the choice-predictive timecourses become low-powered in the latest window of the trial because epochs end at the behavioral response.

This accumulation pattern was clearest and most robust for autobiographical memory itself, and correspondingly weaker for mentalizing. We therefore treat the accumulation interpretation as a working hypothesis for the mentalizing case – supported indirectly by the spatial and temporal structure of choice-predictive effects, and directly by the stronger results in autobiographical memory. That the same posterior-to-anterior, mPFC-culminating architecture appears for both indicates that this choice-predictive organization is not unique to social judgment but is shared with autobiographical retrieval – further evidence that mentalizing draws on autobiographical machinery, with the two converging on a common mPFC apex. A formal test would require modeling the recruitment timecourses against accumulator or attractor dynamics (Wang 2002; Wong & Wang 2006), ideally in a design with sufficient post-decision data to characterize the late window robustly.

Affective valence modulated autobiographical-like recruitment during mentalizing: across the DMN, more positive valence was generally associated with more autobiographical-like processing. This converges with the self-positivity bias we documented behaviorally in Paper I, where participants more readily endorsed positive than negative self- and other-relevant content – a canonical feature of mentalizing (Igou, 2008; Pahl & Eiser, 2005; Silani et al., 2013). The present results suggest a neural counterpart to that behavioral bias: positively-valenced mentalistic content preferentially recruited the autobiographical-like processing that the DMN supplies to mentalizing. One notable exception to this DMN-centered pattern was visual cortex, our control region, which tracked valence despite carrying no mentalizing-specific computation. Although this might appear anomalous, it is consistent with a substantial literature showing that visual cortex receives and represents affective feedback from higher-order regions such as orbitofrontal cortex, including for stimuli with mild or incidental affective value (Barrett & Bar, 2009; Lebrecht et al., 2012; Olofsson et al., 2008; Panichello et al., 2017; Tan & Tarr, 2016). We therefore interpret the visual-cortex valence effect as reflecting top-down affective feedback

rather than bottom-up stimulus valence, particularly as our prompts were not strongly valenced by design.

Several limitations bound our conclusions. First, as noted, cross-classification demonstrates shared population-level spectral signatures between mentalizing and autobiographical memory; it does not directly observe a shared computation, and shared signatures could in principle arise from distinct but spectrally similar processes. Second, our visual-dimension covariates (VD1 and VD2) were derived from all layers of a deep learning image classifier (AlexNet), including higher layers tuned to more complex visual features. This raises the possibility that the covariates were overly aggressive, removing genuine cognitive variance alongside low-level visual confounds. We consider this unlikely: AlexNet is a vision model that cannot read text and therefore cannot represent the semantic content of our written prompts, so its representations of our stimuli are driven by low-level visual properties. Consistent with this, VD1 closely tracked prompt length, and the visual dimensions separated only the visually-distinct arithmetic condition while leaving the semantically-distinct sentence conditions intermingled (Supplementary Fig. 3.1), indicating that the visual dimensions primarily indexed low-level stimulus properties rather than semantic content. We did not formally quantify the VD1–prompt-length relationship, but its strength is apparent in the stimulus feature space.

These limitations point to clear future directions. The qualitative self/other timing difference observed in the whole-brain maps warrants a formal test: trial-by-trial analysis of the onset, peak, and offset latencies of autobiographical recruitment – analogous to the latency metrics we used for high-frequency broadband activity in Paper I – would directly establish whether self- and other-mentalizing differ in the timing of autobiographical recruitment, as the present results suggest. Although we treated autobiographical memory as a unitary domain, autobiographical cognition comprises dissociable processes, most notably semantic and episodic retrieval. In preliminary analyses, individual sites in several participants separated

semantic from episodic conditions cleanly, indicating that the spectral distinction between these subdomains is present in the data; however, the effect did not survive aggregation to ROI or group level. We attribute this to a combination of trial count – our task was not designed to power a semantic/episodic split – and the likelihood that the distinction is carried by a smaller, more specific set of spectral features than the autobio-vs-arithmetic contrast, requiring more sensitive feature selection performed within cross-validation folds to avoid optimistic bias. Pursuing this distinction is a natural extension of the present work, with the single-site observations as encouraging preliminary evidence.

Whether distinct cognitive domains share neural machinery is fundamentally a question about computation rather than co-activation – and co-activation, the currency of fMRI, cannot answer it (Mars et al., 2012). Here we show that mentalizing does not merely co-activate the default mode network alongside autobiographical memory but recruits autobiographical-like computation, graded along the same posterior-to-anterior hierarchy and converging on medial prefrontal cortex, with self and other differing in the timing of that recruitment rather than its content. While much of this computational architecture remains to be characterized, these results show that what the social brain shares with memory is not anatomy alone but – by these measures – the computations performed within it.

Methods

All research activities herein were conducted in accordance with protocol approved by the Stanford Institutional Review Board for human experimentation. All computational procedures and analyses herein were implemented in MATLAB unless otherwise specified. Preprocessing, machine learning and statistical analyses of electrophysiological data were performed using electroCUDA (<https://github.com/kevmtn/electroCUDA>) running on Nvidia GPUs with CUDA 13.1.

Participants

We employed a cohort of thirty-four human participants while they underwent neurosurgical treatment for drug-resistant epilepsy. Each participant provided written informed consent in accordance with the Stanford Institutional Review Board. Participants were not compensated as per Stanford Institutional Review Board guidelines for inpatient clinical research. As part of their presurgical evaluation, participants were implanted with iEEG at Stanford University Medical Center. The anatomical placement of electrode sites was determined according to each participant's clinical needs. Each participant was monitored in hospital for six to ten days prior to surgery, during which the study was conducted.

Behavioral Task

iEEG data was recorded while participants performed an event-related behavioral task with six conditions (trial types; Fig. 3.1a & Supplementary Table 1.4). Five of these conditions required true/false responses to written prompts, while one condition consisted of cued rest. Two conditions featured mentalizing prompts, either about oneself (e.g., "I am honest") or others (e.g., "My neighbor is honest"). Participants were instructed to select a single neighbor (current or past) as the target for other-mentalizing. Cognitive task trials consisted of basic arithmetic (e.g., "9 + 86 = 95"). Two conditions featured memory-related prompts: episodic (e.g., "I ate candy yesterday") and self-semantic (e.g., "I eat a lot of candy"). The cued rest condition required no response and displayed a fixation crosshair for 5-10 seconds. Stimuli were presented in random order and were self-paced, advancing to the next trial after the participants' response, or up to 15 seconds if no response. The inter-trial interval (ITI) occurred -200-0 ms before each trial. The experiment was broken into two separate runs (mean run duration=12.50±1.64 minutes). Participants were allowed a short break between behavioral

runs. On average, each run featured 25 trials of each sentence condition, 40 cognitive trials, and 36 rest trials. Each non-rest trial contained unique prompts; prompts were not repeated within participant. Responses were made via a handheld keypad using either the '1' (true) or '2' (false) key. Participants were instructed to perform the task as accurately and as quickly as possible. All stimuli were presented in white font on a black background (1200x800 pixels) using Psychophysics Toolbox 3 (Kleiner et al., 2007). For all analyses, trials were excluded if meeting any of the following criteria: high-frequency epileptic oscillations, no behavioral responses, irrelevant button presses, or RT_{Behav} under 400 ms.

iEEG data acquisition

Whereas Paper I used electrocorticography (ECoG) alone, the expanded cohort here additionally includes stereo-EEG (sEEG) depth electrodes, which sample sulcal, deep, and subcortical-adjacent structures (e.g., hippocampus) inaccessible to subdural grids and can record across cortical depth rather than the surface alone. We use “iEEG” to refer to both modalities throughout. iEEG recordings were obtained via 4128 subdural and depth electrodes. Electrodes (platinum plates with diameter of 1.2-2.3 mm) were implanted subdurally onto the cortical surface in grids or strips with center-to-center interelectrode spacing of 4-10 mm (Adtech Medical Instruments). Electrodes were connected to a multichannel recording system (Nihon Kohden; Tucker Davis Technologies) with sampling rate of 1,000 Hz or above. Anatomical data was acquired using a GE 3-Tesla SIGNA Magnetic Resonance Imaging (MRI) scanner at Stanford University. A T1-weighted anterior-posterior commissure-aligned pulse sequence was used. T1 data was resampled to 1 mm isotropic voxels, then segmented to distinguish gray and white matter using FreeSurfer (Fischl, 2012). To facilitate electrode localization, postimplant computerized tomography (CT) scans were coregistered to the preoperative MRI anatomical brain volume (Hermes et al., 2010). For each participant,

electrodes sites were localized in BiImage Suite (Papademetris et al., 2006) and displayed on the participants' own reconstructed 3D cortical surface using the iELVis toolbox (Groppe et al., 2017). Electrode positions were corrected for postimplantation brain shift, allowing for the accurate anatomical localization of electrodes sites (Dykstra et al., 2012).

Defining regions of interest (ROIs)

Because the expanded cohort included stereo-EEG (sEEG) depth electrodes sampling sulcal and deep cortex that surface-based native-space parcellation localizes poorly, electrode sites were assigned to ROIs using the Glasser Human Connectome Project multimodal parcellation (HCP-MMP1.0; Glasser et al., 2016), projected from fsaverage onto each participant's native FreeSurfer cortical surface via folding-based registration, using a custom electrode-localization pipeline built on iELVis (Groppe et al., 2017). MNI coordinates were used as a secondary criterion to disambiguate parcels spanning ROI boundaries (specified below). Each site's HCP parcel was assigned to one of seven a priori ROIs – visual cortex (control), ATL, TPJ, PMC, amPFC, dmPFC, and vmPFC – which were organized into three ROI groups: Visual, temporoparietal DMN (tpDMN: ATL, TPJ, PMC), and medial prefrontal cortex (mPFC: dmPFC, amPFC, vmPFC). Parcel assignments were applied bilaterally and are given below without hemispheric prefixes.

Visual cortex: early visual parcels (V1, V2, V3, V3A, V3B, V4, V6, V6A, V7, V8, and the ventral VVC and VMV1), the MT+ complex (MT, MST, FST, V4t), and higher-order occipitotemporal parcels LO1–3, PIT, PH, and FFC. **ATL:** anterior temporal parcels STGa, TGd, TGv, STSda, STSva, TE1a, and TE2a (TE2a restricted to anterior sites, MNI $y > -19$). **TPJ:** inferior-parietal and temporoparietal-junction parcels PGs, PGi, PFm, STV, and TPOJ1–3, together with the posterior portion of PSL (MNI $y < -44$). **PMC:** precuneal, posterior-cingulate, and retrosplenial parcels POS1, POS2, 7Pm, 7m, PCV, v23ab, d23ab, 31pd, 31pv, 31a, and RSC.

amPFC: a24, p32, and 10r, together with the ventral, near-midline portion of 10d ($\text{MNI } z \leq 10$ and $|x| < 12$). **dmPFC:** 9m, 9a, 9p, d32, a32pr, 8BM, and 8BL, together with the dorsal portion of 10d ($\text{MNI } z > 10$); all dmPFC parcels were restricted to anterior, near-midline sites ($\text{MNI } y > 28$ and $|x| < 12$). **vmPFC:** OFC, 25, s32, and 10v.

Sites were assigned to ROIs by this atlas; they were not excluded on the basis of anatomical gray/white tissue classification, which was unreliable for some depth contacts. Instead, site inclusion for each analysis followed a functional response criterion (≥ 100 ms of contiguous activation, $p_{\text{FDR}} < .05$, relative to the pre-stimulus baseline in at least one of the conditions entering that analysis; see Statistics), with a small number of sites reassigned after manual inspection. This atlas-based approach differs from the native-space FreeSurfer (Desikan-Killiany-style) parcellation with manual inspection used in Paper I, which suited that ECoG-only cohort but does not provide the fine, deep-coverage parcels needed for the sEEG sites sampled here.

iEEG preprocessing

Preprocessing was performed on a single-subject basis. First, electrodes marked as unusable in postclinical evaluation (pathological, loose contact, etc.) were discarded. Remaining problematic electrodes were identified using the ImaGIN classifier and other heuristics (Tuyisenge et al., 2018). This left 3827 electrodes for use in analyses.

Recordings from remaining electrodes underwent a computationally-driven robust denoising routine to maximize data quality. First, robust detrending was used to remove slow drifts from recordings (de Cheveigné & Arzounian, 2018). Afterwards, robust average referencing was applied with rank correction. Electrical line noise was then removed using Zapline Plus (Klug & Kloosterman, 2022). Next, non-stationary noise was removed using artifact subspace reconstruction (Chang et al., 2020). Stationary noise was then identified by

independent components analysis and the ImaGIN classifier; artifactual components were subtracted from recordings.

Denoted recordings then underwent time-frequency decomposition via continuous wavelet transform with a Morse mother wavelet (TimeBandwidth = 20), applied to each continuous run to avoid epoching-related edge artifacts. We used a higher time-bandwidth product than the default (TimeBandwidth = 60) to prioritize temporal over spectral specificity, yielding 41 log-spaced frequencies spanning 5–300 Hz. Spectral magnitude was log-transformed into decibel-scaled power. Spectral timecourses were then epoched by trial and baseline-corrected by z-scoring each epoch to its pre-stimulus baseline (–200–0 ms). The resulting single-trial, single-site spectral timecourses (41 frequencies × timepoints) provided the features for cross-classification.

Cross-classification

To test whether mentalizing recruits autobiographical computations, we trained classifiers to distinguish autobiographical memory from the cognitive control task and applied them to held-out mentalizing trials. The autobiographical class pooled semantic and episodic trials; the cognitive class comprised arithmetic trials. Trained classifiers were then applied to self- and other-mentalizing trials, yielding trialwise posterior probabilities of autobiographical versus cognitive class membership. This procedure was performed independently at each 25 ms timebin, generating posterior-probability timecourses.

Classification used regularized linear discriminant analysis (LDA). Two hyperparameters were tuned per participant: Gamma, which regularizes (shrinks) the covariance estimate, and Delta, which thresholds linear coefficients to promote sparsity. To avoid optimistic bias, classifier accuracy on the trained autobiographical-versus-cognitive contrast was assessed using trial-grouped nested cross-validation, in which hyperparameter tuning was confined to inner training folds and never exposed to held-out test data. Posterior probabilities for

mentalizing trials were generated from classifiers trained on the full autobiographical and cognitive data.

Cross-classification was performed at two spatial scales. Single-site classification used the 41-frequency spectral feature vector from each site. ROI classification concatenated these spectral features across all sites within an ROI. Whole-brain analyses were performed at the single-site scale using the same procedure but without nested cross-validation, owing to computational constraints. Observations consisted of the spectral feature frames within each 25 ms timebin; because each timebin contained approximately five frames per trial and approximately 50 trials per class, each classifier was trained on roughly 250 observations per class per timebin.

Analyses spanned 0–2000 ms peri-stimulus. We restricted the analysis window to 2000 ms because of computational constraints and because trial counts decline at later latencies: trials terminated at the behavioral response (RT_{Behav}) plus a 250 ms inter-trial interval included in each epoch, so later timebins drew on progressively fewer trials. Even with all trials included, these analyses were bounded by observation count at later latencies. Classification accuracy was tested for significance via parametric tests, FDR-corrected across timebins and sites or ROIs ($p_{\text{FDR}} < .05$).

Analyses of posterior probability

Posterior-probability (PP) timecourses were summarized using a continuous index, $PP = P(\text{autobiographical}) - P(\text{cognitive})$, ranging from -1 (fully cognitive-like) to $+1$ (fully autobiographical-like), such that the maximum possible difference between two conditions is 2.

To estimate per-condition PPs and their contrasts at the population level, we used unified linear mixed-effects models (LMEMs) that jointly represented all four conditions (autobiographical, cognitive, self-mentalizing, other-mentalizing), nested within subject within

trial to account for between-subject and trial-by-trial variance. Per-condition PP estimates were obtained via no-intercept, full-dummy coding of condition. Condition contrasts were obtained via reference coding: the cognitive condition served as the reference for the autobiographical-versus-cognitive and mentalizing-versus-cognitive contrasts, and self-mentalizing served as the reference for the other-versus-self contrast. Two visual dimensions (VD1 and VD2; see below) were included as nuisance covariates to control for stimulus visual dissimilarity. Random slopes were excluded after they produced singular fits, consistent with recommendations to simplify overparameterized random-effects structures (Matuschek et al., 2017). Models were estimated separately at each timebin and FDR-corrected across timebins and ROIs or sites ($p_{\text{FDR}} < .05$).

To assess the psychobehavioral relevance of autobiographical recruitment, we analyzed behavioral response time (RT_{Behav}), behavioral response choice ($\text{Choice}_{\text{Behav}}$; true/false), and affective valence as predictors of PP timecourses using separate LMEMs. Each model controlled for the original task condition (semantic, episodic, self-, or other-mentalizing) and the visual dimensions, nested within subject within condition within trial. RT_{Behav} and $\text{Choice}_{\text{Behav}}$ were estimated separately for autobiographical (semantic + episodic) and mentalizing (self + other) trials; affective valence was estimated for mentalizing trials only, as autobiographical stimuli were affectively neutral.

The two visual dimensions (VD1 and VD2) were derived from AlexNet, a deep learning image classifier, applied to the visual stimuli (see Data and design); VD1 was highly correlated with prompt length, indicating that the visual dimensions primarily captured low-level stimulus properties.

Paper IV: Probing blind spots in social neuroscience with source-space EEG and MEG

Abstract

Blind spots in social neuroscience, particularly in systems and computational mechanisms, often stem from poor temporal resolution in functional magnetic resonance imaging (fMRI). We show how to probe these blind spots using source-space electroencephalography and magnetoencephalography (M/EEG), noninvasive electrophysiological modalities with millisecond temporal resolution and up to laminar-scale anatomical resolution, also offering favorable costs and naturalism over fMRI. To these ends, we discuss the rich spatiotemporal and spectral neuronal dynamics captured by M/EEG through the framework of distributed hierarchical processing, revealing fundamental ties between signal features and systems-computational neuroarchitecture. We then describe contemporary M/EEG methods and their neurophysics, concluding with a step-by-step guide for applying such methods to the Why/How mentalizing task, providing a template for addressing outstanding questions in the field. Throughout, we detail potential pitfalls and how to avoid them. Altogether, we show that source-space M/EEG provides compelling and incisive mechanistic characterizations of the social brain.

Introduction

A social neuroscientist could be forgiven for believing that cognitive abilities arise from static brain activations like those depicted in functional anatomy maps. Throughout its history, social neuroscience has overwhelmingly relied on functional magnetic resonance imaging (fMRI), a neuroimaging modality with high anatomical resolution but low temporal resolution (Lieberman, 2013). In fMRI, brain regions appear to respond simultaneously on a one-dimensional scale of ‘activation’. However, electrophysiology shows that brain regions respond through intricate spatiotemporal sequences with a multi-dimensional spectrum of activity (Felleman & Van Essen, 1991; Ward, 2003; Fox et al., 2018). As such, overreliance on fMRI leaves our field blind to critical neurocognitive dynamics at millisecond scales, obscuring invaluable systems and computational mechanisms (Tan et al., 2022; Tan & Tarr, 2016).

