

I am pleased to submit a portfolio of work samples, which showcases my skills in science communication, programming/software development, and empirical research.

### **1. Science Communication**

This section includes my published *Nature Neuroscience* commentary, which outlines a theoretical framework for science communication. It also provides practical resources for delivering both didactic and hands-on training in science communication. *Link:* <https://www.nature.com/articles/s41593-024-01646-y.pdf>

### **2. Programming/Software Development**

In this section, I highlight two projects that demonstrate my skills in Python programming and software development:

#### **A. NeuroCluster**

A Python package for linking neural signals with latent cognitive variables extracted from computational models. This project required the implementation of a novel statistical approach while exploring strategies to optimize computing time and resource efficiency. Manuscript is under review at *Journal of Open Source Software*. *GitHub Repository:* <https://github.com/aliefink/NeuroCluster>

#### **B. Python Tutorial for Computational Modeling of Behavior**

Adapted from Wilson & Collins (2019) MATLAB-based tutorial, this resource has been widely used for training at Mount Sinai's Center for Computational Psychiatry. *GitHub Repository:* <https://github.com/christinamaher/10SimpleRulesPythonTutorial>

#### **C. Experimental Paradigm Design and Gamification**

Tutorial on gamifying behavioral paradigms to enhance universal design and optimize user experience, improving both participant retention and data quality. This tutorial has been implemented by researchers at Icahn School of Medicine and University College London, driving collaborative innovation in research applications. *GitHub Repository:* [https://github.com/christinamaher/Canva\\_VideoInstructions\\_Tutorial](https://github.com/christinamaher/Canva_VideoInstructions_Tutorial)

### **3. Empirical Research**

This section highlights three first-author papers:

#### **A. Maher et al. (2024)**

Selected for a talk at the 2024 Conference on Computational Cognitive Neuroscience among 500+ submissions, this study leverages behavioral modeling and direct brain recordings to investigate real-world adaptive decision-making. A manuscript is in preparation. *Publication link:* [https://2024.ccneuro.org/pdf/548\\_Paper\\_authored\\_CCN2024\\_Maheretal.pdf](https://2024.ccneuro.org/pdf/548_Paper_authored_CCN2024_Maheretal.pdf)

#### **B. Maher et al. (2024)**

Published in *PNAS*, this study investigates the relationship between limbic neuromodulation and patient well-being. I coordinated a diverse team of multidisciplinary stakeholders, aligning efforts to leverage scientific inquiry to advance patient wellbeing. *Link:* <https://www.pnas.org/doi/10.1073/pnas.2409423122>

#### **C. Maher et al. (2025)**

Accepted for presentation at RLDM, this work leverages a novel neural network approach for behavioral inference with human behavioral modeling. *Link:* [https://drive.google.com/file/d/1DnHZnwAeAEEtD\\_Ida9FkvuuMhUJw1fz/view?usp=sharing](https://drive.google.com/file/d/1DnHZnwAeAEEtD_Ida9FkvuuMhUJw1fz/view?usp=sharing)

Thank you for considering my portfolio. I look forward to discussing it further during the review process.



Icahn  
School of  
Medicine at  
**Mount  
Sinai**

**Christina Maher**  
*Neuroscience PhD Candidate*  
*Friedman Brain Institute*  
*New York City, NY*

# **Science Communication**

# A guide to science communication training for doctoral students

Christina Maher, Trevonn Gyles, Eric J. Nestler & Daniela Schiller



Effective science communication is necessary for engaging the public in scientific discourse and ensuring equitable access to knowledge. Training doctoral students in science communication will instill principles of accessibility, accountability, and adaptability in the next generation of scientific leaders, who are poised to expand science’s reach, generate public support for research funding, and counter misinformation. To this aim, we provide a guide for implementing formal science communication training for doctoral students.

*“We have not known a single great scientist who could not discourse freely and interestingly with a child.”—John Steinbeck<sup>1</sup>*

Many of us remember our first experience with science, or the first time we heard about a scientific finding that ignited our curiosity. These early experiences draw us in, capture our imagination, and provide us an entry point we may otherwise never know exists. We can attribute these moments that sparked our intellectual curiosity to the science communicators who took the initiative and the accountability to share science with a wider audience. Here, we discuss why and how to set up training for doctoral students to enable them to become the next generation of science communicators, and thereby to expand the future scientific workforce, broaden the public reach of our science, and increase public support for funding for scientific research.