We therefore introduce source-space electroencephalography (EEG) and magnetoencephalography (MEG), collectively referred to as M/EEG, which are noninvasive electrophysiological modalities providing millisecond temporal resolution and up to laminar-scale anatomical resolution (i.e., cortical layer discrimination) while offering favorable costs and naturalism over fMRI (Brookes et al., 2022; Lopes da Silva, 2013). Source-space M/EEG capture the fast spatiotemporal and spectral dynamics of cortical and subcortical regions, enabling incisive and mechanistic characterizations of neurocognitive processes. While many fields of neuroscience use these modalities extensively, social neuroscience remains an outlier with only a handful of studies (Van Duynslaeger et al., 2007; Vistoli et al., 2011; Pavlidou et al., 2014; Bögels et al., 2015; Tucciarelli et al., 2015; Mossad et al., 2016; H. Wang et al., 2016; Yuk et al., 2018). Avoidance of source-space M/EEG may stem from conflation with traditional M/EEG, which social neuroscientists often dismiss as disorienting and uninterpretable. We dispel conflated perceptions by showing that source-space M/EEG can produce rich, compelling, and easily-interpretable results that are analogous to time-locked BOLD responses in fMRI.

We hope this article serves as a practical call-to-action for resolving blind spots and bolstering mechanistic description in social neuroscience. To these ends, we first discuss M/EEG signal features and their fundamental links to systems and computational neuroarchitecture. We then detail contemporary methods in source-space M/EEG, concluding with a step-by-step guide for applying such methods to the Why/How task for mentalizing and action identification (Spunt & Adolphs, 2014), providing a template for addressing outstanding questions in the field.

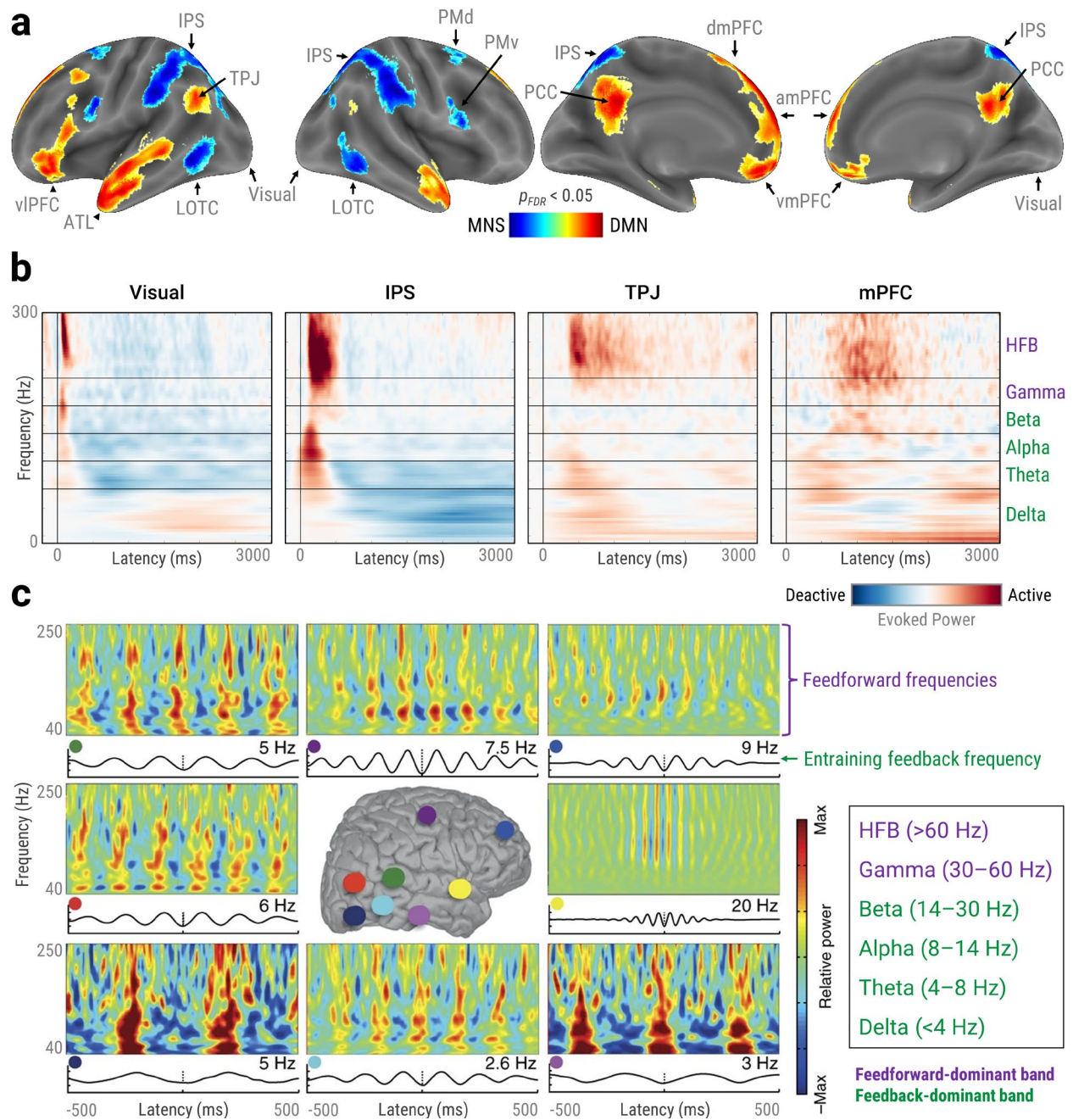


Figure 4.1: Neuronal signals. **(a)** Overview of the social brain derived from the Why/How fMRI localizer (Spunt & Adolphs, 2014). Blue colors represent MNS regions that produce significantly greater activation for 'How' action identification prompts over 'Why' mentalizing prompts (see Fig. 4). Red colors represent DMN regions that produce significantly greater activation for 'Why' prompts ($p_{FDR} < 0.05$, corrected across voxels). **(b)** Trial-averaged event-related spectral perturbations (ERSPs) showing mentalizing-evoked EEG activity across brain regions, time, and spectral frequencies. Warm colors indicate activation over prestimulus baseline. Cool colors indicate deactivation. Spectral bands are delineated. **(c)** Resting-state MEG showing entrainment of feedforward frequencies (ERSP plots) by distinct feedback frequencies (waveform plots) across brain regions. Figure from Baillet, 2017. **Abbreviations:** mirror neuron system (MNS); default mode network (DMN); visual cortex (Visual), lateral occipitotemporal cortex (LOTC), intraparietal sulcus (IPS), dorsal premotor cortex (PMd), ventral premotor cortex (PMv), temporoparietal junction (TPJ), posterior cingulate cortex (PCC), anterior temporal lobe (ATL), ventrolateral prefrontal cortex (vIPFC), dorsomedial prefrontal cortex (dmPFC), anteromedial prefrontal cortex (amPFC), ventromedial prefrontal cortex (vmPFC).

Signals and theory

Electrophysiology is essential for characterizing the systems and computational mechanisms of the social brain. This is exemplified by distributed hierarchical processing, a central organizational principle of neurocognitive systems (Hebb, 1949; Marr & Poggio, 1976; Felleman & Van Essen, 1991; Huntenburg et al., 2018). In this section, we first summarize distributed hierarchical processing and M/EEG physiology before diving into spectral and spatiotemporal neuronal dynamics. Afterwards, we discuss questions suited for source-space M/EEG.

Distributed hierarchical processing

Distributed hierarchical processing provides a cohesive mechanistic framework for functional relationships between brain regions, uncovering invaluable neurocognitive insights in the decades since Hebb's foundational work (Hebb, 1949). Specifically, countless electrophysiological studies show that brain regions are organized hierarchically from low-level sensory to high-level associative at various spatiotemporal scales (Van Essen & Maunsell, 1983; Whittington et al., 1995; Mishkin et al., 1997; Engel et al., 2001; Buzsáki & Draguhn, 2004; Mormann et al., 2008; Kveraga et al., 2011; Jeon, 2014; Fries, 2015; Yang et al., 2019; von Seth et al., 2023). These studies also reveal that cognitive functions are instantiated by feedforward (bottom-up) and feedback (top-down) computations along such hierarchies. Generally, feedforward computations subserve rapid and often stimulus-driven passes of neurocognitive processing, while slower feedback computations expand or fine-tune these passes based on the internal and external milieu (Lamme & Roelfsema, 2000; Alilović et al., 2019). To support complex operations, recurrent processing dynamically updates feedforward and feedback computations via iterative or recursive looping. Critically, these computations are implemented

by electrophysiological signaling, exhibiting crucial spatiotemporal and spectral dynamics that only electrophysiological methods can capture.

In social neuroscience, distributed hierarchical processing has been explored in two recent studies: invasive electrophysiology (Tan et al., 2022) and neuroimaging meta-analysis (Schurz et al., 2021). These studies report tight correspondence between functional selectivity and spatiotemporal dynamics, revealing that the social brain ascends from sensory regions (e.g., fusiform face area), to mirror neuron system (MNS, e.g., intraparietal sulcus), and then to default mode network (DMN), where it traverses temporoparietal regions (tpDMN) like temporoparietal junction (TPJ) before reaching its apex in medial prefrontal cortex (mPFC; Fig. 4.1a). Social cognitive functions appear likewise hierarchical, broadly falling between lower-level sensory cognition (e.g., action identification), intermediate cognition (e.g., empathy), and higher-level abstract cognition (e.g., mentalizing).

Taken together, the hierarchical architecture of the social brain closely mirrors the primary cortical hierarchy established by cytoarchitecture, intrinsic dynamics, and non-social functioning (Margulies et al., 2016; Huntenburg et al., 2018; Fox et al., 2018; Demirtaş et al., 2019; Paquola et al., 2023). This supports domain-general accounts of the social brain, which posit that functional organization is driven by computation rather than cognitive domain, though the two are often correlated (Parkinson & Wheatley, 2015; Spunt & Adolphs, 2017). For example, TPJ may instantiate integrative gestalt construals (Lieberman, 2022) that are later interrogated by mPFC through ‘fuzzy’ propositional logic, ultimately providing high-dimensional probabilistic inferences for social cognition (Mitchell, 2006; Thornton et al., 2019; Tan et al., 2019, 2022) and abstract cognition at large (S. W. Lee et al., 2014; Doll et al., 2015; Tamir et al., 2016; Tan & Tarr, 2016).

While Schurz et al. (2021) and Tan et al. (2022) lay a solid foundation, much remains unknown. Moreover, although fMRI multivariate methods have made substantial inroads on

social computations (Weaverdyck et al., 2020), they fail to capture fast spatiotemporal and spectral dynamics, obscuring crucial features such as feedforward and feedback interrelations across brain regions (Cichy & Pantazis, 2017). As such, the hierarchical functioning of the social brain remains a central blind spot, leading to numerous blind spots in the systems and computational mechanisms of social cognition. Source-space M/EEG could probe these blind spots while being cheaper and more naturalistic than fMRI (see *methods*).

Physiology

A subtle choice of adjective indicates the stark physiological difference between targets of measurement in M/EEG and fMRI. M/EEG directly measure *neuronal* activity, while fMRI measures *neural* hemodynamics that reflect both neuronal and non-neuronal metabolism (Jackson & Bolger, 2014). ‘Neuronal’ refers specifically to neurons, while ‘neural’ refers broadly to the nervous system. Specifically, M/EEG detect electromagnetic fields generated by as few as 10,000 neurons, while a single fMRI voxel samples hemodynamics over a million cells, including neurons, glia, and vascular cells (Baillet, 2017). However, given signal physiology, the neurons sampled by M/EEG might be interdigitated among non-contributing neurons over a wider area than an fMRI voxel.

EEG and MEG signals have identical neuronal origins, namely oscillatory post-synaptic potentials (PSPs). Broadly, PSPs are excitatory or inhibitory modulations of cellular membrane potential caused by neurotransmitter receptor binding (Buzsáki et al., 2012). Dendritic integration sums multiple PSPs together to regulate action potential firing rate. Oscillatory PSPs specifically mediate ‘long-range’ coordination beyond adjacent cortical columns, flexibly assembling distributed neuronal ensembles to meet neurocognitive demands (Whittington et al., 1995). Oscillatory PSPs are mainly generated by pyramidal neurons with extensive long-range innervation, which can link even the furthest cortical and subcortical regions (Thio & Grill, 2023).

Oscillatory PSPs are therefore remarkably efficient at long-range signaling, with one action potential evoking thousands of PSPs across the brain. As such, cognitive computations across the cortical hierarchy are largely instantiated by oscillatory PSPs (Buzsáki & Draguhn, 2004; Gu et al., 2021).

Action potentials and non-oscillatory PSPs are local to single neurons and discrete neuronal populations, respectively, constituting only a minor proportion of M/EEG signals (Thio & Grill, 2023). Currently, invasive electrophysiology is required to reliably measure these local signals, while M/EEG is largely limited to oscillatory PSPs that robustly escape the scalp (Parvizi & Kastner, 2018). Moreover, M/EEG cannot discern underlying anatomy without source localization algorithms, hence ‘source-space’ versus traditional unlocalized M/EEG (Wendel et al., 2009). Nevertheless, while source-space M/EEG are coarser than invasive modalities, M/EEG still capture rich spatiotemporal and spectral neuronal dynamics that are obscured from fMRI (Lopes da Silva, 2013).

Spectral dynamics

Neuronal activity occurs through a spectrum where frequency distinguishes both physiology and function (Engel et al., 2001; Ward, 2003; Jensen & Colgin, 2007; Mendoza-Halliday et al., 2024). This is akin to the radio spectrum, where frequency differentiates equipment (e.g., cellular or Wi-Fi) and function (e.g., voice or internet). In electrophysiology, frequency refers to the spatiotemporal interval (i.e., wavelength) at which oscillatory PSPs coordinate computations. Lower frequencies are slower and direct larger neuronal ensembles, while higher frequencies are faster and direct smaller ensembles (Buzsáki, 2010). M/EEG signals are summed oscillatory PSPs across frequencies mainly under 200 Hz but can reach over 1000 Hz (Onton & Makeig, 2009; Baillet, 2017). Spectral decomposition isolates these

frequencies, providing numerous distinct measures of neuronal activity for a given anatomical location (Kim & Im, 2018).

The electrophysiological spectrum is divided into major bands (Fig. 4.1b): delta (<4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-30 Hz), gamma (30-60 Hz), and high-frequency broadband (HFB; >60 Hz), the latter also called high gamma (Buzsáki, 2006). Generally, faster bands (gamma and HFB) carry feedforward signals from superficial cortical layers, while slower bands (delta through beta) carry feedback signals from deep cortical layers, though exceptions occur (Lamme & Roelfsema, 2000; Markov et al., 2014; Mejias et al., 2016; Bonaiuto, Meyer, et al., 2018; Mendoza-Halliday et al., 2024). Critically, each band is found in nearly every brain region, sometimes showing marked regional variability in both physiology and function. For example, Fig. 4.1b shows event-related spectral perturbations (ERSPs) of mentalizing-evoked activity relative to prestimulus baseline, revealing distinct spectral responses across brain regions.

Gamma and HFB, the main feedforward bands, play the most general cognitive roles and facilitate the brunt of neurocomputation by tightly modulating neuronal firing (Uhlhaas et al., 2011; Adaikkan & Tsai, 2020). Gamma regulates firing by balancing excitation and inhibition, rapidly alternating between them to synchronize neuronal ensembles (Ray & Maunsell, 2011). HFB enacts multiple mechanisms of firing modulation and operates at finer spatiotemporal scales than gamma (Leszczyński et al., 2020). Consequently, of any band, HFB shows the most correlation with firing rate and fMRI activation (Muthukumaraswamy, 2013). The initial feedforward sweep of gamma and HFB activity marks a coarse first pass of neurocognitive processing before any feedback (Alilović et al., 2019). Critically, while gamma and HFB carry feedforward signals, these bands can indirectly reflect feedback after sustained processing, as feedforward and feedback processes are looped together by the recurrent computations that instantiate much of cognition (Fries, 2015; Xiong et al., 2023). Moreover, feedback is easily

integrated by gamma and HFB due to small neuronal ensemble sizes (Buzsáki & Wang, 2012). Social cognition strongly recruits HFB for both feedforward and recurrent processing (Tan et al., 2022). Altogether, gamma and HFB are considered broad signals of neuronal activity.

Alpha, a feedback band, facilitates predictive coding and active inference in cognition, whether in sensory or semantic systems (Alamia & VanRullen, 2019; Wendiggensen et al., 2023). At the circuit level, alpha gates irrelevant representations from propagating downstream (Engel et al., 2001; Fries, 2015; Bonaiuto, Meyer, et al., 2018; M. Halgren et al., 2019). Social neuroscientists are often interested in sensorimotor alpha (μ), as it can be suppressed by either observing or performing actions (Perry et al., 2011). While μ suppression is a tempting index of MNS engagement, μ appears largely independent of social cognition (Hobson & Bishop, 2017; Brunsdon et al., 2019). To add confusion, μ suppression studies often use traditional M/EEG, which cannot reliably distinguish μ from non-sensorimotor alpha (see *source localization*). Overall, μ gates corticospinal excitability in a differential manner across somatosensory and motor neurons, which is inconsistent with mirroring as canonically described (Thies et al., 2018; Wischniewski et al., 2022).

Beta, another feedback band, might better reflect MNS activity relative to μ . Specifically, action identification recruits beta in more MNS regions than seen in μ suppression (Pavlidou et al., 2014; Tucciarelli et al., 2015). Moreover, beta is prominent in sensorimotor processing overall (Barone & Rossiter, 2021). However, like all bands, beta is domain-general and helps support a wide variety of cognitive functions (Xiong et al., 2023; Wessel & Anderson, 2024). At the systems level, beta assembles and regulates neuronal ensembles in a top-down manner, often at the behest of computations in other bands (Buzsáki, 2010; Kopell et al., 2011). As such, beta appears to set the 'neurocognitive status quo' by balancing distributed network representations (Engel & Fries, 2010; Spitzer & Haegens, 2017).

Theta facilitates a mix of feedforward and feedback computations, playing a role in balancing between the two, despite having slower oscillations than alpha and beta (Pastoll et al., 2013; E. Halgren et al., 2015; Fries, 2015). Broadly, theta enacts sensory and attentional sampling in thalamocortical circuits. In sensory cortex, theta primarily subserves feedforward sampling (Helfrich et al., 2018). In associative cortex, theta also facilitates feedback computations. For example, theta carries semantic feedback from tpDMN to sensory and other regions (Clarke et al., 2018). Theta also facilitates goal-directed feedback from frontoparietal control network (Phillips et al., 2014). Moreover, theta is critical for memory (Herweg et al., 2020) and internally-oriented attention (Kam et al., 2019), both of which recruit theta in DMN. Social cognition recruits theta in MNS but can either recruit or suppress theta in DMN (Bögels et al., 2015; Foster et al., 2015; Fox et al., 2018). Interestingly, theta-enhancing neurofeedback and brain stimulation can result in greater mentalization and DMN connectivity (H. Wang et al., 2016; Imperatori et al., 2017). Taken together, theta feedback appears to play a major role in sculpting social cognitive computations across MNS and DMN, the details of which deserve much more investigation.

Faster frequencies are often nested within slower ones through phase-locking, whereby feedback entrains fast feedforward computations to efficiently meet neurocognitive demands (Engel et al., 2001; Ward, 2003; Jensen & Colgin, 2007; Fries, 2015). For example, Fig. 4.1c shows entrainment of fast feedforward frequencies by distinct feedback frequencies across brain regions. Critically, a given brain location can have multiple entraining feedback frequencies, which vary dynamically with neurocognitive demands. Taken together, distinct feedback computations flexibly guide feedforward computations, jumping across brain regions to best meet demands.