According to UNESCO’s 2021 Science Report, there are approximately nine million full-time researchers world-wide<sup>2</sup>. However, with a global population of approximately eight billion, it is imperative that science is accessible to far more than just about 0.1% of the population, creating a global need for communicators. The commodity that researchers and educators exchange is information, and the quality of this exchange depends on our ability to communicate effectively.

We live in an interconnected world with tremendous access to information. This brings both opportunity and risk, as scientific (or pseudo-scientific) information can easily spread through various media platforms. It is imperative that we train scientists to be accountable, accessible, and adaptable, so that they may effectively engage with the public through varied platforms, promote scientific literacy, and mitigate the risks of misinformation<sup>3,4</sup>. Training researchers to communicate accessibly and accurately is instrumental for building trust with a wider audience<sup>5,6</sup>. Indeed, trust is the foundation for greater

## Three As of effective science communicators

**Accessible**  
Convey scientific concepts clearly to a wide range of audiences, regardless of age and background

**Accountable**  
Foster trust through transparent communication and responsible dissemination of information

**Adaptable**  
Use diverse communication tools to reach audiences across media and contexts

## Key modules

- Science education
- Science writing
- Science podcasting

- Data visualization
- Public outreach
- Science policy

- Social media
- Improvisation
- On-camera presentation

**Fig. 1 | The three As of effective science communicators.** Examples of the modules used to instill each are given.

public awareness, interest, and policy adherence to science-based issues, such as climate change<sup>7</sup> and mental health<sup>8</sup>.

If we instill good communication practices in doctoral students, they will be better equipped to inform and influence policy makers and the public. In this way, scientists can foster a rich and vibrant discourse that extends our work’s reach and helps to strengthen our society. This is particularly important because many students trained as basic scientists will pursue non-academic careers. According to the US National Science Foundation’s Survey of Doctorate Recipients, only 23% of life and health sciences PhD graduates held tenured or tenure-track positions in academia in 2017. Training in effective science communication prepares students for a wide range of careers, including education, the biotechnology and pharmaceutical industries, government, and publishing.

We believe that the essential attributes of effective science communicators can be summarized by the ‘three As’: accessible, accountable, and adaptable. Accessible communication practices entail conveying scientific concepts to a wide range of audiences regardless of age and background<sup>9</sup>. Accountability requires gaining the trust of the target audience through transparent communication and responsible dissemination of information. Accessible and accountable communication practices construct a bridge of trust between communicators and their audience, facilitating a reciprocal exchange of knowledge<sup>10</sup>. These practices are crucial for sharing knowledge beyond institutional and geographic borders<sup>11</sup>. Proficiency in this process allows communicators to be adaptable to variety of media and contexts for disseminating science<sup>12</sup>. These attributes build on each another to ultimately enhance the effectiveness of science communication efforts (Fig. 1).

Currently, science communication is not a universal component of the curriculum in neuroscience doctoral programs. Our goal in writing this piece is to explain why PhD-level science communication courses are needed, and to share practical information about their implementation. We aim to encourage funding agencies to promote the development of this training, and to allocate resources to support it. The example syllabus we provide (Table 1) is designed to promote accessible, accountable, and adaptable science communication through

**Table 1 | Effective science communication course breakdown**

Module topic	Objective	Project	Evaluation
<b>Written communication modules</b>			
<b>Science on social media</b>	Learn to develop professional identity and create an online presence for scientific discourse	Creation of a personal website or a social media account for the purpose of conveying scientific knowledge	Feedback on ways to improve their social media presence to appeal to their professional interest
<b>Creating a science curriculum</b>	Learn inclusive teaching techniques for middle school science education	Design of a lesson plan, application of teaching techniques learned during workshop	Quality and relevance of lesson plan, effectiveness of teaching techniques applied
<b>Illustration and visualization</b>	Learn techniques for visualizing data and scientific projects	Creation of graphical abstracts of student's dissertation project	Feedback on figures to improve schematic quality
<b>Science writing for non-science audiences</b>	Learn principles of popular science writing and journalistic reporting	Creation of a science blog post, op-ed, or commentary geared toward non-scientists	Peer review of written pieces and implementation of feedback to improve reader accessibility
<b>Oral communication modules</b>			
<b>Improvisation and on-stage performance</b>	Develop skills in science communication through improvisation, role play, and storytelling	One time participation in on-stage performance of improvisation, storytelling or similar	Improved ability to think creatively about science communication, increased confidence in public speaking
<b>Public engagement in neuroscience</b>	Learn skills for engaging with the public, brainstorm ideas for public engagement initiatives	Development of a strategic plan for a public engagement initiative	Effectiveness and relevance of public engagement initiative
<b>Science podcasting</b>	Learn interview techniques, practice interviewing skills	Create one podcast episode and/or participate in podcast interviewing	Understanding of key elements of science podcasting, quality of interview skills demonstrated
<b>Science policy and advocacy</b>	Learn how to approach and pitch to local government representatives	Plan or participate in a meeting with your local representative to advocate for science	Efficacy of the communication strategy and feedback from meeting with local representative

An exemplar syllabus based on the Effective Science Communication course held at the Icahn School of Medicine at Mount Sinai. This course structure could be adapted to meet the needs of any institution and its students.

didactic lectures and hands-on activities, designed to engage students in best practices for effective science communication.

We have specifically created the 'Effective Science Communication' course at the Icahn School of Medicine at Mount Sinai, which was initiated by students in 2018, and has since been completed by more than 100 students. This is a semester-long course delivered annually to approximately 20 graduate students under the auspices of the Friedman Brain Institute at Icahn Mount Sinai. We strive for this material to be delivered in a classroom setting that is respectful of each student's comfort and experience with science communication. The class is intended to be a safe space for students to practice new skills while receiving supportive feedback from experts across various domains.

As an exemplar course structure (Table 1), there are two sets of modules, representing the two forms of communication: written and oral. In each set, four modules focus on different media for communicating one's science. Each module consists of two parts – theoretical work and a hand-on workshop, which could be delivered within a 3-h session. We emphasize hosting speakers who reflect diverse personal and professional backgrounds and identities, and who can best connect with, inspire, and support the students in their development as effective science communicators. Previous lecturers include journal editors, podcasters, journalists, lobbyists, press officers, industry leaders, TV presenters, improvisation and storytelling coaches, and social media experts. Speakers are typically scientists who pursue their preferred mode of science communication either full-time or to supplement their academic research. This allows students to explore alternative career paths or be inspired in ways to productively integrate science communication in their research careers.

Each speaker provides pre-class materials, an in-class hands-on workshop, and an optional post-class assignment. Students can choose one of these assignments as their final class project, which they will develop into a complete product, such as a blog post, podcast episode, science website, elementary school class, or storytelling performance. We expect that by completing this course, students will be able to identify the needs of their diverse audiences, effectively and flexibly convey their research's significance using different media, and maintain the integrity of scientific principles. These skills create a holistic scientist, poised to serve as a changemaker on a global scale.

This course structure could be adapted to meet the needs of any institution and its students. For example, different weights could be put on oral or written communication, on education, or on public interfacing projects. In fact, each of these topics could constitute a course in itself. The emphasis on lectures versus hand-on workshops can also be adjusted according to institutional needs and capabilities.

We hope that these resources will decrease the barrier to entry for institutions seeking to implement such educational programs for trainees. We also hope that students will take this practical framework as a guide for academic development. Students provide feedback before and after the course as well as after each module to track the course's efficacy and to ensure that material is tailored to the students' unique needs and interests. Our goal is to inspire scientists across disciplines to uphold the three As of science communication – accessibility, accountability, and adaptability – when pursuing opportunities to communicate their research and to effectively embed science in everyday life.



Icahn  
School of  
Medicine at  
**Mount  
Sinai**

**Christina Maher**  
*Neuroscience PhD Candidate*  
*Friedman Brain Institute*  
*New York City, NY*

# **Programming/ Software Development**

1 NeuroCluster: A Python toolbox for nonparametric  
2 cluster-based statistical testing of neurophysiological  
3 data with respect to continuous predictors.