Phase-locking can occur at distinct phases of an oscillatory cycle. For example, hippocampus uses numerous fast feedforward frequencies to disambiguate information from

different sources, which are differentially phase-locked by theta to meet working memory demands (Lopes-dos-Santos et al., 2018; L. Zhang et al., 2019; Michon et al., 2021; Aguilera et al., 2022). In the context of working memory, activity in some feedforward frequencies (e.g., 35 and 170 Hz) can be locked to theta troughs, facilitating retrieval-encoding cycles for distinct neurocognitive processes. In contrast, activity in other feedforward frequencies (e.g., 100 Hz) can be locked to theta peaks, mediating representational integration and pruning. Critically, active feedforward frequencies vary with each theta cycle to meet ongoing demands. Exploring these rich hippocampal mechanisms could, among other possibilities, reveal the systems-computational architecture of social working memory (Meyer et al., 2012, 2015). Altogether, hippocampus provides a well-characterized example of the intricate spectral mechanisms used throughout the brain, illustrating that specific frequencies within bands are assigned dynamic computational roles (Jensen & Colgin, 2007; Fries, 2015; Xiong et al., 2023; Mendoza-Halliday et al., 2024).

In sum, the electrophysiological spectrum is itself hierarchical, with slow feedback frequencies (delta, theta, alpha, and beta) directing fast feedforward frequencies (gamma and HFB). Critically, each band is found throughout the brain, sometimes showing marked regional variability in physiology and function

Spatiotemporal dynamics

To incisively characterize systems-computational neuroarchitecture, spatiotemporal and spectral dynamics should be considered holistically, ideally incorporating functional connectivity and laminar structure (Parvizi & Kastner, 2018). Spectral dynamics alone leave ambiguity in distinguishing feedforward and feedback signals, providing no information about where such signals originate or lead to. Temporal dynamics also furnish critical details like conscious awareness, which occurs after roughly 250 ms of sustained stimulus-evoked neuronal activity

(Olofsson et al., 2008). Moreover, differences between experimental conditions could manifest in the timing (e.g., duration), not just the intensity (e.g., spectral power), of neuronal activity. Source-space M/EEG particularly excels at temporal dynamics given its millisecond precision.

Stimulus-evoked cortical activity begins in primary sensory regions and then propagates as intricate spatiotemporal sequences across the cortical hierarchy, reflecting functional interdependencies between brain regions (Gu et al., 2021). Different spectral bands may have unique spatiotemporal trajectories (Fig. 4.1bc), suggesting distinct pathways of feedforward and feedback computation (DiCarlo & Maunsell, 2005). Putative pathways can then be concretely characterized by effective connectivity, which provides a time-varying and directional view of information flow across brain regions and spectral frequencies (Sakkalis, 2011). As such, causal analyses like Granger have greater credibility using M/EEG over fMRI. Lastly, given that feedforward and feedback computations are implemented by distinct cortical layers (Mendoza-Halliday et al., 2024), laminar source localization would maximize mechanistic characterization and precision (see source localization below).

Enacting these recommendations would chart neuronal activity and connectivity across space, time, spectral frequency, and cortical layer. This would reveal distinct feedforward, feedback, and recurrent computations with remarkable specificity and minimal ambiguity. Such results could chart distinct neuronal pathways of social cognitive computation, potentially providing an immense level of mechanistic characterization for the field.

Questions for source-space M/EEG

Source-space M/EEG excels at characterizing functional timing, connectivity, and cross-regional interrelations (Michel & Brunet, 2019; Zorzos et al., 2021). As such, many questions in social neuroscience can at least be complemented with these modalities, though fine-grained functional anatomy is best revealed with fMRI (Fig. 4.2). For example, how do neurocognitive

computations progress across the social brain with time? Is hierarchical structure consistent across social cognitive functions? What are the functional interrelations between each MNS and DMN region? How are social representations transformed as they traverse these regions? What feedforward and feedback computations enact such transformations? Do MNS and DMN act in a strictly feedforward manner to transform action representations into mentalistic inferences, or do they act recurrently to enhance both types of representation? Questions need not be directly neuroscientific. For example, are social inferences anchored-and-adjusted away from the self? How do social cognition and affect interplay, is affect fundamental to social cognition?

Hyperscanning and interpersonal neuronal dynamics are exciting avenues for source-space M/EEG, which are cheaper and more naturalistic than fMRI (see *methods*). A few hyperscanning studies have used traditional EEG, revealing mu suppression in social interactions but mu enhancement during spontaneous coordination (reviewed in Tognoli & Kelso, 2015). However, like other unlocalized mu results, sensorimotor specificity is tenuous (see *spectral dynamics*). We only found one hyperscanning study using source-space M/EEG, a proof-of-concept study showing a viable and fruitful path of research (Holmes et al., 2023). For example, at what timescales does interpersonal neuronal synchrony operate, and does this vary across brain regions? What feedforward, feedback, and recurrent computations subserve interpersonal synchrony and how do they differ from non-interactive social cognition? How do neurocognitive representations propagate across individuals, are shared representations contained to certain brain regions and timescales? Altogether, these are but limited examples of the incisive and intricate mechanistic characterization possible with source-space M/EEG.

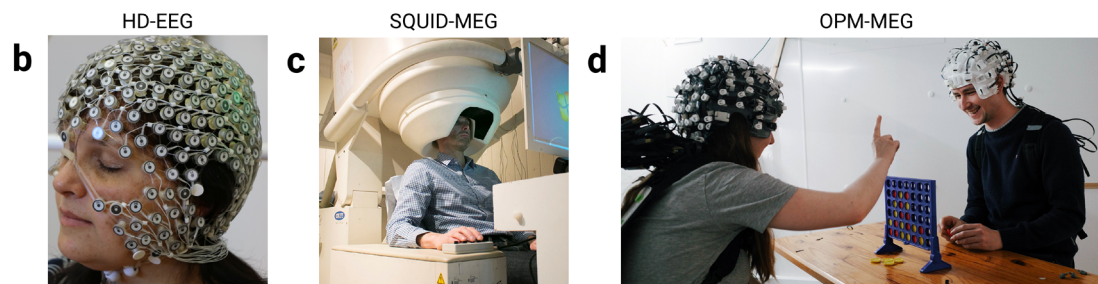
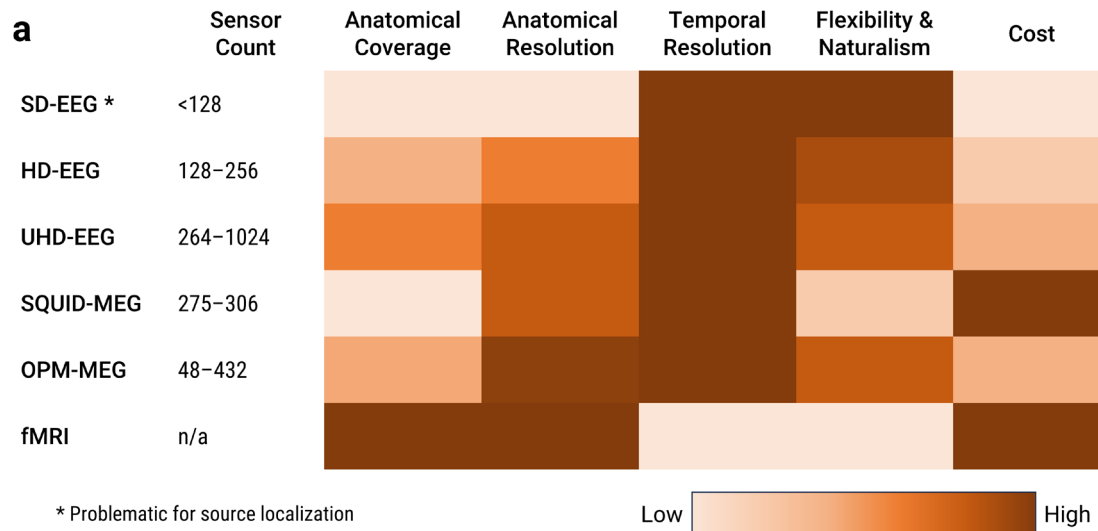


Figure 4.2: EEG and MEG equipment. **(a)** Comparison of M/EEG subtypes and fMRI. Ratings are approximate and based on literature and experience. Anatomical coverage refers to signal sensitivity topology (see Fig. 3b). Anatomical resolution refers to distances between differentiable sources (see Fig. 3a). Temporal resolution refers to length between differentiable timepoints. Flexibility and naturalism refer to the range of viable subjects and experimental paradigms. Note that SD-EEG is problematic for source localization. UHD-EEG and OPM-MEG are considered next-generation as of 2023. **(b)** HD-EEG system with 256 sensors. Facial sensors improve source localization of inferior brain regions. Photo from NIMH Czechia. **(c)** SQUID-MEG system with 275 sensors in a cryogenic helmet. **(d)** OPM-MEG in a naturalistic hyperscanning paradigm. Photos in panels C and D are from UCL Wellcome Trust Centre for Human Neuroimaging. **Abbreviations:** electroencephalography (EEG); magnetoencephalography (MEG); standard-density EEG (SD-EEG); high-density EEG (HD-EEG); ultra-high-density EEG (UHD-EEG); superconducting quantum interference device MEG (SQUID-MEG); optically-pumped magnetometer MEG (OPM-MEG).

Methods and physics

EEG and MEG are highly similar electrophysiological modalities that use noninvasive scalp sensors that discern neuronal activity but not anatomy. The phrase ‘source-space’ refers to source localization, which reconstructs dynamic functional anatomy by decomposing sensor signals into their constituent neuronal sources. This section provides a practical and neurophysical primer on M/EEG equipment, costs, logistics, and source localization.

Equipment and data collection

The age of neuroimaging began in 1875 with the birth of EEG, when electrodes placed on the skull of a rabbit measured stimulus-evoked electrical field potentials (Stone & Hughes, 2013). At that time, electromagnetism was already known to be unified force (Maxwell, 1865), and EEG signals were assumed to have corresponding magnetic fields. However, over a century passed until magnetometers became sufficiently sensitive for scalp readings, a feat accomplished by superconducting quantum interference device MEG (SQUID-MEG), a technology with major drawbacks (D. Cohen, 1972). As of 2026, SQUID remains standard in MEG but is being superseded by optically-pumped magnetometers (OPM-MEG), which resolve many drawbacks of SQUID-MEG (Brookes et al., 2022). This history reflects key differences between EEG and MEG that stem from neurophysics: electrical fields radiate much further and are thus easier to detect than magnetic fields.

The costs and logistics of EEG are considered the best in neuroimaging (Fig. 4.2ab), aided by 150 years of maturation (Cooper et al., 2014). Modern EEG is wearable, conforms to individual head geometry, and is mobile to varying degrees. These features dampen artifacts and intersubject variability while enabling diverse use cases such as infants, interactive hyperscanning, and naturalistic paradigms (Bell & Cuevas, 2012). Moreover, EEG no longer requires electromagnetically-shielded rooms, as interference is well-handled by active shielding or denoising algorithms (Klug & Kloosterman, 2022). However, wet electrodes remain a logistical drawback, as conductive solution must be applied at least once per session to facilitate scalp contact, which is messy and time-consuming (Xu et al., 2017). This can also lead to electrical bridging, which arises from spillover of conductive solution between electrodes, resulting in identical recordings from affected sites. Dry electrodes are often contraindicated for

source-space EEG due to reduced signal-to-noise ratio (SNR). Though not free of challenges, EEG is exponentially cheaper and more flexible than SQUID-MEG or fMRI.

MEG is in the middle of a paradigm shift on its way to being as cheap and flexible as EEG (Fig. 4.2acd). Although the costs and logistics of standard SQUID-MEG are akin to fMRI, those of next-generation OPM-MEG are more comparable with high-end EEG systems (Brookes et al., 2022). Specifically, SQUID-MEG uses fixed-position sensors in a closet-sized cryogenic helmet (Fig. 4.2c), limiting use cases to adults and non-naturalistic paradigms (Hansen et al., 2010). Moreover, SQUID-MEG requires on-site technicians and specialized infrastructure such as extreme environmental shielding, much like fMRI. In contrast, OPM-MEG leverages quantum mechanical loopholes to provide EEG-like flexibility and costs while offering superior fidelity over SQUID-MEG (Hill et al., 2020). Specifically, OPM-MEG is wearable, head-conforming, mobile to varying extents, and requires no specialized staff or infrastructure (Iivanainen et al., 2019). Lastly, all MEG sensors are dry, conferring a logistical advantage over EEG.

The intrinsic fidelity of EEG, SQUID-MEG, and OPM-MEG generally ascends in that order (Fig. 4.2), with OPM-MEG needing the fewest sensors to achieve a given level of performance in metrics like SNR, functional discernment, and anatomical resolution (see Fig. 4.3b; Pratt et al., 2021; Brookes et al., 2022). Fidelity differences between EEG and MEG are often attributable to the neurophysics of conductance, which enhances propagation, distortion, and noise for electrical but not magnetic fields (Cho et al., 2015). Thus, when sensor density is equivalent, MEG might have better fidelity than EEG (Baillet, 2017), though differences can be nonsignificant (Hedrich et al., 2017). However, lack of conductance in MEG limits sensitivity to distant sources, resulting in worse anatomical coverage than EEG, especially in sulci and deeper brain regions (Fig. 4.3b; Samuelsson et al., 2019). SQUID-MEG systems often integrate EEG to fill coverage gaps. Meanwhile, OPM-MEG improves coverage and fidelity over SQUID-MEG by minimizing scalp-to-sensor distance among other enhancements (Iivanainen et al., 2019). As such, the

performance of 50-sensor OPM-MEG appears comparable to 275-sensor SQUID-MEG (Hill et al., 2020).

Additional sensors improve fidelity when comparing within but not between scanning modalities (i.e., EEG vs. SQUID-MEG) (Hansen et al., 2010; Cooper et al., 2014; Song et al., 2015). This is most apparent with EEG, which suffers from aliasing and other issues when using standard-density montages (SD-EEG, <128 sensors), leading to inaccurate source localization (see *source localization* below). Source-space EEG studies should therefore use high-density montages (HD-EEG, 128 to 256 sensors) or ultra-high-density montages (UHD-EEG, 264 to 1024 sensors). More sensors can overcome EEG's intrinsic fidelity limitations, with top UHD-EEG systems surpassing SQUID-MEG in many fidelity metrics (Petrov et al., 2014; H. S. Lee et al., 2022; Schreiner et al., 2023; J. Zhang et al., 2023). However, additional sensors worsen costs, logistics, and naturalism. For example, UHD-EEG is costlier and far more difficult with hair than HD-EEG. Moreover, sensor positions have a major impact on fidelity. For example, facial sensors improve coverage of deep and inferior brain regions (Samuelsson et al., 2019; Seeber et al., 2019), while sensors placed on the roof of the mouth provide excellent amygdala and hippocampus coverage (Tierney et al., 2021). Critically, only EEG and OPM-MEG have the advantage of movable sensors.

In sum, EEG is often cheaper, more flexible, and has wider anatomical coverage than MEG. However, MEG tends to have greater fidelity than EEG. Critically, SD-EEG should be avoided due to poor source localization performance. Both EEG and MEG transitioning to next-generation systems, namely UHD-EEG and OPM-MEG, which provide the best fidelity while offering favorable costs and flexibility over SQUID-MEG or fMRI. As of 2023, next-generation systems remain less common but are growing fast, particularly OPM-MEG.

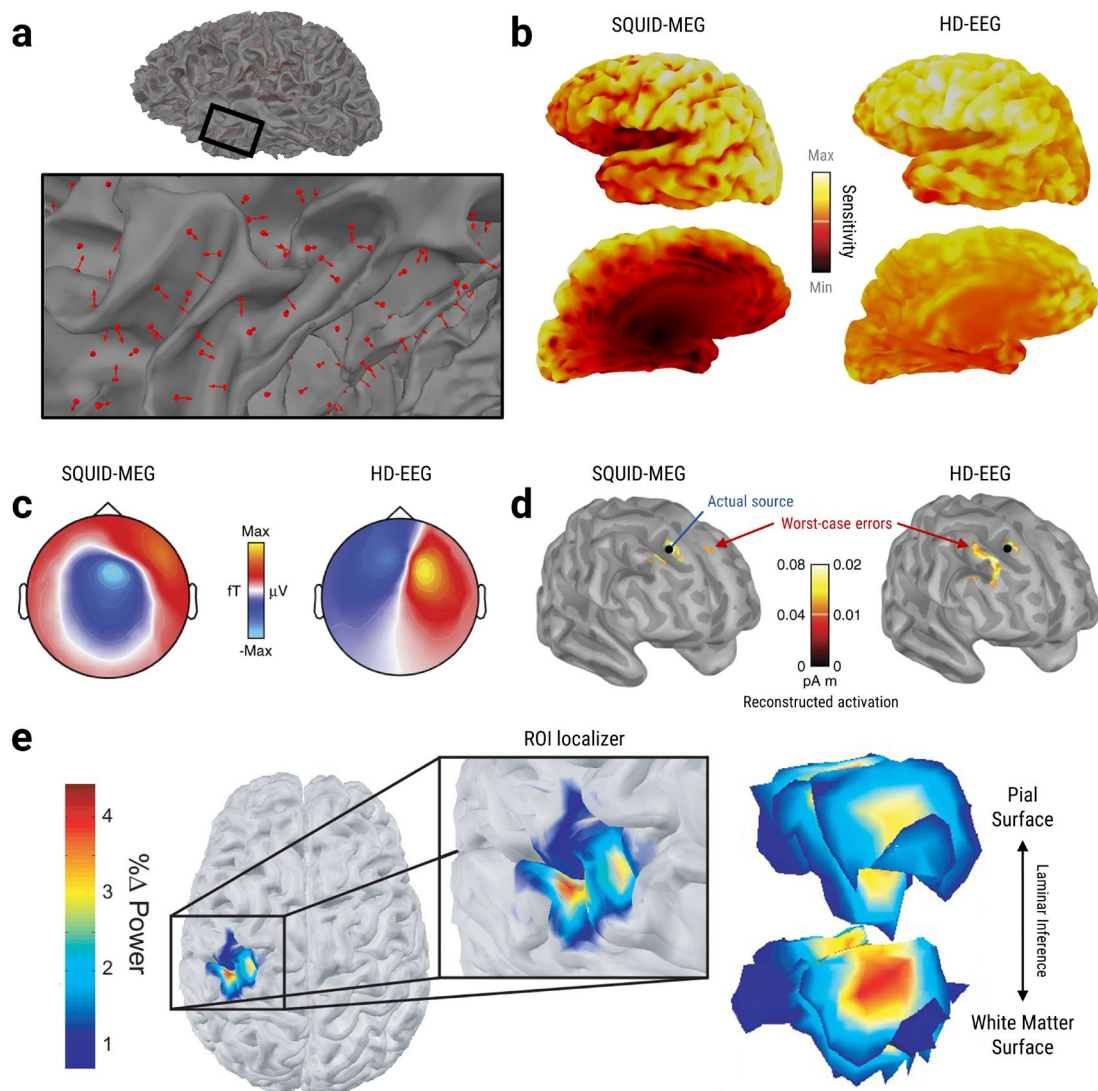


Figure 4.3: Source localization. **(a)** Example M/EEG source space showing candidate sources (red dots) and their dipole orientations (red arrows). Sources are typically restricted to grey matter and can also be restricted by the anatomical coverage (see panel B) and resolution of a given M/EEG system. Anatomical resolution is spatiotemporally-variable and defined by distances between differentiable sources. **(b)** Typical signal sensitivity of SQUID-MEG and HD-EEG. Figure from Samuelsson et al., 2019. **(c)** Simulated scalp topography of the neuronal source shown in panel D. Colors reflect electromagnetic polarity, not activation and deactivation. EEG and MEG have orthogonal scalp topography due to the “right-hand rule” of electromagnetism (Maxwell, 1865). Figure from Baillet, 2017. **(d)** Source localization of the scalp topography in panel C. Black dots indicate the actual neuronal source. Heatmaps reflect reconstructed activation strength. Red arrows highlight worst-case localization errors with realistic head models. **(e)** Laminar inference of source activity across superficial and deep cortical layers, which implement feedforward and feedback signaling, respectively. Laminar inference is bounded by the pial and white matter surfaces. Shown are spectral responses from an auditory localizer. Figure from Bonaiuto et al., 2018.