4 Alexandra Fink Skular <sup>1,2,3\*</sup>, Christina Maher <sup>1,2,3\*</sup>, Salman Qasim <sup>2,3</sup>,  
5 and Ignacio Saez <sup>1,2,3,4,5</sup> 

6 **1** The Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, NY, NY **2**  
7 The Nash Family Center for Advanced Circuit Therapeutics, The Mount Sinai Hospital, NY, NY **3** The  
8 Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai, NY, NY **4** Department  
9 of Neurosurgery, The Mount Sinai Hospital, NY, NY **5** Department of Neurology, The Mount Sinai  
10 Hospital, NY, NY  Corresponding author \* These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright  
and release the work under a  
Creative Commons Attribution 4.0  
International License ([CC BY 4.0](#)).

11 Summary

12 Cognitive neurophysiology offers a unique framework for studying cognitive brain-behavior  
13 relationships by relating electrophysiological signals to complex behaviors. With the advent of  
14 new technical and behavioral paradigms, researchers can design cognitive experiments that  
15 leverage both the spatiotemporal resolution of electrophysiological data and the complexity  
16 of continuous behavioral variables. Analyzing these data requires sophisticated statistical  
17 methods that can interpret multidimensional neurophysiological data and dynamic, continuous  
18 behavioral variables. Often used statistical frameworks for nonparametric, cluster-based  
19 statistical tests are specifically focused on the contrast between discrete behavioral conditions  
20 but are not suitable for assessing how continuous variables predict the occurrence of clusters  
21 in neurophysiological data. NeuroCluster is an open-source Python toolbox for analysis  
22 of two-dimensional electrophysiological data (e.g. time-frequency representations) related to  
23 multivariate and continuous behavioral variables. NeuroCluster introduces a statistical approach  
24 which uses nonparametric cluster-based permutation testing in tandem with linear regression  
25 to identify two-dimensional clusters of neurophysiological activity that significantly encodes  
26 time-varying, continuous behavioral variables. Uniquely, it also supports multivariate analyses  
27 by allowing for multiple behavioral predictors to model neural activity. NeuroCluster addresses  
28 a methodological gap in statistical approaches to relate continuous, cognitive predictors to  
29 underlying electrophysiological activity with time and frequency resolution, to determine the  
30 neurocomputational processes giving rise to complex behaviors. # Statement of need

31 NeuroCluster addresses a methodological gap in cognitive and behavioral neuroscience, by  
32 providing a Python-based statistical toolbox to relate continuous predictors to two-dimensional  
33 neurophysiological activity. Continuous predictors vary over an experimental session, reflecting  
34 dynamic behaviors, underlying cognitive processes, complex movements, trial-varying experi-  
35 mental conditions, perceptual signals, or value-based trial outcomes ([Collins & Shenhav, 2022](#);  
36 [Hoy et al., 2021](#); [Mathis & Mathis, 2020](#); ?). Standard analytical approaches for relating  
37 complex behavioral variables to neuronal activity sacrifice the complexity of neurophysiological  
38 signals by reducing the dimensionality of neuronal timeseries data (e.g., averaging across  
39 temporal, spectral, or spatial domains, or dimensionality reduction) ([Crosse et al., 2016](#); [Rey  
40 et al., 2015](#); [Saez et al., 2018](#); [Stokes & Spaak, 2016](#); ?; ?). Conversely, analysis methods  
41 that preserve the complexity of neurophysiological data (i.e., two-dimensional timeseries)  
42 constrain behavioral predictors to discrete conditions ([Domenech et al., 2020](#); [Kosciessa et al.,  
43 2020](#); [Maris & Oostenveld, 2007](#); ?; ?). Directly linking continuous experimental variables to

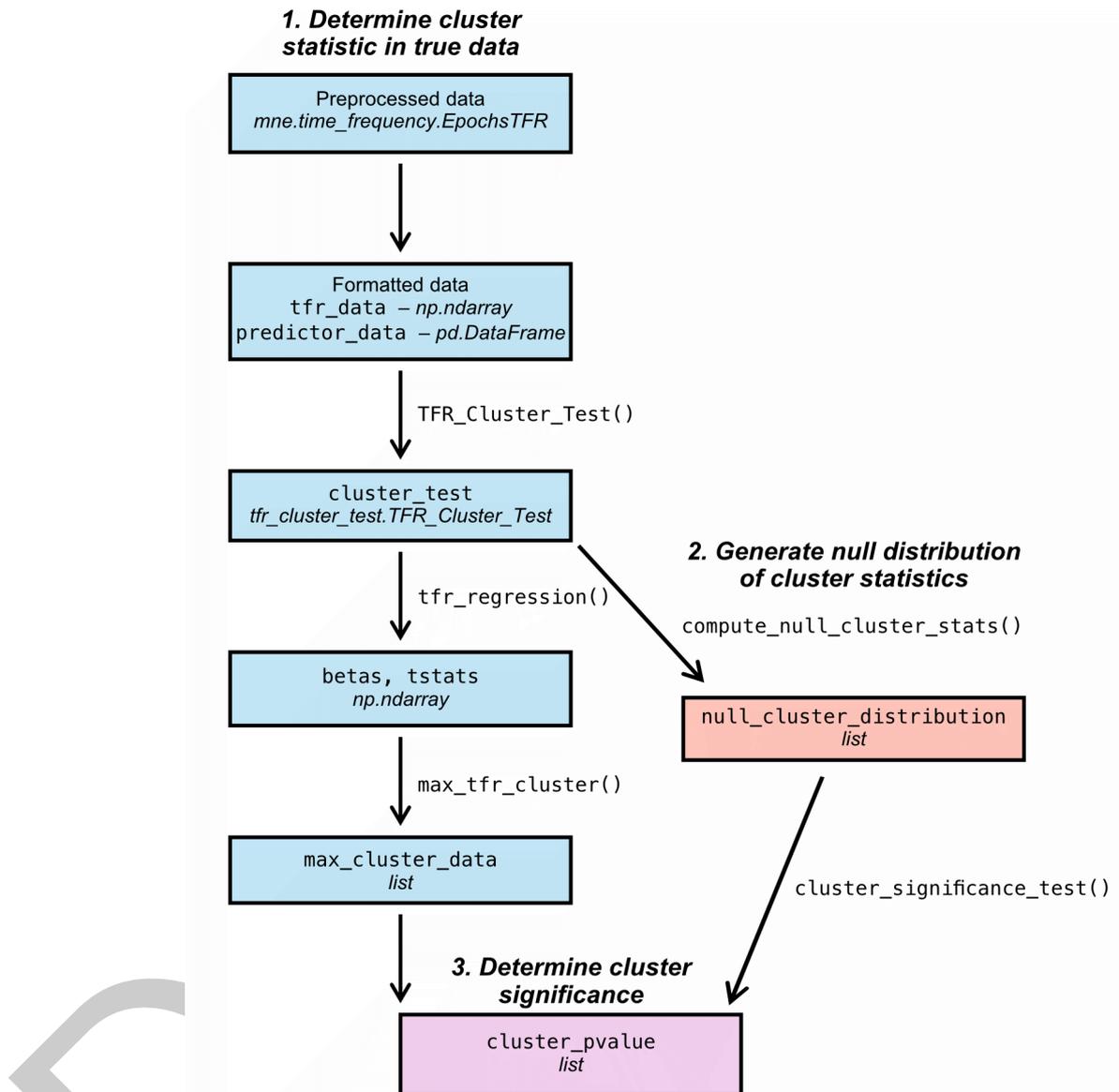
two-dimensional physiological timeseries data offers a rigorous way to study brain-behavior relationships, by maintaining the complexity of dynamic behavior, without sacrificing the resolution of event-related neurophysiological activity.