Source localization

Given the passive nature of M/EEG sensors (not to be confused with active shielding), source localization is needed to reveal the functional anatomy of M/EEG signals. Specifically,

M/EEG passively detect electromagnetic fields that escape the skull, while fMRI actively resonates each voxel location to detect structure and function (Lopes da Silva, 2013). Without source localization, M/EEG anatomical sources are not apparent from scalp topology, even for one source near the scalp (Fig. 4.3cd). Namely, electromagnetic fields are dipoles that radiate far from a source; positive and negative polarity (colors in Fig. 4.3c) are often mistaken for widespread activation and deactivation (Jackson & Bolger, 2014). Scalp topology also varies idiosyncratically with source location, depth, and orientation (Grech et al., 2008). Moreover, real-world scalp topology arises from countless overlapping dipoles, such that a single M/EEG sensor detects summed activity from much of the brain. Lastly, EEG and MEG have orthogonal scalp topography given the 'right-hand rule' of electromagnetism (Maxwell, 1865), complicating unlocalized cross-modal comparisons. Unsurprisingly, unlocalized M/EEG often disorients social neuroscientists.

EEG and MEG use near-identical methods and procedure for source localization, which involves leveraging multiple sensors to reconstruct n-dimensional functional anatomy (space, time, spectral frequency, and cortical laminae; Dalal et al., 2008; Wendel et al., 2009). Source localization is performed as a late preprocessing step in M/EEG software packages, the most source-focused of which include Fieldtrip (Oostenveld et al., 2010), MNE-Python (Gramfort et al., 2013), and Nutmeg (Hinkley et al., 2020). At minimum, reliable source localization requires sensor position measurements per recording, typically using stylus-based or photometric digitizers (Michel & Brunet, 2019). Moreover, structural MRI of each participant is recommended, though template MRI (e.g., MNI-305) can be used with reduced source localization performance (Cho et al., 2015). If using template MRI, participant head shape should be digitized along with sensor positions, as head shape is used to warp template MRI to approximate native MRI. Performance metrics include anatomical resolution (distances between differentiable sources;

Fig. 4.3a), coverage (source sensitivity distribution; Fig. 4.3b), and error (deviation from actual sources; Fig. 4.3d).

Source localization consists of two main components – the forward and inverse solutions – that are unique to each recording (Michel & Brunet, 2019). The forward solution models electromagnetic propagation from neuronal sources to scalp sensors (Hallez et al., 2007). First, a head model is generated from structural MRI, which is then coregistered with sensor positions. MEG typically uses single-shell head models, while EEG requires multi-shell models with conductance estimates (Cho et al., 2015). Head modeling is the only major modal difference in localization methods. Afterwards, the source space is defined within the head model (Zorzos et al., 2021). Source spaces consist of candidate sources (red dots in Fig. 4.3a) and their dipole orientations (red arrows). Candidate sources are usually restricted to grey matter for enhanced performance and biological realism. Laminar inference (Fig. 4.3e) places sources between the pial and white matter surfaces, acting as proxies for superficial and deep cortical layers, respectively (Bonaiuto, Rossiter, et al., 2018). Candidate sources can also be constrained by the anatomical coverage and resolution of a given M/EEG system. Lastly, Maxwell equations model the sensor-space presentation of each candidate source, producing the forward solution (Maxwell, 1865), which model the sensor-space presentation of dipoles from each candidate source. Packages like MNE-Python also estimate source-space signal sensitivity (Fig. 4.3b), which is useful for distinguishing genuine lack of activity from signal dropout (Gramfort et al., 2013).

The inverse solution decomposes sensor signals into the source space (Grech et al., 2008). Popular inverse algorithms often fall into two classes: minimum-norm estimate (MNE) and beamforming. MNE simultaneously estimates activity across candidate sources using the minimum current (L1/L2-norm) needed to explain sensor data with least error (Hämäläinen & Ilmoniemi, 1994). While classical MNE remains prevalent, common MNE variants include depth-

weighted MNE (wMNE) for mitigating depth bias (Lin et al., 2006), local autoregressive average (LAURA) for field potential modeling (Menendez et al., 2004), and dynamic statistical parametric mapping (dSPM) for fMRI integration (Dale et al., 2000). Meanwhile, beamformers independently estimate activity at each candidate source by applying spatial filters that pass certain signals while dampening crosstalk from others (Westner et al., 2022). Unlike MNE, beamformers falsely assume that sources are uncorrelated, though practical impacts of this are unclear (Jaiswal et al., 2020). Prevalent M/EEG beamformers include linearly-constrained minimum-variance (LCMV; van Drongelen et al., 1996) and time-frequency beamforming (TFBF) for spectral localization (Dalal et al., 2008). Of note, beamformers are well-developed from prolific everyday applications like cell phones, Wi-Fi, and acoustics. There are also many inverse algorithms that fall outside MNE and beamforming.

Inverse algorithms are far less standardized than forward algorithms, fomenting dizzying choices in algorithms and their parameters. Ideally, data should be set aside for trial-and-error. Tentatively, MNE may be best for diffuse sources and high-SNR recordings, while beamformers might be better for focal sources and low-SNR recordings (Hincapié et al., 2017; Michel & Brunet, 2019). Regardless, both MNE and beamforming are well-validated by invasive electrophysiology and simulation, which reveal reliable performance in many circumstances (Song et al., 2015; Hedrich et al., 2017; Liu et al., 2018; Jaiswal et al., 2020; Allouch et al., 2023). Beyond MNE and beamforming, recent deep learning and hierarchical Bayesian approaches show much promise (Zorzos et al., 2021). We strongly discourage the use of equivalent dipole fitting, the original inverse algorithm (Brody, 1968), which persists as the default option in popular software like EEGLAB (Delorme & Makeig, 2004), despite being biologically-implausible and largely inaccurate (Wendel et al., 2009). Note that laminar inference can use any inverse algorithm but involves unique considerations (see Bonaiuto, Rossiter, et al., 2018).

While fMRI anatomical resolution generally remains steady across voxels and time (typically 3 mm³), the same cannot be said for source-space M/EEG, whose performance varies with brain location (Fig. 4.3b) and neuronal dynamics. Using appropriate methods, anatomical resolution might average 10 mm³ for HD-EEG, 5 mm³ for SQUID-MEG, and 2 mm³ for top OPM-MEG systems (Fig. 4.2a), with worst-case errors limited to mislocalization of adjacent gyri or sulci (see Fig. 4.3d; Wendel et al., 2009; Hedrich et al., 2017; Bonaiuto, Rossiter, et al., 2018; Brookes et al., 2022). In contrast, insufficient sensor density (e.g., SD-EEG) could lead to errors larger than brain lobes (>8 cm³) in certain conditions, such as diffuse and highly-correlated sources (Song et al., 2015; Liu et al., 2018; Soler et al., 2022; Allouch et al., 2023). SD-EEG is highly affected by cross-source interference, which can misconstrue actual neuronal dynamics. Combining SD-EEG with template MRI is particularly deleterious (Brodbeck et al., 2011). However, SD-EEG might be sufficient for localizing focal activity, such as fusiform sources during face perception (Muukkonen et al., 2020). Localization issues become more severe with fewer sensors; 64-sensor EEG may often approach passable performance (e.g., <1 cm³ errors), while 16-sensor EEG might be unusable for source localization.

In sum, forward computation uses largely standardized methods, while inverse computation can involve dizzying choices. The most popular and well-validated inverse algorithms are often based on MNE and beamforming, both providing reliable performance in many situations. Critical pitfalls include omitted procedures (e.g., sensor position digitization), insufficient sensor density (e.g., SD-EEG), and obsolete algorithms (e.g., equivalent dipole fitting).

Implementation

In this section, we provide a step-by-step guide for applying source-space M/EEG methods to the Why/How task, a functional localizer for MNS and DMN regions (Spunt & Adolphs, 2014). Note that we did not collect this data but are giving a primer on what such a protocol would look like when modifying an fMRI paradigm for use with M/EEG. Each stimulus is shown with two prompt types: action identification (How) and mentalizing (Why). Stimuli are emotional facial expressions and intentional hand actions. We examine the spatiotemporal and spectral dynamics of evoked neuronal activity and connectivity through univariate and multivariate analyses. We include predicted results for the analyses described below.

To start, literature review of neurocognitive constructs should include invasive electrophysiology studies, which essentially provide a finer view of source-space M/EEG results (see Buzsáki et al., 2012). As for preprocessing and analysis, developing the best possible pipeline may entail mix-and-matching software packages (see prior section for prominent packages). To minimize human error and bias, automation should be maximized other than during development and exploration

(Baroumand et al., 2018). We have implemented some of the methods detailed below in electroCUDA, a GPU-accelerated electrophysiology package, though it is not currently a substitute for a dedicated M/EEG package (Tan, 2024).

		PROMPT	
		Why	How
STIMULUS	Emotions	Is the person feeling self-doubt? 	Is the person looking sideways? 
	Actions	Is the person competing? 	Is the person throwing a ball? 

Figure 4.4: Why/How task. This task is a popular fMRI localizer for DMN and MNS regions. Each stimulus is paired with two prompt types: mentalizing (Why) and action identification (How). There are two stimulus types: emotional facial expressions (Emotions) and intentional hand actions (Actions).

Task design

Electrophysiological tasks generally have fast event-related designs with as many trials as possible – blocked designs should be avoided (Cooper et al., 2014). Electrophysiological analyses typically use trial-by-trial baselines that better discern evoked activity compared to fMRI’s averaged baselines (Parvizi & Kastner, 2018). Baseline duration (i.e., intertrial intervals) should be longer than the slowest frequency-of-interest (FOI) to enable full-cycle oscillatory measurements (Kim & Im, 2018). Moreover, given that M/EEG crisply captures single-trial activity, incorporating rich trialwise data is fruitful, such as detailed stimulus metrics, behavioral response times (RT_{Behav}), and physiological measures like eye-tracking. Variable-duration trials can cause epoching issues in popular M/EEG software, however this may be worth it for added attentional capture and naturalism. Moreover, including a separate functional localizer may be even more important in source-space M/EEG than fMRI, given the multiple comparisons across space, time, spectral frequency, and cortical laminae.

We modify the Why/How task with the above recommendations, including switching from a blocked design to a fast event-related design. To obtain visual and semantic metrics for stimuli, we use CORnet, a brain-like recurrent convolutional model (Kubilius et al., 2019). Recurrent models better capture neurocognitive dynamics than non-recurrent designs (Kietzmann et al., 2019). Since this task is already a functional localizer, we use a priori region-of-interests (ROIs) from fMRI results (see Fig. 4.1a; Spunt & Adolphs, 2014).

Data collection

Since structural MRI can be the largest cost in source-space M/EEG, recruitment can focus on participants with existing structural MRIs, including former fMRI study participants (Wendel et al., 2009). To maximize datapoints, temporal precision, and spectral frequencies, M/EEG systems should be set to the highest feasible sampling rate (Kim & Im, 2018). Sampling

rate must be double the fastest FOI to avoid aliasing. Since most M/EEG signals are under 200 Hz, this will be our fastest FOI, with the closest appropriate sampling rate being 512 Hz in most systems (Muthukumaraswamy, 2013). If possible, sensors should be placed around the eye, jaw, and heart, which helps denoising algorithms subtract ocular, muscle, and cardiac artifacts. Make sure to digitize sensor positions per recording (see prior section). During experimental sessions, participant drowsiness can be inferred from delta and alpha activity on recording readouts (Johnson et al., 1969). Motion, blinks, and clenching artifacts can also be spotted on readouts, though these are generally well-handled *post-hoc* by our suggested preprocessing pipeline. It may be beneficial to allow tasks to be paused in case of drowsiness or other issues.

Early-stage preprocessing

For more on preprocessing, see these reviews (Peyk et al., 2011; Sheoran et al., 2015; Islam et al., 2016; Peng, 2019). Early-stage preprocessing includes steps like bad channel detection and referencing, the latter only necessary in EEG. Referencing involves subtracting a baseline timecourse from each sensor timecourse (channel), which drastically reduces noise (Cooper et al., 2014). Robust average referencing is generally preferred and involves subtracting the robust average of all good channels from each channel in a montage. Early-stage preprocessing can be performed with automated pipelines like PREP (Bigdely-Shamlo et al., 2015). Lastly, make sure to synchronize task metadata (e.g., timings) with M/EEG recordings.

Denoising

Denoising involves removing non-neuronal components in M/EEG signals, such as powerline, muscle, cardiac, ocular, and motion artifacts. HFB frequencies tend to be most affected by artifacts (Onton & Makeig, 2009). Moreover, electrophysiological signals become weaker at higher frequencies ($1/f$ decay) due to smaller neuronal ensembles and less efficient signal propagation, including greater scalp attenuation. As such, denoising is particularly

important for higher frequencies. Thankfully, even HFB can be reliably measured by M/EEG when using appropriate preprocessing (see Muthukumaraswamy, 2013).

Our suggested denoising pipeline is fully-automated but computationally intensive. Many less intensive pipelines are also available (see Pedroni et al., 2019; Miyakoshi et al., 2021). Our suggested denoising methods work best with ≥ 1 -Hz high-pass filtering, but such an aggressive high-pass distorts event-related potentials, a traditional M/EEG analysis that we will not perform (Winkler et al., 2015). Temporal filtering and spectral decomposition should be performed for each continuous recording with no disaggregation or concatenation (e.g., epoching) to avoid edge artifacts among other critical benefits.

First, we perform 1-Hz high-passing using a finite infinite response filter with low steepness to minimize artifacts. Next, we clean powerline noise from the electrical supply with Zapline-Plus, a nonstationary (temporally-variable) technique with data-driven parameter optimization (Klug & Kloosterman, 2022). Afterwards, we perform artifact subspace reconstruction (ASR) to remove nonstationary noise (Chang et al., 2020). Next, we perform independent components analysis (ICA) to reveal persistent noise sources, which comes with a choice of algorithms. Adaptive-mixture ICA provides nonstationary decomposition (Hsu et al., 2018), while *cudaICA* provides stationary decomposition but runs 25x faster due to GPU acceleration (Raimondo et al., 2012). Lastly, we use ICLabel to identify ICA noise components, which we then subtract from the data (Pion-Tonachini et al., 2019).

Source and spectral decomposition

Source localization is detailed in *methods* above (reviewed in Wendel et al., 2009; Michel & Brunet, 2019; Zorzos et al., 2021). Spectral decomposition isolates different frequencies (reviewed in Kim & Im, 2018). We will extract 4–200 Hz, encompassing theta, alpha, beta, gamma, and HFB bands. By default, source localization is performed on frequency-summed signals (i.e., normal M/EEG signals) without spectral decomposition. Performing spectral

decomposition first and then localizing FOIs can be more precise at the expense of flexibility. The opposite tradeoff occurs for post-localization spectral decomposition. Alternatively, inverse algorithms like time-frequency beamforming (TFBF) directly integrate spectral decomposition with localization computation to maximize precision (Dalal et al., 2008).

For spectral source localization, we use TFBF with Morse wavelets that account for unequal variance-covariance across frequencies, helping mitigate $1/f$ decay (Dalal et al., 2008). Additionally, we use wMNE (Lin et al., 2006) for localization of frequency-summed signals. Analyses should be performed in native anatomical space whenever feasible, as performance detriments from warping into standardized space (e.g., Montreal Neurological Institute; MNI) are larger with M/EEG than fMRI (Hallez et al., 2007). For example, ROIs should be anatomically or functionally defined in native space. Timecourses from native-space ROIs can then be extracted for further analysis.

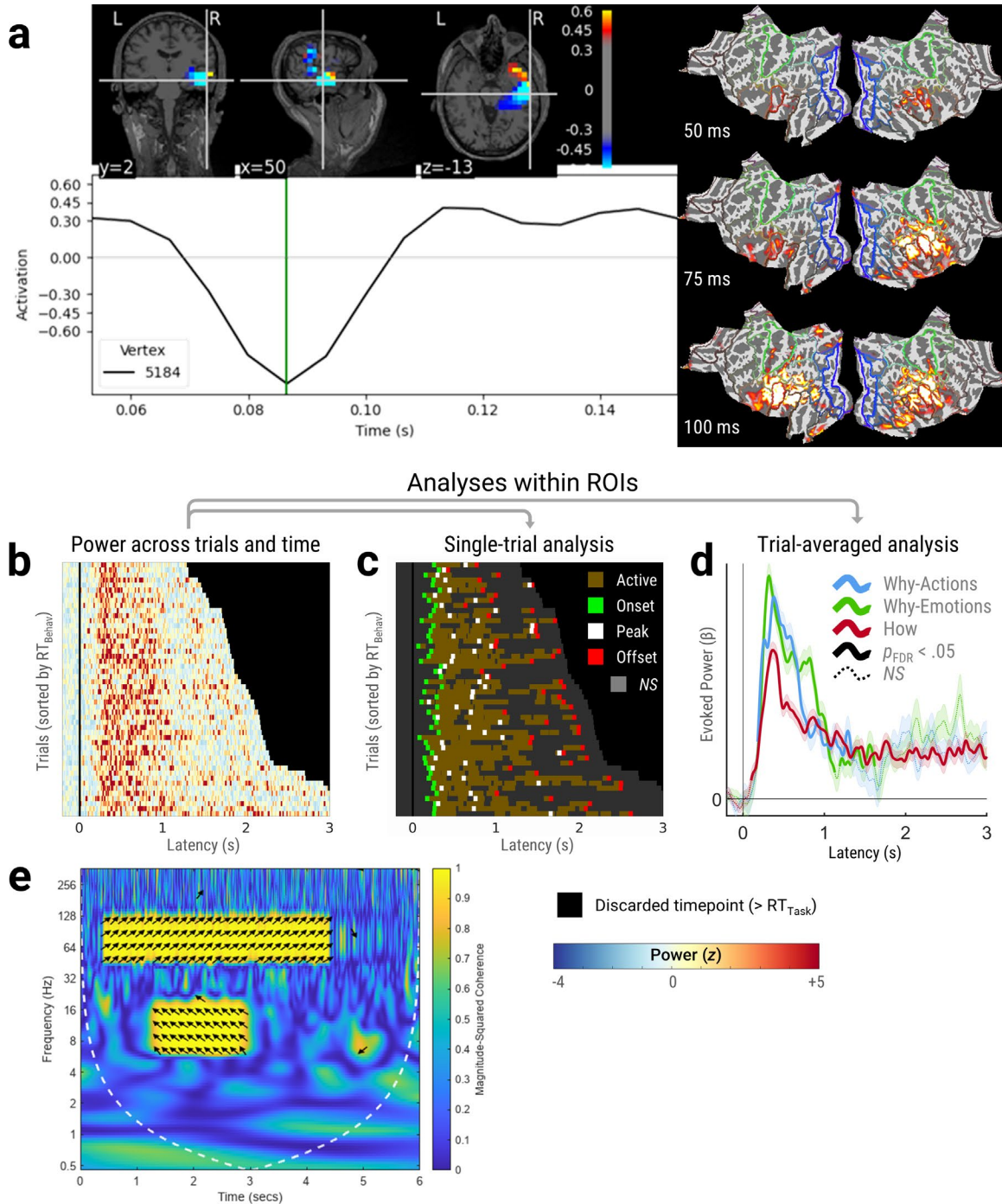


Figure 4.5: Univariate analyses. **(a)** Whole-brain MEG source analysis of task-evoked activity relative to prestimulus baseline. The MNE-Python interface is shown. The timecourse depicts trial-averaged source activity for the location indicated by crosshairs. The right side shows task-evoked activity at specific timepoints using flattened surface maps. **(b)** Trial-by-trial HFB power in an ROI. Black areas indicate post-trial timepoints. **(c)** Single-trial analysis of Panel B data. Timepoints with significant activity over prestimulus baseline are colored brown ($p_{FDR} < 0.05$, corrected across timepoints and trials). Onset latency (green) is the earliest significant timepoint, offset latency (red) is the latest significant timepoint, and duration is total significant timepoints. Peak latency and power (white) are the timepoint and magnitude, respectively, of the strongest activity. **(d)** Trial-averaged analysis of Panel B data. Solid colored lines indicate significant activity over from prestimulus baseline ($p_{FDR} < 0.05$, corrected across timepoints). Dashed lines indicate nonsignificant activity. **(e)** Effective connectivity (wavelet coherence) between ROIs. Arrows indicate directionality. This chart depicts recurrent processing consisting of feedforward gamma and HFB connectivity alongside theta and alpha feedback.

Whole-brain univariate analysis

Standard fMRI analyses are generally based on condition contrasts, with the Why-How contrast used as a standard localizer for the social brain. In this contrast, How prompts appear to activate MNS and deactivate DMN, while Why prompts evoke the opposite pattern (Spunt & Adolphs, 2014). However, this pattern could arise from relative differences in activation strength without actual deactivation from baseline. To reveal spatiotemporal neuronal dynamics without condition contrasts, we perform whole-brain univariate analysis of evoked activity over the prestimulus baseline in standard MNI space. For example, Fig. 4.5a shows the timecourse of a particular neuronal source (left) alongside whole-brain activity at given latencies (right). For the widest view, we first chart frequency-summed activity, perhaps revealing that both prompt types activate MNS, but only Why prompts activate DMN. Next, we repeat the prior analysis for each spectral band. This could chart the initial feedforward sweep of gamma and HFB from visual cortex, then MNS, followed by tpDMN, and lastly mPFC regions. Distinct spatiotemporal sequences might be found for theta, alpha, and beta. Critically, we use cluster-based multiple comparisons correction as appropriate across space, time, and spectral frequency (see Sassenhagen & Draschkow, 2019).

ROI univariate analyses

For more fine-grained spatiotemporal and spectral dynamics, we then perform univariate analyses using native-space ROIs. To reveal significant deviations from prestimulus baseline, we use linear mixed models nested within trial, providing trial-averaged and single-trial timecourses of evoked activity and their significance. Single-trial timecourses (Fig. 4.5b) can be used to extract key metrics of evoked responses in a particular ROI (Fig. 4.5c). Onset latency is the earliest significant timepoint, offset latency is the latest significant timepoint, and duration is total significant timepoints. Peak latency and power are the timepoint and magnitude,

respectively, of largest deviation from baseline. Trial-averaged timecourses for a single ROI (Fig. 4.5d) succinctly show temporal dynamics of each condition, revealing timepoints with significant activations or deactivations from baseline.

To reveal evoked activity across the electrophysiological spectrum, we perform trial-averaged analysis of spectral activations compared to baseline. This estimates ERSPs (see Fig. 4.1b) for each ROI and condition. We then perform a Why-How contrast to produce ERSP difference maps, which might show that significant differences first arise as increased HFB activity in DMN, followed by gradually increased theta and diminishing HFB in MNS. This would suggest that mentalizing-evoked DMN activity gradually deactivates MNS action representations activity via theta feedback.

To show significant latency differences across ROIs, we use single-trial activation metrics of non-spectral data. This analysis could show that activations first begin in visual cortex and propagate across MNS, tpDMN, and then mPFC regions, with significant onset differences between them. This may be followed by substantial activation overlap across all regions, suggesting recurrent processing.

ROI effective connectivity

To reveal information flow across ROIs, we measure time-varying effective connectivity with wavelet coherence (see Schoffelen & Gross, 2009; Sakkalis, 2011). For example, Fig. 4.5e shows feedforward HFB alongside a shorter period of theta feedback between two ROIs. To calculate significance from baseline, we compare single-trial connectivity strength at each timepoint versus prestimulus baseline. This analysis could show feedforward information flow mirroring the spatiotemporal sequence of onsets in the gamma and HFB bands, followed by distinct sequences of feedback bands across theta, alpha, and beta bands. Feedback might

begin with the first active MNS region to visual cortex, followed by intra-MNS feedback, succeeded by tpDMN-to-MNS feedback, and finally mPFC-to-tpDMN feedback.

Multivariate analyses

Source-space M/EEG is particularly suited for multivariate analyses given that data are multidimensional across space, time, spectral frequency, and cortical laminae – these dimensions can be used as input features (Fig. 4.6a; reviewed in Chichy & Pantazis, 2017). In contrast, multivariate fMRI typically only uses spatial features (see Weaverdyck et al., 2020). A common multivariate technique is classification, which predicts categories from features. Another popular technique is representational similarity analysis (RSA), which compares categories (e.g., stimuli) by input feature similarity. Here we focus on RSA but similar approaches can be used for classification.

To reveal how low-level visual features give rise to high-level mentalistic representations, we compared evoked neuronal activity with layers of the CORnet model, a brain-like recurrent convolutional neural network (Kubilius et al., 2019). Each CORnet layer represents stimuli distinctly, with

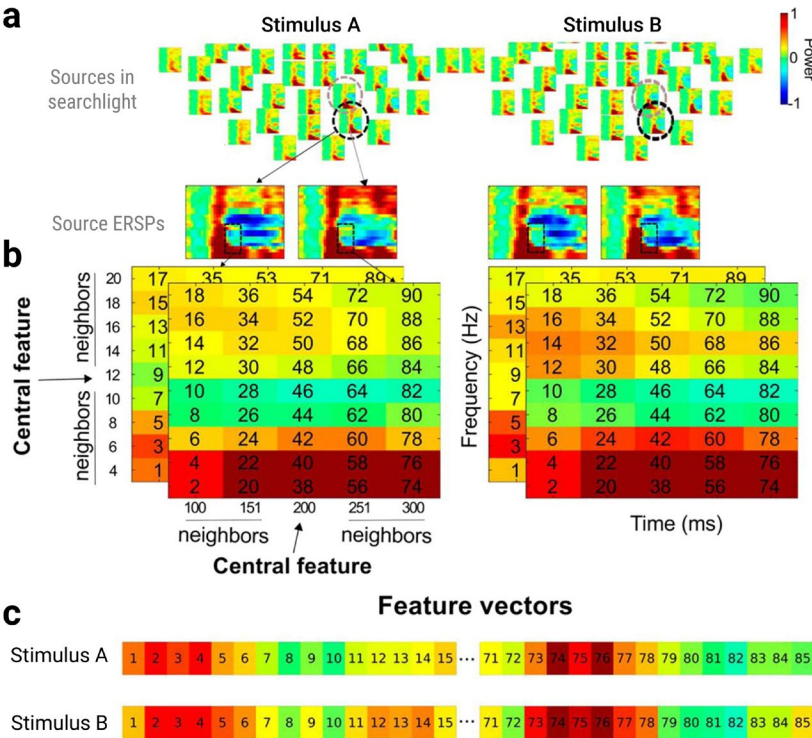


Figure 4.6: Multivariate analyses. **(a)** Sources within a 50 mm spherical searchlight. Source activity evoked by two stimuli are shown as separate ERSPs. **(b)** Searchlight features are extracted per timepoint, consisting of spectral power at each frequency from each neuronal source. **(c)** Concatenation of searchlight features into feature vectors for each stimulus. Figure adapted from Tucciarelli et al., 2015.

representations gradually transforming from visual to semantic across succeeding layers. We aim to chart an analogous transformation across the brain. To this end, we use RSA to compare neuronal stimulus encoding with each CORnet layer. The approach of comparing neuronal responses with machine learning layers has been used by many source-space M/EEG studies (Clarke et al., 2018; Kuzovkin et al., 2018; Kietzmann et al., 2019; Yang et al., 2019; von Seth et al., 2023).

To chart the broad spatiotemporal progression of visual-to-semantic transformation, we perform whole-brain searchlights across space, time, and frequency. Each searchlight encompasses sources within a 50 mm sphere of the central feature (Fig. 4.6a). For all sources within a searchlight, RSA is performed across time using spectral power at each frequency as features (Fig. 4.6b). Features are concatenated into feature vectors per stimulus, providing the basis for RSA. Statistical significance is determined by comparing similarity at each timepoint versus prestimulus baseline. This analysis could reveal that low-level visual CORnet layers are encoded similarly during Why and How prompts, propagating along the dorsal and ventral visual streams across time, echoing studies like Yang et al., 2019. Mid-level semantic layer encoding might arise in MNS and propagate to DMN, reflecting conceptual action representations as seen in Tucciarelli et al., 2015. High-level semantic layers might be better encoded during Why prompts. The highest semantic layer might be first encoded by mPFC, indicating mPFC sits atop the representational hierarchy during mentalizing.

To characterize underlying information flow of the observed visual-to-semantic transformation, we compare ROI effective connectivity and CORnet model layers. This approach was pioneered by von Seth et al., 2023. This analysis could reveal that recurrent MNS-DMN connectivity enhances mid-level semantic representations in MNS and high-level mentalistic representations in DMN. Specifically, alpha feedback from DMN might gate certain action

representations in MNS, in turn refining mentalistic representations in DMN, with discrete gamma and HFB frequencies encoding distinct types of feedforward information. Moreover, mPFC might send theta feedback based on highest-level mentalistic representations, guiding psycho-semantic processes in tpDMN regions (Lieberman, 2022), in turn enhancing feedforward inputs for propositional mentalistic reasoning in mPFC.

Conclusion

These predicted results are but limited examples of the level of mechanistic characterization possible with source-space M/EEG, providing a template for probing the social brain from a systems and computational perspective. Overall, we hope this article inspires social neuroscience to incorporate these modalities into its methodological repertoire.

General Discussion

The four papers of this dissertation approach a single question from complementary directions: what does the social brain look like when the dimensions fMRI collapses are restored? Paper I restores time; Papers II and III restore frequency and computation; Paper IV provides the route by which the field can pursue these dimensions without intracranial recording. Read together, they argue that what fMRI represents as a single activation is – on direct electrophysiological measurement – a sequence, a spectrum, and a computation. Recovering each one revises, rather than merely refines, the received account of mentalizing. Across the four, a single architectural theme recurs: the social brain operates as a parallel distributed processing system, whose organization becomes legible only once time and frequency are restored.

One activation, three dimensions

Functional MRI represents brain activity as one value per region per condition. That value faithfully summarizes one component of the underlying electrophysiology – high-frequency broadband (HFB) activity, which tracks local population spiking and to which the BOLD signal couples most tightly (Logothetis et al., 2001; Mukamel et al., 2005; Parvizi & Kastner, 2018). What it cannot represent is *when* that activity occurs at sub-second scales, *what frequencies* outside HFB carry the rest of the signal, and *what computations* the activity performs beyond the fact of engagement. Each empirical paper restores one of these. Adding time (Paper I) replaces co-activated DMN regions with a posterior-to-anterior sequence and recasts self/other differences as differences in duration rather than intensity. Adding frequency (Paper II) embeds that HFB sequence within a fuller spatiotemporospectral architecture and yields the dissertation's sharpest reinterpretation: the medial-prefrontal deactivation fMRI reports during

cognitive tasks corresponds, in dmPFC, to sustained HFB suppression co-occurring with sustained theta consistent with frontal-midline theta – two bands moving in opposite directions in the same sites. Adding computation (Paper III) shows that the architecture mentalizing shares with autobiographical memory carries shared computational content: mentalizing recruits autobiographical-like computation, graded along the same hierarchy and reliably short of full autobiographical engagement.

A parallel distributed architecture

Paper I framed mentalizing as a form of distributed hierarchical processing – a central organizing principle of neurocognitive systems, in which no single region computes a function alone (Hebb, 1949; Felleman & Van Essen, 1991; Huntenburg et al., 2018; Hilgetag & Goulas, 2020). The dimensions restored in Papers II and III extend this framing toward the parallel distributed processing (PDP) tradition, in which cognition emerges from graded, distributed representations and the parallel interactions among many units rather than from localized, all-or-none computations (Rumelhart & McClelland, 1986; McClelland et al., 1986). Each restored dimension supplies a piece this account requires. The posterior-to-anterior sequence (Paper I) is the distributed substrate: mentalizing is computed across a hierarchy of regions, not localized to a single ‘mentalizing area’. The dual-phase spatiotemporospectral architecture (Paper II) supplies the coordination such a distributed system requires – an early, broadly synchronous low-frequency phase that engages the network in parallel before any region performs its specific computation, followed by the content-specific cascade, with the frequency spectrum itself providing the multiplexed channels through which distributed regions communicate. And the graded recruitment of autobiographical-like computation (Paper III) – continuous rather than categorical, distributed across the DMN, and ordered as Autobiographical > Self $\approx$ Other >

Cognitive – is the graded, distributed representation a PDP account predicts, rather than the discrete category-membership a localist account would expect.

Read together, the three dimensions describe mentalizing as parallel distributed processing made visible: distributed across regions, coordinated and multiplexed across the spectrum, and representing computational content in graded form. This reframing connects the social brain to a broader principle of neural computation, and it reads the dimensions fMRI collapses not as incidental detail but as the very features that distinguish a parallel distributed account from the localist, modular one that single-value measurement tends to invite. We offer this as an organizing framework the data render newly legible, not as a mechanistic claim the present analyses test.

Two axes for self and other

Across Papers I and III, self- and other-mentalizing emerge as organized along two largely orthogonal axes. On a content axis – what computation is recruited – they are equivalent, engaging autobiographical-like computation to the same degree. On a temporal axis, they differ, with other-mentalizing reaching its processing later. Two methods sensitive to different things – activation timing in Paper I, computational content in Paper III – converge on the same conclusion: self and other differ in *when* they compute, not in *what*. This is a more parsimonious account than the magnitude- and selectivity-based framing inherited from fMRI.

Architecture and behavior

The dimensions restored here are not merely descriptive; they map onto behavior in a way that links the spatiotemporospectral architecture of Paper II to the computational dynamics of Paper III. The early, broadly synchronous low-frequency phase – shared across conditions – carries an early behavioral signature: at short latencies, recruitment toward the cognitive,

domain-general pole predicted faster responses, consistent with a fast, content-general engagement of the network preceding any specific computation. Later in the trial this relationship reversed, with autobiographical-like recruitment predicting faster responses as the content-specific cascade came online. Choice-predictive activity followed the same posterior-to-anterior hierarchy that carries autobiographical recruitment, escalating through the DMN and converging on medial prefrontal cortex near the time of response – a pattern consistent with evidence accumulating toward a decision, though we treat accumulation as a working hypothesis rather than a demonstrated mechanism. That this gradient was clearest for autobiographical memory itself, and weaker but present for mentalizing, is further evidence that the two draw on common machinery. Architecture and behavior thus tell one story: a fast domain-general phase that speeds responding, followed by a slower content-specific phase whose recruitment both distinguishes mentalizing from arithmetic and carries the information on which the behavioral choice is formed.

From apparent contradiction to reconciliation

Where the iEEG findings seem to contradict the fMRI account, the contradiction dissolves into a question of measurement. Frontal-midline theta was characterized for decades at the scalp, where its topography could not be cross-referenced against the specific DMN regions fMRI was independently mapping; source localization placed it in medial frontal cortex, and intracranial recordings have since confirmed those generators. The deactivation BOLD reports and the cognitive-control signal the electrophysiology literature describes are, at least in dmPFC, two aspects of the same activity, seen through instruments with complementary spectral sensitivity. The iEEG results thus do not overturn the canonical fMRI finding so much as resolve an apparent contradiction between the fMRI and cognitive-control literatures.

A methodological argument

The empirical claims of Papers I–III rest on temporal and spectral resolution that fMRI lacks by construction, and that the scalp-level M/EEG literature in social neuroscience has lacked the anatomical precision to exploit. Paper IV argues this is not a necessary limit: source-localized EEG and MEG can recover these dimensions non-invasively, at scale, and with anatomical precision sufficient to cross-reference the DMN regions fMRI has characterized. The intracranial work here is best understood as a demonstration of the structure that becomes visible when time and frequency are restored – structure the wider field can pursue with accessible tools.

Beyond mentalizing

The reinterpretations offered here concern mentalizing, but the errors they correct are not specific to it – they are failure modes of inferring neural computation from a measurement that collapses time, frequency, and computation into a single value. Each empirical paper supplies an existence proof of a distinct such error: a timing difference read as a magnitude difference (self versus other, Paper I); low-frequency activity invisible to hemodynamics read as disengagement (the dmPFC deactivation, Paper II); and co-activation read as shared computation (Paper III). None of these is unique to the social brain. Much of the fMRI literature in social and affective neuroscience rests on regional activations, deactivations, and selectivities inferred from the same one-dimensional signal, and the present results suggest that some unknown fraction of those inferences may be under-determined by the measurement that produced them – not necessarily wrong, but resting on a single dimension where the others might tell a different story. We do not claim to know how many; we claim that the question is now answerable, and worth asking. The broader contribution of this dissertation is therefore a template as much as a set of findings: a demonstration that the dimensions fMRI collapses can

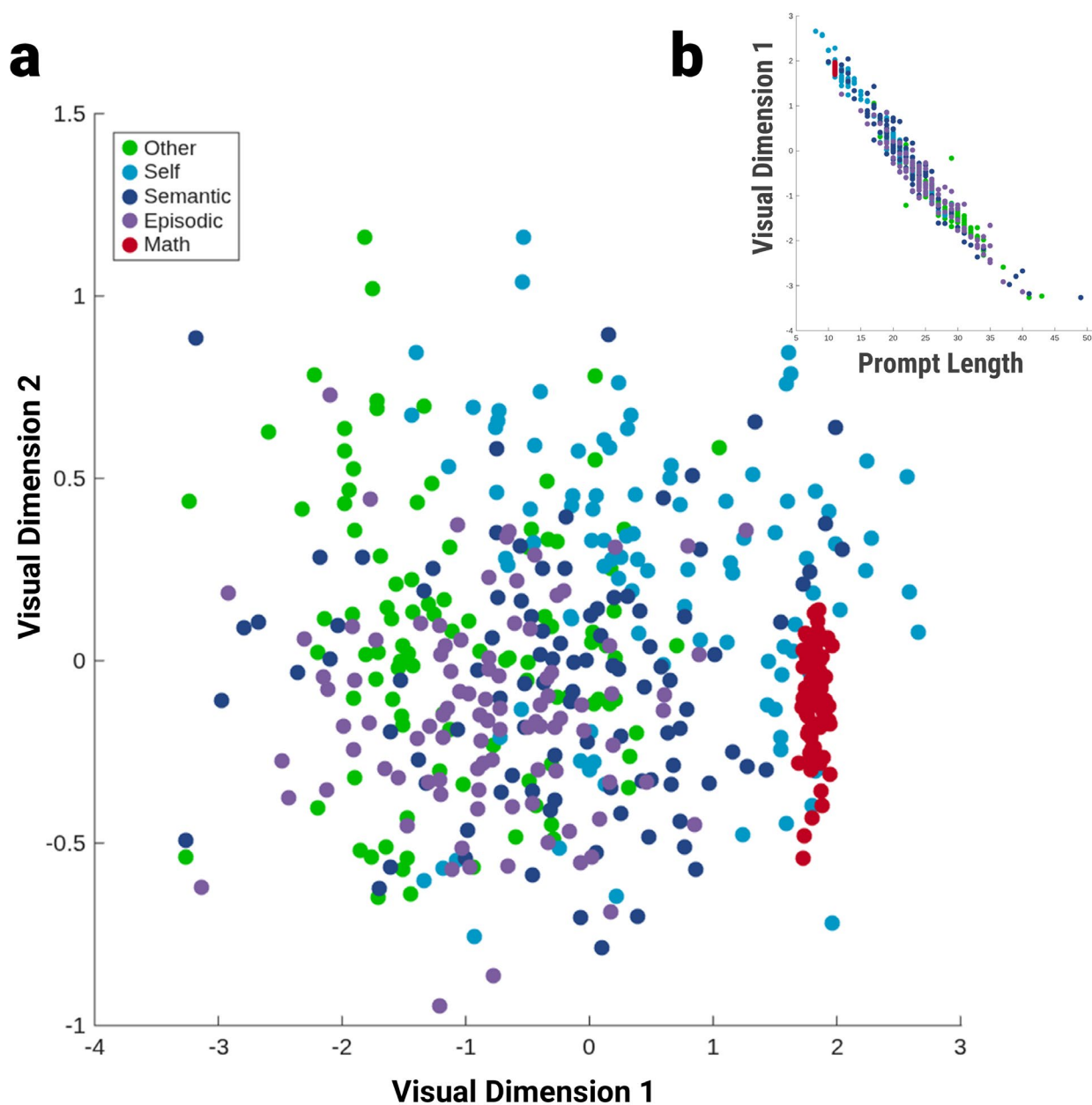
be restored, that doing so sometimes changes the conclusion, and that the tools to do it at scale are within the field's reach.