NeuroCluster uses cluster-based permutation testing to identify significant two-dimensional clusters with respect to continuous task variables. Cluster-based nonparametric statistical testing is a standard approach to analyze two-dimensional event-related time series data, while controlling for multiple comparisons and reducing family-wise error rates (Cohen, 2014; Groppe et al., 2011; Maris, 2012; Maris & Oostenveld, 2007; Nichols & Holmes, 2002). Neurophysiological activity is typically aggregated by condition to perform a two-sample cluster-based permutation test, which tests whether the neuronal encoding patterns differ between two discrete task conditions, rather than continuous, trial-varying features (Bullmore et al., 1999; Maris & Oostenveld, 2007). While two-sample cluster-based permutation tests provide a nonparametric statistical inference tool for identifying the presence of significant clusters of activity between two conditions, they are insufficient for identifying the presence of clusters as a function of continuously varying predictors. NeuroCluster provides a solution to this analytical gap by performing linear regressions at individual points across the 2D neural matrix. This approach enables users to quantify the degree to which a continuous predictor is related to neurophysiological activity at the pixel-level and to perform analyses with multivariate behavioral data, by incorporating multiple continuous or categorical covariates in the regression models. The t-statistics corresponding to the predictor of interest from the pixel-wise regressions are thresholded by a critical t-statistic to control for the FDR, creating a binary 2D matrix (Genovese et al., 2002). The binary 2D matrix is then used to identify putative 2D clusters of activation related to the continuous predictor of interest. This process is repeated many times with the predictor of interest randomly permuted to produce a surrogate distribution of 2D clusters. Clusters that survive cluster-based permutation testing are classified as significant regions of activation with respect to the specified continuous predictor.

NeuroCluster is applicable for numerous analysis goals; the major use cases are performing an initial exploratory analysis to generate specific hypotheses, determine data-driven windows of interest, or to identify regional patterns of significant clusters within and between subjects. Future adaptations of NeuroCluster may implement mixed effects regressions, nonlinear mapping models, or group-level analysis frameworks (Bianchi et al., 2019; Ivanova et al., 2022; König et al., 2024; Yu et al., 2022). We demonstrate our approach with human intracranial local field potential data, but NeuroCluster is applicable for all types of two-dimensional neurophysiological measures (e.g., spatiotemporal clusters from EEG/MEG, cross-frequency interactions). To our knowledge, NeuroCluster presents a novel Python-based statistical software package. NeuroCluster is designed to supplement existing Python-based electrophysiological analysis toolboxes (Donoghue et al., 2020; Gramfort, 2013; Koscieska et al., 2020; Whitten et al., 2011), particularly MNE-Python.

## NeuroCluster Documentation

NeuroCluster is accompanied by a detailed tutorial which outlines the workflow (Fig. 1) for implementing this approach with time-frequency power estimates from multi-region LFP recording.



**Figure 1:** NeuroCluster workflow. This approach involves three key steps: (1) determine cluster statistic in true data, (2) generate a null distribution of cluster statistics by permuting dataset, (3) determine significance of true cluster statistic against null distribution.

86 Below we outline the statistical approach implemented by NeuroCluster for performing  
 87 nonparametric permutation-based cluster testing using time-frequency resolved power estimates  
 88 from neural data estimated using (?) and continuous predictors (i.e., latent cognitive processes,  
 89 behavior, or experimental conditions). In these example data, we are testing the hypothesis  
 90 that RPEs are significantly encoded in the electrophysiological signal from a given iEEG channel  
 91 time-frequency representation (TFR). The following methodological description is based on  
 92 data collected from a neurosurgical epilepsy patient undergoing stereotactic EEG (sEEG)  
 93 monitoring for treatment-resistant depression. During the monitoring period, the patient  
 94 performed a value-based decision-making task while local field potentials (LFPs) were recorded

95 from both cortical and subcortical sites. By analyzing the patient's behavior during the task,  
96 we derived continuous variables representing hypothesized latent cognitive processes—such as  
97 the trial-by-trial computation of reward prediction errors (RPEs)—to examine their relationship  
98 with neural activity.