What the dimensions reveal

These reinterpretations are bounded. Cross-classification establishes shared spectral signatures, not directly observed shared computation; the frontal-midline theta identification is consistent with the cognitive-control literature but is not a mechanistic dissociation; the self/other timing difference in Paper III is qualitative; and the cognitive reading of the dual-phase architecture is a framework applied rather than tested. Each marks a tractable next experiment rather than an overreach. What remains across all of them is a single reframing: the social brain is not a set of regions that light up, but a set of regions organized in time, communicating across a spectrum of frequencies, and performing graded, distinguishable computations – a parallel distributed system whose organization becomes legible only when those dimensions are restored. The map fMRI built represents one dimension of that structure faithfully. This dissertation argues, by demonstration, that the rest of it is recoverable – and that recovering it does not erase the map, but shows what the map leaves out.

Supplemental Materials

Supplemental Figure 3.1



a) Stimulus visual dissimilarity as represented by visual dimensions (VDs) derived from AlexNet representational dissimilarity multidimensional scaling (see Methods). **b)** Comparison of VD1 and prompt length.

References

- Adaikkan, C., & Tsai, L.-H. (2020). Gamma Entrainment: Impact on Neurocircuits, Glia, and Therapeutic Opportunities. *Trends in Neurosciences*, 43(1), 24–41.
<https://doi.org/10.1016/j.tins.2019.11.001>
- Aguilera, M., Douchamps, V., Battaglia, D., & Goutagny, R. (2022). How Many Gammas? Redefining Hippocampal Theta-Gamma Dynamic During Spatial Learning. *Frontiers in Behavioral Neuroscience*, 16.
<https://www.frontiersin.org/articles/10.3389/fnbeh.2022.811278>
- Akalin Acar, Z., & Makeig, S. (2013). Effects of Forward Model Errors on EEG Source Localization. *Brain Topography*, 26(3), 378–396. <https://doi.org/10.1007/s10548-012-0274-6>
- Alamia, A., & VanRullen, R. (2019). Alpha oscillations and traveling waves: Signatures of predictive coding? *PLOS Biology*, 17(10), e3000487.
<https://doi.org/10.1371/journal.pbio.3000487>
- Alilović, J., Timmermans, B., Reteig, L. C., van Gaal, S., & Slagter, H. A. (2019). No Evidence that Predictions and Attention Modulate the First Feedforward Sweep of Cortical Information Processing. *Cerebral Cortex*, 29(5), 2261–2278. <https://doi.org/10.1093/cercor/bhz038>
- Allouch, S., Kabbara, A., Duprez, J., Khalil, M., Modolo, J., & Hassan, M. (2023). Effect of channel density, inverse solutions and connectivity measures on EEG resting-state networks reconstruction: A simulation study. *NeuroImage*, 271, 120006.
<https://doi.org/10.1016/j.neuroimage.2023.120006>
- Amft, M., Bzdok, D., Laird, A. R., Fox, P. T., Schilbach, L., & Eickhoff, S. B. (2015). Definition and characterization of an extended social-affective default network. *Brain Structure and Function*, 220(2), 1031–1049. <https://doi.org/10.1007/s00429-013-0698-0>

- Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, 65(4), 550–562.
<https://doi.org/10.1016/j.neuron.2010.02.005>
- Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, 1316(1), 29–52.
<https://doi.org/10.1111/nyas.12360>
- Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and Temporo-Parietal Lobe Contributions to Theory of Mind: Neuropsychological Evidence from a False-Belief Task with Reduced Language and Executive Demands. *Journal of Cognitive Neuroscience*, 16(10), 1773–1784. <https://doi.org/10.1162/0898929042947928>
- Asada, H., Fukuda, Y., Tsunoda, S., Yamaguchi, M., & Tonoike, M. (1999). Frontal midline theta rhythms reflect alternative activation of prefrontal cortex and anterior cingulate cortex in humans. *Neuroscience Letters*, 274(1), 29–32. [https://doi.org/10.1016/S0304-3940\(99\)00679-5](https://doi.org/10.1016/S0304-3940(99)00679-5)
- Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, 20(3), Article 3. <https://doi.org/10.1038/nn.4504>
- Barone, J., & Rossiter, H. E. (2021). Understanding the Role of Sensorimotor Beta Oscillations. *Frontiers in Systems Neuroscience*, 15.
<https://www.frontiersin.org/articles/10.3389/fnsys.2021.655886>
- Baroumand, A. G., van Mierlo, P., Strobbe, G., Pinborg, L. H., Fabricius, M., Rubboli, G., Leffers, A.-M., Uldall, P., Jespersen, B., Brennum, J., Henriksen, O. M., & Beniczky, S. (2018). Automated EEG source imaging: A retrospective, blinded clinical validation study. *Clinical Neurophysiology*, 129(11), 2403–2410.
<https://doi.org/10.1016/j.clinph.2018.09.015>

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
<https://doi.org/10.1016/j.jml.2012.11.001>
- Barrett, L. F., & Bar, M. (2009). See it with feeling: Affective predictions during object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1325–1334. <https://doi.org/10.1098/rstb.2008.0312>
- Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology, Social and Emotional Neuroscience*, 23(3), 361–372.
<https://doi.org/10.1016/j.conb.2012.12.012>
- Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00016>
- Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational Interaction in the Scanner: Mentalizing during Language Processing as Revealed by MEG. *Cerebral Cortex*, 25(9), 3219–3234. <https://doi.org/10.1093/cercor/bhu116>
- Bonaiuto, J. J., Meyer, S. S., Little, S., Rossiter, H., Callaghan, M. F., Dick, F., Barnes, G. R., & Bestmann, S. (2018). Lamina-specific cortical dynamics in human visual and sensorimotor cortices. *eLife*, 7, e33977. <https://doi.org/10.7554/eLife.33977>
- Bonaiuto, J. J., Rossiter, H. E., Meyer, S. S., Adams, N., Little, S., Callaghan, M. F., Dick, F., Bestmann, S., & Barnes, G. R. (2018). Non-invasive laminar inference with MEG: Comparison of methods and source inversion algorithms. *NeuroImage*, 167, 372–383.
<https://doi.org/10.1016/j.neuroimage.2017.11.068>
- Bowman, A. D., Griffis, J. C., Visscher, K. M., Dobbins, A. C., Gawne, T. J., DiFrancesco, M. W., & Szaflarski, J. P. (2017). Relationship Between Alpha Rhythm and the Default Mode

- Network: An EEG-fMRI Study. *Journal of Clinical Neurophysiology*, 34(6), 527.
<https://doi.org/10.1097/WNP.0000000000000411>
- Brody, D. A. (1968). The Inverse Determination of Simple Generator Configurations from Equivalent Dipole and Multipole Information. *IEEE Transactions on Biomedical Engineering, BME-15*(2), 106–110. *IEEE Transactions on Biomedical Engineering*.
<https://doi.org/10.1109/TBME.1968.4502545>
- Brookes, M. J., Leggett, J., Rea, M., Hill, R. M., Holmes, N., Boto, E., & Bowtell, R. (2022). Magnetoencephalography with optically pumped magnetometers (OPM-MEG): The next generation of functional neuroimaging. *Trends in Neurosciences*, 45(8), 621–634.
<https://doi.org/10.1016/j.tins.2022.05.008>
- Brunsdon, V. E. A., Bradford, E. E. F., & Ferguson, H. J. (2019). Sensorimotor mu rhythm during action observation changes across the lifespan independently from social cognitive processes. *Developmental Cognitive Neuroscience*, 38, 100659.
<https://doi.org/10.1016/j.dcn.2019.100659>
- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49–57. <https://doi.org/10.1016/j.tics.2006.11.004>
- Buckner, R. L., & DiNicola, L. M. (2019). The brain's default network: Updated anatomy, physiology and evolving insights. *Nature Reviews Neuroscience*, 20(10), 593–608.
<https://doi.org/10.1038/s41583-019-0212-7>
- Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.
- Buzsáki, G. (2010). Neural Syntax: Cell Assemblies, Synapsembles, and Readers. *Neuron*, 68(3), 362–385. <https://doi.org/10.1016/j.neuron.2010.09.023>
- Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6), Article 6.
<https://doi.org/10.1038/nrn3241>

- Buzsáki, G., & Draguhn, A. (2004). Neuronal Oscillations in Cortical Networks. *Science*, 304(5679), 1926–1929. <https://doi.org/10.1126/science.1099745>
- Buzsáki, G., & Wang, X.-J. (2012). Mechanisms of Gamma Oscillations. *Annual Review of Neuroscience*, 35(1), 203–225. <https://doi.org/10.1146/annurev-neuro-062111-150444>
- Bzdok, D., Langner, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., & Eickhoff, S. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00232>
- Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11), 506–515. <https://doi.org/10.1016/j.tics.2010.09.001>
- Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition*, 36, 426–433. <https://doi.org/10.1016/j.concog.2015.03.012>
- Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, 18(8), 414–421. <https://doi.org/10.1016/j.tics.2014.04.012>
- Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., & Jung, T.-P. (2020). Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. *IEEE Transactions on Biomedical Engineering*, 67(4), 1114–1121. IEEE Transactions on Biomedical Engineering. <https://doi.org/10.1109/TBME.2019.2930186>
- Cichy, R. M., & Pantazis, D. (2017). Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *NeuroImage*, 158, 441–454. <https://doi.org/10.1016/j.neuroimage.2017.07.023>
- Clarke, A., Devereux, B. J., & Tyler, L. K. (2018). Oscillatory Dynamics of Perceptual to Conceptual Transformations in the Ventral Visual Pathway. *Journal of Cognitive Neuroscience*, 30(11), 1590–1605. https://doi.org/10.1162/jocn_a_01325

- Cohen, D. (1972). Magnetoencephalography: Detection of the Brain's Electrical Activity with a Superconducting Magnetometer. *Science*, 175(4022), 664–666.
<https://doi.org/10.1126/science.175.4022.664>
- Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends in Neurosciences*, 37(9), 480–490. <https://doi.org/10.1016/j.tins.2014.06.004>
- Cohen, M. X., Ridderinkhof, K. R., Haupt, S., Elger, C. E., & Fell, J. (2008). Medial frontal cortex and response conflict: Evidence from human intracranial EEG and medial frontal cortex lesion. *Brain Research*, 1238, 127–142. <https://doi.org/10.1016/j.brainres.2008.07.114>
- Cona, G., Chiossi, F., Di Tomasso, S., Pellegrino, G., Piccione, F., Bisiacchi, P., & Arcara, G. (2020). Theta and alpha oscillations as signatures of internal and external attention to delayed intentions: A magnetoencephalography (MEG) study. *NeuroImage*, 205, 116295.
<https://doi.org/10.1016/j.neuroimage.2019.116295>
- Cooper, R., Osselton, J. W., & Shaw, J. C. (2014). *EEG Technology*. Butterworth-Heinemann.
- Dalal, S. S., Guggisberg, A. G., Edwards, E., Sekihara, K., Findlay, A. M., Canolty, R. T., Berger, M. S., Knight, R. T., Barbaro, N. M., Kirsch, H. E., & Nagarajan, S. S. (2008). Five-dimensional neuroimaging: Localization of the time–frequency dynamics of cortical activity. *NeuroImage*, 40(4), 1686–1700. <https://doi.org/10.1016/j.neuroimage.2008.01.023>
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, 26(1), 55–67.
[https://doi.org/10.1016/S0896-6273\(00\)81138-1](https://doi.org/10.1016/S0896-6273(00)81138-1)
- Das, A., de los Angeles, C., & Menon, V. (2022). Electrophysiological foundations of the human default-mode network revealed by intracranial-EEG recordings during resting-state and cognition. *NeuroImage*, 250, 118927.
<https://doi.org/10.1016/j.neuroimage.2022.118927>

- Das, A., & Menon, V. (2024). Frequency-specific directed connectivity between the hippocampus and parietal cortex during verbal and spatial episodic memory: An intracranial EEG replication. *Cerebral Cortex*, 34(7), bhae287. <https://doi.org/10.1093/cercor/bhae287>
- de Cheveigné, A., & Arzounian, D. (2018). Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *NeuroImage*, 172, 903–912. <https://doi.org/10.1016/j.neuroimage.2018.01.035>
- de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016). Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *The Journal of Neuroscience*, 36(24), 6553–6562. <https://doi.org/10.1523/JNEUROSCI.4402-15.2016>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <http://www.sciencedirect.com/science/article/pii/S0165027003003479>
- Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., Essen, D. C. V., Sotiropoulos, S. N., Anticevic, A., & Murray, J. D. (2019). Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics. *Neuron*, 0(0). <https://doi.org/10.1016/j.neuron.2019.01.017>
- Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A Meta-Analysis of Functional Neuroimaging Studies of Self and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, 24(8), 1742–1752. https://doi.org/10.1162/jocn_a_00233
- Diaconescu, A. O., Litvak, V., Mathys, C., Kasper, L., Friston, K. J., & Stephan, K. E. (2017). A computational hierarchy in human cortex. *arXiv:1709.02323 [q-Bio]*. <http://arxiv.org/abs/1709.02323>

- DiCarlo, J. J., & Maunsell, J. H. R. (2005). Using Neuronal Latency to Determine Sensory–Motor Processing Pathways in Reaction Time Tasks. *Journal of Neurophysiology*, 93(5), 2974–2986. <https://doi.org/10.1152/jn.00508.2004>
- Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, 18(5), Article 5. <https://doi.org/10.1038/nn.3981>
- Dykstra, A. R., Chan, A. M., Quinn, B. T., Zepeda, R., Keller, C. J., Cormier, J., Madsen, J. R., Eskandar, E. N., & Cash, S. S. (2012). Individualized localization and cortical surface-based registration of intracranial electrodes. *NeuroImage*, 59(4), 3563–3570. <https://doi.org/10.1016/j.neuroimage.2011.11.046>
- Engel, A. K., & Fries, P. (2010). Beta-band oscillations—Signalling the status quo? *Current Opinion in Neurobiology, Cognitive Neuroscience*, 20(2), 156–165. <https://doi.org/10.1016/j.conb.2010.02.015>
- Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, 2(10), 704–716. <https://doi.org/10.1038/35094565>
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1), 1–47. <https://doi.org/10.1093/cercor/1.1.1>
- Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62(2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Foster, B. L., Rangarajan, V., Shirer, W. R., & Parvizi, J. (2015). Intrinsic and Task-Dependent Coupling of Neuronal Population Activity in Human Parietal Cortex. *Neuron*, 86(2), 578–590. <https://doi.org/10.1016/j.neuron.2015.03.018>

- Fox, K. C. R., Foster, B. L., Kucyi, A., Daitch, A. L., & Parvizi, J. (2018). Intracranial Electrophysiology of the Human Default Network. *Trends in Cognitive Sciences*, 22(4), 307–324. <https://doi.org/10.1016/j.tics.2018.02.002>
- Friedrich, E. V. C., Hilla, Y., Sterner, E. F., Ostermeier, S. S., Behnke, L., & Sauseng, P. (2026). Theta-gamma phase amplitude coupling serves as a marker of social cognition and visual working memory deficits in individuals with elevated autistic traits. *Communications Psychology*. <https://doi.org/10.1038/s44271-025-00392-6>
- Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1), 220–235. <https://doi.org/10.1016/j.neuron.2015.09.034>
- Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, 50(4), 531–534. <https://doi.org/10.1016/j.neuron.2006.05.001>
- Frith, C. D., & Frith, U. (2008). Implicit and Explicit Processes in Social Cognition. *Neuron*, 60(3), 503–510. <https://doi.org/10.1016/j.neuron.2008.10.032>
- Gaylor, D. W., & Hopper, F. N. (1969). Estimating the Degrees of Freedom for Linear Combinations of Mean Squares by Satterthwaite's Formula. *Technometrics*, 11(4), 691–706. <https://doi.org/10.2307/1266893>
- Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, 7(4), 374–385. <https://doi.org/10.1093/cercor/7.4.374>
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535–574. <https://doi.org/10.1146/annurev.neuro.29.051605.113038>
- Grave de Peralta Menendez, R., Murray, M. M., Michel, C. M., Martuzzi, R., & Gonzalez Andino, S. L. (2004). Electrical neuroimaging based on biophysical constraints. *NeuroImage*, 21(2), 527–539. <https://doi.org/10.1016/j.neuroimage.2003.09.051>

- Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 5(1), 25.
<https://doi.org/10.1186/1743-0003-5-25>
- Groppe, D. M., Bickel, S., Dykstra, A. R., Wang, X., Mégevand, P., Mercier, M. R., Lado, F. A., Mehta, A. D., & Honey, C. J. (2017). iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. *Journal of Neuroscience Methods*, 281, 40–48. <https://doi.org/10.1016/j.jneumeth.2017.01.022>
- Gu, Y., Sainburg, L. E., Kuang, S., Han, F., Williams, J. W., Liu, Y., Zhang, N., Zhang, X., Leopold, D. A., & Liu, X. (2021). Brain Activity Fluctuations Propagate as Waves Traversing the Cortical Hierarchy. *Cerebral Cortex*, 31(9), 3986–4005.
<https://doi.org/10.1093/cercor/bhab064>
- Halgren, E., Kaestner, E., Marinkovic, K., Cash, S. S., Wang, C., Schomer, D. L., Madsen, J. R., & Ulbert, I. (2015). Laminar profile of spontaneous and evoked theta: Rhythmic modulation of cortical processing during word integration. *Neuropsychologia, Special Issue: Semantic Cognition*, 76, 108–124.
<https://doi.org/10.1016/j.neuropsychologia.2015.03.021>
- Halgren, M., Ulbert, I., Bastuji, H., Fabó, D., Erőss, L., Rey, M., Devinsky, O., Doyle, W. K., Mak-McCully, R., Halgren, E., Wittner, L., Chauvel, P., Heit, G., Eskandar, E., Mandell, A., & Cash, S. S. (2019). The generation and propagation of the human alpha rhythm. *Proceedings of the National Academy of Sciences*, 116(47), 23772–23782.
<https://doi.org/10.1073/pnas.1913092116>
- Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D’Asseler, Y., Camilleri, K. P., Fabri, S. G., Van Huffel, S., & Lemahieu, I. (2007). Review on solving the forward

- problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 4(1), 46.
<https://doi.org/10.1186/1743-0003-4-46>
- Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1), 35–42.
<https://doi.org/10.1007/BF02512476>
- Hansen, P., Kringelbach, M., & Salmelin, R. (2010). *MEG: An Introduction to Methods*. Oxford University Press.
- Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. Wiley.
- Hedrich, T., Pellegrino, G., Kobayashi, E., Lina, J. M., & Grova, C. (2017). Comparison of the spatial resolution of source imaging techniques in high-density EEG and MEG. *NeuroImage*, 157, 531–544. <https://doi.org/10.1016/j.neuroimage.2017.06.022>
- Helfrich, R. F., Fiebelkorn, I. C., Szczepanski, S. M., Lin, J. J., Parvizi, J., Knight, R. T., & Kastner, S. (2018). Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron*, 99(4), 854-865.e5. <https://doi.org/10.1016/j.neuron.2018.07.032>
- Hermes, D., Miller, K. J., Noordmans, H. J., Vansteensel, M. J., & Ramsey, N. F. (2010). Automated electrocorticographic electrode localization on individually rendered brain surfaces. *Journal of Neuroscience Methods*, 185(2), 293–298.
<https://doi.org/10.1016/j.jneumeth.2009.10.005>
- Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta Oscillations in Human Memory. *Trends in Cognitive Sciences*, 24(3), 208–227. <https://doi.org/10.1016/j.tics.2019.12.006>
- Hilgetag, C. C., & Goulas, A. (2020). ‘Hierarchy’ in the organization of brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1796), 20190319. <https://doi.org/10.1098/rstb.2019.0319>
- Hincapié, A.-S., Kujala, J., Mattout, J., Pascarella, A., Daligault, S., Delpuech, C., Mery, D., Cosmelli, D., & Jerbi, K. (2017). The impact of MEG source reconstruction method on

- source-space connectivity estimation: A comparison between minimum-norm solution and beamforming. *NeuroImage*, 156, 29–42.
- <https://doi.org/10.1016/j.neuroimage.2017.04.038>
- Hobson, H. M., & Bishop, D. V. M. (2017). The interpretation of mu suppression as an index of mirror neuron activity: Past, present and future. *Royal Society Open Science*, 4(3), 160662. <https://doi.org/10.1098/rsos.160662>
- Holmes, N., Rea, M., Hill, R. M., Boto, E., Leggett, J., Edwards, L. J., Rhodes, N., Shah, V., Osborne, J., Fromhold, T. M., Glover, P., Montague, P. R., Brookes, M. J., & Bowtell, R. (2023). Naturalistic Hyperscanning with Wearable Magnetoencephalography. *Sensors*, 23(12), Article 12. <https://doi.org/10.3390/s23125454>
- Hox, J. J., Moerbeek, M., & Schoot, R. van de. (2017). *Multilevel Analysis: Techniques and Applications, Third Edition*. Routledge.
- Hsieh, L.-T., & Ranganath, C. (2014). Frontal Midline Theta Oscillations during Working Memory Maintenance and Episodic Encoding and Retrieval. *NeuroImage*, 85(0 2), 10.1016/j.neuroimage.2013.08.003. <https://doi.org/10.1016/j.neuroimage.2013.08.003>
- Hsu, S.-H., Pion-Tonachini, L., Palmer, J., Miyakoshi, M., Makeig, S., & Jung, T.-P. (2018). Modeling brain dynamic state changes with adaptive mixture independent component analysis. *NeuroImage*, 183, 47–61. <https://doi.org/10.1016/j.neuroimage.2018.08.001>
- Hughes, C., Setton, R., Mwilambwe-Tshilobo, L., Baracchini, G., Turner, G. R., & Spreng, R. N. (2024). Precision mapping of the default network reveals common and distinct (inter) activity for autobiographical memory and theory of mind. *Journal of Neurophysiology*. (Rockville, MD). <https://doi.org/10.1152/jn.00427.2023>
- Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience*, 15(3), 470–476. <https://doi.org/10.1038/nn.3017>

- Huntenburg, J. M., Bazin, P.-L., & Margulies, D. S. (2018). Large-Scale Gradients in Human Cortical Organization. *Trends in Cognitive Sciences*, 22(1), 21–31.
<https://doi.org/10.1016/j.tics.2017.11.002>
- Imperator, C., Della Marca, G., Amoroso, N., Maestoso, G., Valenti, E. M., Massullo, C., Carbone, G. A., Contardi, A., & Farina, B. (2017). Alpha/Theta Neurofeedback Increases Mentalization and Default Mode Network Connectivity in a Non-Clinical Sample. *Brain Topography*, 30(6), 822–831. <https://doi.org/10.1007/s10548-017-0593-8>
- Ishii, R., Shinosaki, K., Ukai, S., Inouye, T., Ishihara, T., Yoshimine, T., Hirabuki, N., Asada, H., Kihara, T., Robinson, S. E., & Takeda, M. (1999). Medial prefrontal cortex generates frontal midline theta rhythm. *NeuroReport*, 10(4), 675.
https://journals.lww.com/neuroreport/abstract/1999/03170/medial_prefrontal_cortex_generates_frontal_midline.3.aspx
- Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 46(4), 287–305.
<https://doi.org/10.1016/j.neucli.2016.07.002>
- Jackson, A. F., & Bolger, D. J. (2014). The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. *Psychophysiology*, 51(11), 1061–1071.
<https://doi.org/10.1111/psyp.12283>
- Jaiswal, A., Nenonen, J., Stenroos, M., Gramfort, A., Dalal, S. S., Westner, B. U., Litvak, V., Mosher, J. C., Schoffelen, J.-M., Witton, C., Oostenveld, R., & Parkkonen, L. (2020). Comparison of beamformer implementations for MEG source localization. *NeuroImage*, 216, 116797.
<https://doi.org/10.1016/j.neuroimage.2020.116797>
- Jarrett, K., Kavukcuoglu, K., Ranzato, M., & LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? *2009 IEEE 12th International Conference on Computer Vision*, 2146–2153. <https://doi.org/10.1109/ICCV.2009.5459469>

- Jensen, O., & Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences*, 11(7), 267–269. <https://doi.org/10.1016/j.tics.2007.05.003>
- Jeon, H.-A. (2014). Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*, 8. <https://doi.org/10.3389/fnsys.2014.00223>
- Johnson, L., Lubin, A., Naitoh, P., Nute, C., & Austin, M. (1969). Spectral analysis of the EEG of dominant and non-dominant alpha subjects during waking and sleeping. *Electroencephalography and Clinical Neurophysiology*, 26(4), 361–370. [https://doi.org/10.1016/0013-4694\(69\)90086-8](https://doi.org/10.1016/0013-4694(69)90086-8)
- Kam, J. W. Y., Lin, J. J., Solbakk, A.-K., Endestad, T., Larsson, P. G., & Knight, R. T. (2019). Default network and frontoparietal control network theta connectivity supports internal attention. *Nature Human Behaviour*, 3(12), Article 12. <https://doi.org/10.1038/s41562-019-0717-0>
- Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, 116(43), 21854–21863. <https://doi.org/10.1073/pnas.1905544116>
- Kim, D.-W., & Im, C.-H. (2018). EEG Spectral Analysis. In C.-H. Im (Ed.), *Computational EEG Analysis: Methods and Applications* (pp. 35–53). Springer. https://doi.org/10.1007/978-981-13-0908-3_3
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & others. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14), 1.
- Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, 43(9), 2743–2758. <https://doi.org/10.1002/hbm.25832>

- Kopell, N., Whittington, M. A., & Kramer, M. A. (2011). Neuronal assembly dynamics in the beta1 frequency range permits short-term memory. *Proceedings of the National Academy of Sciences*, 108(9), 3779–3784. <https://doi.org/10.1073/pnas.1019676108>
- Kubilius, J., Schrimpf, M., Kar, K., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Bashivan, P., Prescott-Roy, J., Schmidt, K., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2019). *Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs* (arXiv:1909.06161). arXiv. <https://doi.org/10.48550/arXiv.1909.06161>
- Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J.-P., Baciú, M., Kahane, P., Rheims, S., Vidal, J. R., & Aru, J. (2018). Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. *Communications Biology*, 1(1), Article 1. <https://doi.org/10.1038/s42003-018-0110-y>
- Kveraga, K., Ghuman, A. S., Kassam, K. S., Aminoff, E. A., Hämäläinen, M. S., Chaumon, M., & Bar, M. (2011). Early onset of neural synchronization in the contextual associations network. *Proceedings of the National Academy of Sciences*, 108(8), 3389–3394. <https://doi.org/10.1073/pnas.1013760108>
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571–579. <http://www.sciencedirect.com/science/article/pii/S016622360001657X>
- Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-Valences: Perceiving Affective Valence in Everyday Objects. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00107>
- Lee, H. S., Schreiner, L., Jo, S.-H., Sieghartsleitner, S., Jordan, M., Pretl, H., Guger, C., & Park, H.-S. (2022). Individual finger movement decoding using a novel ultra-high-density electroencephalography-based brain-computer interface system. *Frontiers in Neuroscience*, 16. <https://www.frontiersin.org/articles/10.3389/fnins.2022.1009878>

- Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, 81(3), 687–699.
<https://doi.org/10.1016/j.neuron.2013.11.028>
- Leeds, D. D., Seibert, D. A., Pyles, J. A., & Tarr, M. J. (2013). Comparing visual representations across human fMRI and computational vision. *Journal of Vision*, 13(13), 25–25.
<http://jov.arvojournals.org/article.aspx?articleid=2193844>
- Leszczyński, M., Barczak, A., Kajikawa, Y., Ulbert, I., Falchier, A. Y., Tal, I., Haegens, S., Melloni, L., Knight, R. T., & Schroeder, C. E. (2020). Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. *Science Advances*, 6(33), eabb0977.
<https://doi.org/10.1126/sciadv.abb0977>
- Lieberman, M. D. (2010). Social Cognitive Neuroscience. In *Handbook of Social Psychology*. American Cancer Society. <https://doi.org/10.1002/9780470561119.socpsy001005>
- Lieberman, M. D. (2013). *Social: Why Our Brains Are Wired to Connect*. Crown.
- Lieberman, M. D. (2022). Seeing minds, matter, and meaning: The CEEing model of pre-reflective subjective construal. *Psychological Review*, 129(4), 830–872.
<https://doi.org/10.1037/rev0000362>
- Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, Self, (Situational), and Affective Processes in Medial Prefrontal Cortex (MPFC): Causal, Multivariate, and Reverse Inference Evidence. *Neuroscience & Biobehavioral Reviews*.
<https://doi.org/10.1016/j.neubiorev.2018.12.021>
- Lin, F.-H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., & Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage*, 31(1), 160–171.
<https://doi.org/10.1016/j.neuroimage.2005.11.054>

- Liu, Q., Ganzetti, M., Wenderoth, N., & Mantini, D. (2018). Detecting Large-Scale Brain Networks Using EEG: Impact of Electrode Density, Head Modeling and Source Localization. *Frontiers in Neuroinformatics*, 12.
<https://www.frontiersin.org/article/10.3389/fninf.2018.00004>
- Lopes da Silva, F. (2013). EEG and MEG: Relevance to Neuroscience. *Neuron*, 80(5), 1112–1128.
<https://doi.org/10.1016/j.neuron.2013.10.017>
- Lopes-dos-Santos, V., van de Ven, G. M., Morley, A., Trouche, S., Campo-Urriza, N., & Dupret, D. (2018). Parsing Hippocampal Theta Oscillations by Nested Spectral Components during Spatial Exploration and Memory-Guided Behavior. *Neuron*, 100(4), 940-952.e7.
<https://doi.org/10.1016/j.neuron.2018.09.031>
- Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, 113(44), 12574–12579.
<https://doi.org/10.1073/pnas.1608282113>
- Marr, D., & Poggio, T. (1976). *From Understanding Computation to Understanding Neural Circuitry*. <https://dspace.mit.edu/handle/1721.1/5782>
- Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the “default mode network” and the “social brain.” *Frontiers in Human Neuroscience*, 6, 189. <https://doi.org/10.3389/fnhum.2012.00189>
- Marzetti, L., Di Lanzo, C., Zappasodi, F., Chella, F., Raffone, A., & Pizzella, V. (2014). Magnetoencephalographic alpha band connectivity reveals differential default mode network interactions during focused attention and open monitoring meditation. *Frontiers in Human Neuroscience*, 8, 832. <https://doi.org/10.3389/fnhum.2014.00832>

Mathworks. (2019). *MATLAB 2019b* [Computer software].

<https://www.mathworks.com/company/newsroom/mathworks-announces-release-r2019b-of-matlab-and-simulink.html>

Maxwell, J. C. (1865). A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, 155, 459–512.

<https://doi.org/10.1098/rstl.1865.0008>

Meltzer, J. A., Negishi, M., Mayes, L. C., & Constable, R. T. (2007). Individual differences in EEG theta and alpha dynamics during working memory correlate with fMRI responses across subjects. *Clinical Neurophysiology*, 118(11), 2419–2436.

<https://doi.org/10.1016/j.clinph.2007.07.023>

Mendoza-Halliday, D., Major, A. J., Lee, N., Lichtenfeld, M. J., Carlson, B., Mitchell, B., Meng, P. D., Xiong, Y. (Sophy), Westerberg, J. A., Jia, X., Johnston, K. D., Selvanayagam, J., Everling, S., Maier, A., Desimone, R., Miller, E. K., & Bastos, A. M. (2024). A ubiquitous spectrolaminar motif of local field potential power across the primate cortex. *Nature Neuroscience*, 1–14. <https://doi.org/10.1038/s41593-023-01554-7>

Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., & Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences*, 109(6), 1883–1888.

<https://doi.org/10.1073/pnas.1121077109>

Meyer, M. L., Taylor, S. E., & Lieberman, M. D. (2015). Social working memory and its distinctive link to social cognitive ability: An fMRI study. *Social Cognitive and Affective Neuroscience*, 10(10), 1338–1347. <https://doi.org/10.1093/scan/nsv065>

Michel, C. M., & Brunet, D. (2019). EEG Source Imaging: A Practical Review of the Analysis Steps. *Frontiers in Neurology*, 10.

<https://www.frontiersin.org/articles/10.3389/fneur.2019.00325>

- Michon, F., Krul, E., Sun, J.-J., & Kloosterman, F. (2021). Single-trial dynamics of hippocampal spatial representations are modulated by reward value. *Current Biology*.
<https://doi.org/10.1016/j.cub.2021.07.058>
- Mishkin, M., Suzuki, W. A., Gadian, D. G., & Vargha-Khadem, F. (1997). Hierarchical organization of cognitive memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1360), 1461–1467. <https://doi.org/10.1098/rstb.1997.0132>
- Mitchell, J. P. (2006). Mentalizing and Marr: An information processing approach to the study of social cognition. *Brain Research, Multiple Perspectives on the Psychological and Neural Bases of Understanding Other People's Behavior*, 1079(1), 66–75.
<https://doi.org/10.1016/j.brainres.2005.12.113>
- Miyakoshi, M., Schmitt, L. M., Erickson, C. A., Sweeney, J. A., & Pedapati, E. V. (2021). Can We Push the “Quasi-Perfect Artifact Rejection” Even Closer to Perfection? *Frontiers in Neuroinformatics*, 14, 597079. <https://doi.org/10.3389/fninf.2020.597079>
- Mizuhara, H., Wang, L.-Q., Kobayashi, K., & Yamaguchi, Y. (2004). A long-range cortical network emerging with theta oscillation in a mental task. *Neuroreport*, 15(8), 1233–1238.
<https://doi.org/10.1097/01.wnr.0000126755.09715.b3>
- Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, 65, 276–291. <https://doi.org/10.1016/j.neubiorev.2016.03.020>
- Morel, P. (2018). Gramm: Grammar of graphics plotting in Matlab. *Journal of Open Source Software*, 3(23), 568. <https://doi.org/10.21105/joss.00568>
- Mormann, F., Kornblith, S., Quiroga, R. Q., Kraskov, A., Cerf, M., Fried, I., & Koch, C. (2008). Latency and Selectivity of Single Neurons Indicate Hierarchical Processing in the Human Medial Temporal Lobe. *Journal of Neuroscience*, 28(36), 8865–8872.
<https://doi.org/10.1523/JNEUROSCI.1640-08.2008>

- Mossad, S. I., AuCoin-Power, M., Urbain, C., Smith, M. L., Pang, E. W., & Taylor, M. J. (2016). Thinking about the thoughts of others; temporal and spatial neural activation during false belief reasoning. *NeuroImage*, 134, 320–327.
<https://doi.org/10.1016/j.neuroimage.2016.03.053>
- Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science*, 309(5736), 951–954. <https://doi.org/10.1126/science.11110913>
- Mullally, S. L., Hassabis, D., & Maguire, E. A. (2012). Scene Construction in Amnesia: An fMRI Study. *Journal of Neuroscience*, 32(16), 5646–5653.
<https://doi.org/10.1523/JNEUROSCI.5522-11.2012>
- Muthukumaraswamy, S. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Frontiers in Human Neuroscience*, 7.
<https://www.frontiersin.org/articles/10.3389/fnhum.2013.00138>
- Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, 77(3), 247–265.
<https://doi.org/10.1016/j.biopsycho.2007.11.006>
- Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage*, 27(2), 341–356.
<https://doi.org/10.1016/j.neuroimage.2005.04.014>
- Onton, J., & Makeig, S. (2009). High-frequency Broadband Modulations of Electroencephalographic Spectra. *Frontiers in Human Neuroscience*, 3.
<https://doi.org/10.3389/neuro.09.061.2009>
- Panichello, M. F., Kveraga, K., Chaumon, M., Bar, M., & Barrett, L. F. (2017). Internal valence modulates the speed of object recognition. *Scientific Reports*, 7(1).
<https://doi.org/10.1038/s41598-017-00385-4>

- Papademetris, X., Jackowski, M. P., Rajeevan, N., DiStasio, M., Okuda, H., Constable, R. T., & Staib, L. H. (2006). BiImage Suite: An integrated medical image analysis suite: An update. *The Insight Journal*, 2006, 209.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4213804/>
- Paquola, C., Garber, M., Frässle, S., Royer, J., Tavakol, S., Rodriguez-Cruces, R., Cabalo, D. G., Valk, S. L., Eickhoff, S., Margulies, D. S., Evans, A. C., Amunts, K., Jefferies, E., Smallwood, J., & Bernhardt, B. (2023). *The Unique Cytoarchitecture and Wiring of the Human Default Mode Network* (p. 2021.11.22.469533). bioRxiv.
<https://doi.org/10.1101/2021.11.22.469533>
- Parkinson, C., & Wheatley, T. (2015). The repurposed social brain. *Trends in Cognitive Sciences*, 19(3), 133–141. <https://doi.org/10.1016/j.tics.2015.01.003>
- Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nature Neuroscience*, 21(4), 474.
<https://doi.org/10.1038/s41593-018-0108-2>
- Pastoll, H., Solanka, L., van Rossum, M. C. W., & Nolan, M. F. (2013). Feedback Inhibition Enables Theta-Nested Gamma Oscillations and Grid Firing Fields. *Neuron*, 77(1), 141–154.
<https://doi.org/10.1016/j.neuron.2012.11.032>
- Pavlidou, A., Schnitzler, A., & Lange, J. (2014). Distinct spatio-temporal profiles of beta-oscillations within visual and sensorimotor areas during action recognition as revealed by MEG. *Cortex*, 54, 106–116. <https://doi.org/10.1016/j.cortex.2014.02.007>
- Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, 200, 460–473. <https://doi.org/10.1016/j.neuroimage.2019.06.046>
- Peng, W. (2019). EEG Preprocessing and Denoising. In L. Hu & Z. Zhang (Eds.), *EEG Signal Processing and Feature Extraction* (pp. 71–87). Springer. https://doi.org/10.1007/978-981-13-9113-2_5

- Perry, A., Stein, L., & Bentin, S. (2011). Motor and attentional mechanisms involved in social interaction—Evidence from mu and alpha EEG suppression. *NeuroImage*, 58(3), 895–904. <https://doi.org/10.1016/j.neuroimage.2011.06.060>
- Petrov, Y., Nador, J., Hughes, C., Tran, S., Yavuzcetin, O., & Sridhar, S. (2014). Ultra-dense EEG sampling results in two-fold increase of functional brain information. *NeuroImage*, 90, 140–145. <https://doi.org/10.1016/j.neuroimage.2013.12.041>
- Peyk, P., De Cesarei, A., & Junghöfer, M. (2011). ElectroMagnetoEncephalography Software: Overview and Integration with Other EEG/MEG Toolboxes. *Computational Intelligence and Neuroscience*, 2011, e861705. <https://doi.org/10.1155/2011/861705>
- Phillips, J. M., Vinck, M., Everling, S., & Womelsdorf, T. (2014). A Long-Range Fronto-Parietal 5- to 10-Hz Network Predicts “Top-Down” Controlled Guidance in a Task-Switch Paradigm. *Cerebral Cortex*, 24(8), 1996–2008. <https://doi.org/10.1093/cercor/bht050>
- Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198, 181–197. <https://doi.org/10.1016/j.neuroimage.2019.05.026>
- Pourtois, G., Vocat, R., N'Diaye, K., Spinelli, L., Seeck, M., & Vuilleumier, P. (2010). Errors recruit both cognitive and emotional monitoring systems: Simultaneous intracranial recordings in the dorsal anterior cingulate gyrus and amygdala combined with fMRI. *Neuropsychologia*, 48(4), 1144–1159. <https://doi.org/10.1016/j.neuropsychologia.2009.12.020>
- Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., Madsen, J. R., & Lisman, J. E. (2001). Gating of Human Theta Oscillations by a Working Memory Task. *Journal of Neuroscience*, 21(9), 3175–3183. <https://doi.org/10.1523/JNEUROSCI.21-09-03175.2001>

- Raimondo, F., Kamienkowski, J. E., Sigman, M., & Slezak, D. F. (2012). CUDAICA: GPU optimization of infomax-ICA EEG analysis. *Computational Intelligence and Neuroscience*, 2012, 2:2. <https://doi.org/10.1155/2012/206972>
- Ray, S., & Maunsell, J. H. R. (2011). Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biology*, 9(4), e1000610. <https://doi.org/10.1371/journal.pbio.1000610>
- Roehri, N., Bréchet, L., Seeber, M., Pascual-Leone, A., & Michel, C. M. (2022). Phase-Amplitude Coupling and Phase Synchronization Between Medial Temporal, Frontal and Posterior Brain Regions Support Episodic Autobiographical Memory Recall. *Brain Topography*, 35(2), 191–206. <https://doi.org/10.1007/s10548-022-00890-4>
- Samogin, J., Liu, Q., Marino, M., Wenderoth, N., & Mantini, D. (2019). Shared and connection-specific intrinsic interactions in the default mode network. *NeuroImage*, 200, 474–481. <https://doi.org/10.1016/j.neuroimage.2019.07.007>
- Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7(5), Article 5. <https://doi.org/10.1038/nn1223>
- Samuelsson, J. G., Khan, S., Sundaram, P., Peled, N., & Hämäläinen, M. S. (2019). Cortical Signal Suppression (CSS) for Detection of Subcortical Activity Using MEG and EEG. *Brain Topography*, 32(2), 215–228. <https://doi.org/10.1007/s10548-018-00694-5>
- Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6), e13335. <https://doi.org/10.1111/psyp.13335>
- Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research, Multiple Perspectives on the*

- Psychological and Neural Bases of Understanding Other People's Behavior*, 1079(1), 86–97. <https://doi.org/10.1016/j.brainres.2006.01.005>
- Saxe, R. (2010). The right temporo-parietal junction: A specific brain region for thinking about thoughts. *Handbook of Theory of Mind*, 1–35. <http://dosen.narotama.ac.id/wp-content/uploads/2013/09/The-right-temporo-parietal-junction-a-specific-brain-region-for-thinking-about-thoughts.pdf>
- Saxe, R., & Powell, L. J. (2006). It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind. *Psychological Science*, 17(8), 692–699. <https://doi.org/10.1111/j.1467-9280.2006.01768.x>
- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, 19(2), 65–72. <https://doi.org/10.1016/j.tics.2014.11.007>
- Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Alaguela, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9), 1141–1152. <https://doi.org/10.1111/2041-210X.13434>
- Schoffelen, J.-M., & Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Human Brain Mapping*, 30(6), 1857–1865. <https://doi.org/10.1002/hbm.20745>
- Schreiner, L., Sieghartsleitner, S., Mayr, K., Pretl, H., & Guger, C. (2023). Hand gesture decoding using ultra-high-density EEG. *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, 01–04. <https://doi.org/10.1109/NER52421.2023.10123901>
- Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, 147(3), 293. <https://doi.org/10.1037/bul0000303>

- Seeber, M., Cantonas, L.-M., Hoevels, M., Sesia, T., Visser-Vandewalle, V., & Michel, C. M. (2019). Subcortical electrophysiological activity is detectable with high-density EEG source imaging. *Nature Communications*, 10(1), Article 1. <https://doi.org/10.1038/s41467-019-08725-w>
- Seymour, R. A., Wang, H., Rippon, G., & Kessler, K. (2018). Oscillatory networks of high-level mental alignment: A perspective-taking MEG study. *NeuroImage*, 177, 98–107. <https://doi.org/10.1016/j.neuroimage.2018.05.016>
- Sheoran, M., Kumar, S., & Chawla, S. (2015). Methods of denoising of electroencephalogram signal: A review. *International Journal of Biomedical Engineering and Technology*, 18(4), 385–395. <https://doi.org/10.1504/IJBET.2015.071012>
- Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, 13(2), 121–134. <https://doi.org/10.1038/nrn3137>
- Song, J., Davey, C., Poulsen, C., Luu, P., Turovets, S., Anderson, E., Li, K., & Tucker, D. (2015). EEG source localization: Sensor density and head surface coverage. *Journal of Neuroscience Methods*, 256, 9–21. <https://doi.org/10.1016/j.jneumeth.2015.08.015>
- Soni, S., Overton, J., Kam, J. W. Y., Pexman, P., Prabhu, A., Garza, N., Saez, I., & Girgis, F. (2024). Intracranial recordings reveal high-frequency activity in the human temporal-parietal cortex supporting non-literal language processing. *Frontiers in Neuroscience*, 17. <https://doi.org/10.3389/fnins.2023.1304031>
- Spitzer, B., & Haegens, S. (2017). Beyond the Status Quo: A Role for Beta Oscillations in Endogenous Content (Re)Activation. *eNeuro*, 4(4). <https://doi.org/10.1523/ENEURO.0170-17.2017>
- Spreng, R. N., & Grady, C. L. (2010). Patterns of Brain Activity Supporting Autobiographical Memory, Prospection, and Theory of Mind, and Their Relationship to the Default Mode

- Network. *Journal of Cognitive Neuroscience*, 22(6), 1112–1123.
<https://doi.org/10.1162/jocn.2009.21282>
- Sprengh, R. N., Mar, R. A., & Kim, A. S. N. (2008). The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience*, 21(3), 489–510.
<https://doi.org/10.1162/jocn.2008.21029>
- Spunt, R. P., & Adolphs, R. (2014). Validating the Why/How contrast for functional MRI studies of Theory of Mind. *NeuroImage*, 99, 301–311.
<https://doi.org/10.1016/j.neuroimage.2014.05.023>
- Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: Insights from social neuroscience. *Nature Reviews Neuroscience*, 18(9), 559–567.
<https://doi.org/10.1038/nrn.2017.76>
- Spunt, R. P., & Lieberman, M. D. (2012). Dissociating Modality-Specific and Supramodal Neural Systems for Action Understanding. *Journal of Neuroscience*, 32(10), 3575–3583.
<https://doi.org/10.1523/JNEUROSCI.5715-11.2012>
- Stone, J. L., & Hughes, J. R. (2013). Early History of Electroencephalography and Establishment of the American Clinical Neurophysiology Society. *Journal of Clinical Neurophysiology*, 30(1), 28–44. <https://doi.org/10.1097/WNP.0b013e31827edb2d>
- Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3(4), 151–162.
[https://doi.org/10.1016/S1364-6613\(99\)01299-1](https://doi.org/10.1016/S1364-6613(99)01299-1)
- Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: The role of simulation in the default network. *Social Cognitive and Affective Neuroscience*, 11(2), 215–224. <https://doi.org/10.1093/scan/nsv114>

- Tan, K. M., Burklund, L. J., Craske, M. G., & Lieberman, M. D. (2019). Posttraumatic stress disorder and the social brain: Affect-related disruption of the default and mirror networks. *Depression and Anxiety*, 36(11), 1058–1071.
<https://doi.org/10.1002/da.22953>
- Tan, K. M., Daitch, A. L., Pinheiro-Chagas, P., Fox, K. C. R., Parvizi, J., & Lieberman, M. D. (2022). Electrocorticographic evidence of a common neurocognitive sequence for mentalizing about the self and others. *Nature Communications*, 13(1), Article 1.
<https://doi.org/10.1038/s41467-022-29510-2>
- Tan, K. M., & Tarr, M. (2016, April 1). *Valence-related visual encoding is dependent on valence strength prior to frontal affective processing*. Social and Affective Neuroscience Society.
<https://doi.org/10.13140/RG.2.2.31730.91840>
- Thies, M., Zrenner, C., Ziemann, U., & Bergmann, T. O. (2018). Sensorimotor mu-alpha power is positively related to corticospinal excitability. *Brain Stimulation*, 11(5), 1119–1122.
<https://doi.org/10.1016/j.brs.2018.06.006>
- Thio, B. J., & Grill, W. M. (2023). Relative contributions of different neural sources to the EEG. *NeuroImage*, 275, 120179. <https://doi.org/10.1016/j.neuroimage.2023.120179>
- Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019). The Social Brain Automatically Predicts Others' Future Mental States. *Journal of Neuroscience*, 39(1), 140–148.
<https://doi.org/10.1523/JNEUROSCI.1431-18.2018>
- Tierney, T. M., Levy, A., Barry, D. N., Meyer, S. S., Shigihara, Y., Everatt, M., Mellor, S., Lopez, J. D., Bestmann, S., Holmes, N., Roberts, G., Hill, R. M., Boto, E., Leggett, J., Shah, V., Brookes, M. J., Bowtell, R., Maguire, E. A., & Barnes, G. R. (2021). Mouth magnetoencephalography: A unique perspective on the human hippocampus. *Neuroimage*, 225, 117443. <https://doi.org/10.1016/j.neuroimage.2020.117443>

- Tognoli, E., & Kelso, J. A. S. (2015). The coordination dynamics of social neuromarkers. *Frontiers in Human Neuroscience*, 9.
<https://www.frontiersin.org/articles/10.3389/fnhum.2015.00563>
- Tucciarelli, R., Turella, L., Oosterhof, N. N., Weisz, N., & Lingnau, A. (2015). MEG Multivariate Analysis Reveals Early Abstract Action Representations in the Lateral Occipitotemporal Cortex. *Journal of Neuroscience*, 35(49), 16034–16045.
<https://doi.org/10.1523/JNEUROSCI.1422-15.2015>
- Tuyisenge, V., Trebaul, L., Bhattacharjee, M., Chanteloup-Forêt, B., Saubat-Guigui, C., Mîndruță, I., Rheims, S., Maillard, L., Kahane, P., Taussig, D., & David, O. (2018). Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning. *Clinical Neurophysiology*, 129(3), 548–554.
<https://doi.org/10.1016/j.clinph.2017.12.013>
- Uhlhaas, P. J., Pipa, G., Neuenschwander, S., Wibral, M., & Singer, W. (2011). A new look at gamma? High- (>60 Hz) γ -band activity in cortical networks: Function, mechanisms and impairment. *Progress in Biophysics and Molecular Biology, BrainModes: The Role of Neuronal Oscillations in Health and Disease*, 105(1), 14–28.
<https://doi.org/10.1016/j.pbiomolbio.2010.10.004>
- van Drongelen, W., Yuchtman, M., Van Veen, B. D., & van Huffelen, A. C. (1996). A spatial filtering technique to detect and localize multiple sources in the brain. *Brain Topography*, 9(1), 39–49. <https://doi.org/10.1007/BF01191641>
- Van Duynslaeger, M., Van Overwalle, F., & Verstraeten, E. (2007). Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience*, 2(3), 174–188. <https://doi.org/10.1093/scan/nsm016>

- Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6, 370–375. [https://doi.org/10.1016/0166-2236\(83\)90167-4](https://doi.org/10.1016/0166-2236(83)90167-4)
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, 48(3), 564–584. <https://doi.org/10.1016/j.neuroimage.2009.06.009>
- Vistoli, D., Brunet-Gouet, E., Baup-Bobin, E., Hardy-Bayle, M.-C., & Passerieux, C. (2011). Anatomical and temporal architecture of theory of mind: A MEG insight into the early stages. *NeuroImage*, 54(2), 1406–1414. <https://doi.org/10.1016/j.neuroimage.2010.09.015>
- von Seth, J., Nicholls, V. I., Tyler, L. K., & Clarke, A. (2023). Recurrent connectivity supports higher-level visual and semantic object representations in the brain. *Communications Biology*, 6(1), Article 1. <https://doi.org/10.1038/s42003-023-05565-9>
- Wang, C., Ulbert, I., Schomer, D. L., Marinkovic, K., & Halgren, E. (2005). Responses of Human Anterior Cingulate Cortex Microdomains to Error Detection, Conflict Monitoring, Stimulus-Response Mapping, Familiarity, and Orienting. *Journal of Neuroscience*, 25(3), 604–613. <https://doi.org/10.1523/JNEUROSCI.4151-04.2005>
- Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., & Kessler, K. (2016). Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex*, 75, 68–81. <https://doi.org/10.1016/j.cortex.2015.11.011>
- Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7(12), 553–559. <https://doi.org/10.1016/j.tics.2003.10.012>
- Weaverdyck, M. E., Lieberman, M. D., & Parkinson, C. (2020). Tools of the Trade Multivoxel pattern analysis in fMRI: A practical introduction for social and affective neuroscientists.

- Social Cognitive and Affective Neuroscience*, 15(4), 487–509.
<https://doi.org/10.1093/scan/nsaa057>
- Wendel, K., Väisänen, O., Malmivuo, J., Gencer, N. G., Vanrumste, B., Durka, P., Magjarević, R., Supek, S., Pascu, M. L., Fontenelle, H., & Grave de Peralta Menendez, R. (2009). EEG/MEG Source Imaging: Methods, Challenges, and Open Issues. *Computational Intelligence and Neuroscience*, 2009, e656092. <https://doi.org/10.1155/2009/656092>
- Wendiggensen, P., Prochnow, A., Pscherer, C., Münchau, A., Frings, C., & Beste, C. (2023). Interplay between alpha and theta band activity enables management of perception-action representations for goal-directed behavior. *Communications Biology*, 6(1), Article 1. <https://doi.org/10.1038/s42003-023-04878-z>
- Wessel, J. R., & Anderson, M. C. (2024). Neural mechanisms of domain-general inhibitory control. *Trends in Cognitive Sciences*, 28(2), 124–143.
<https://doi.org/10.1016/j.tics.2023.09.008>
- Westner, B. U., Dalal, S. S., Gramfort, A., Litvak, V., Mosher, J. C., Oostenveld, R., & Schoffelen, J.-M. (2022). A unified view on beamformers for M/EEG source reconstruction. *NeuroImage*, 246, 118789. <https://doi.org/10.1016/j.neuroimage.2021.118789>
- Whittington, M. A., Traub, R. D., & Jefferys, J. G. (1995). Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature*, 373(6515), 612–615. <https://doi.org/10.1038/373612a0>
- Winkler, I., Debener, S., Muller, K.-R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 4101–4105.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7319296
- Wischnewski, M., Haigh, Z. J., Shirinpour, S., Alekseichuk, I., & Opitz, A. (2022). The phase of sensorimotor mu and beta oscillations has the opposite effect on corticospinal

- excitability. *Brain Stimulation*, 15(5), 1093–1100.
<https://doi.org/10.1016/j.brs.2022.08.005>
- Xiong, Y., Fries, P., & Bastos, A. M. (2023). Which Rhythms Reflect Bottom-Up and Top-Down Processing? In N. Axmacher (Ed.), *Intracranial EEG: A Guide for Cognitive Neuroscientists* (pp. 389–414). Springer International Publishing. https://doi.org/10.1007/978-3-031-20910-9_25
- Yang, Y., Tarr, M. J., Kass, R. E., & Aminoff, E. M. (2019). Exploring spatiotemporal neural dynamics of the human visual cortex. *Human Brain Mapping*, 40(14), 4213–4238.
<https://doi.org/10.1002/hbm.24697>
- Yekutieli, D., & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82(1), 171–196. [https://doi.org/10.1016/S0378-3758\(99\)00041-5](https://doi.org/10.1016/S0378-3758(99)00041-5)
- Yuk, V., Urbain, C., Pang, E. W., Anagnostou, E., Buchsbaum, D., & Taylor, M. J. (2018). Do you know what I'm thinking? Temporal and spatial brain activity during a theory-of-mind task in children with autism. *Developmental Cognitive Neuroscience*, 34, 139–147.
<https://doi.org/10.1016/j.dcn.2018.08.001>
- Zhang, J., Li, J., Huang, Z., Huang, D., Yu, H., & Li, Z. (2023). Recent Progress in Wearable Brain–Computer Interface (BCI) Devices Based on Electroencephalogram (EEG) for Medical Applications: A Review. *Health Data Science*, 3, 0096. <https://doi.org/10.34133/hds.0096>
- Zhang, L., Gan, J. Q., Zheng, W., & Wang, H. (2018). Spatiotemporal Phase Synchronization in Adaptive Reconfiguration from Action Observation Network to Mentalizing Network for Understanding Other's Action Intention. *Brain Topography*, 31(3), 447–467.
<https://doi.org/10.1007/s10548-017-0614-7>
- Zhang, L., Lee, J., Rozell, C., & Singer, A. C. (2019). Sub-second dynamics of theta-gamma coupling in hippocampal CA1. *eLife*, 8, e44320. <https://doi.org/10.7554/eLife.44320>

Zorzos, I., Kakkos, I., Ventouras, E. M., & Matsopoulos, G. K. (2021). Advances in Electrical Source Imaging: A Review of the Current Approaches, Applications and Challenges. *Signals*, 2(3), Article 3. <https://doi.org/10.3390/signals2030024>