## 99 **1. Determine cluster statistic in true data**

100 A. Define clusters: At each time-frequency index, we perform a linear univariate (or multivariate)  
101 regression using behaviorally-derived independent variables (e.g., latent cognitive variables,  
102 behavioral measures, task conditions) to predict neuronal activity (i.e., power). The  $\beta$  coefficient  
103 represents the strength and direction of the relationship between each independent variable  
104 and the dependent variable. It is estimated from the regression model and reflects how changes  
105 in the independent variable are associated with changes in power at the specific time-frequency  
106 pair. Pixel-wise regressions are parallelized for speed. For each time-frequency pair, the  $\beta$   
107 coefficient for the regressor of interest (the independent variable of primary interest) is extracted  
108 from the regression results (Fig 2A). A t-statistic is computed for the  $\beta$  coefficient to capture  
109 how significantly different it is from zero (Fig 2B). A significance threshold is applied to the  
110 t-statistics of the  $\beta$  coefficient for the regressor of interest. If the t-statistic for a time-frequency  
111 pair exceeds the significance threshold, the pair is deemed significant. Clusters are then defined  
112 as adjacent time-frequency pairs where all pairs within the cluster have t-statistics exceeding  
113 the threshold, according to the test's desired tails (Fig 2C).

114 B. Compute cluster statistics: For each identified cluster, sum the t-statistics of all time-  
115 frequency pairs within the cluster. In a two-tailed test (the default), compute both the  
116 maximum and minimum cluster sums (Fig 2D).

## 117 **2. Generate null distribution of cluster statistics**

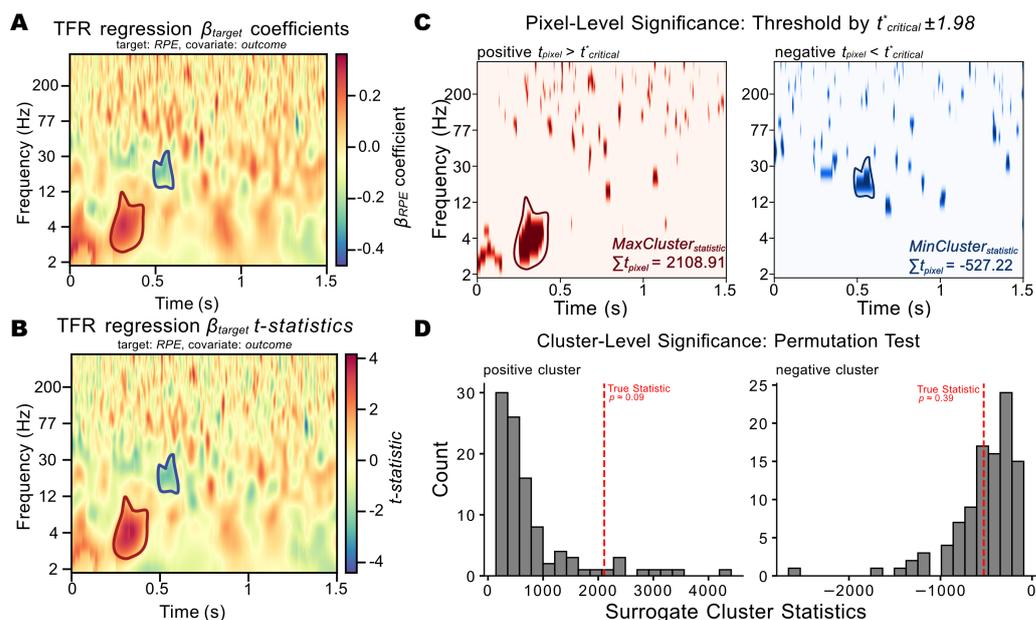
118 A. Permutation procedure: Labels for the behavioral predictor of interest are shuffled for the  
119 desired number of permutations.

120 B. Recalculate cluster statistic: Steps 1A/1B are repeated to define clusters and compute  
121 cluster statistics for each permuted dataset.

122 C. Construct null distribution: The cluster statistics from all permutations are compiled to  
123 create a null distribution, representing the distribution of cluster statistics under the null  
124 hypothesis (Fig 2E). The permuted TFR regressions are also parallelized at the pixel-level,  
125 while each permutation is performed sequentially. We tested many iterations of these functions  
126 with different parallelization approaches and sequential permutation-level computations with  
127 pixel-level parallelization within each TFR regression was the fastest method.

## 128 **3. Determine cluster significance**

129 A. Compare true cluster statistic to null distribution to compute p-values: The proportion of  
130 cluster statistics in the null distribution falling above (or below) the true cluster statistic(s)  
131 determines the p-value associated with the cluster(s) identified in the true data (Fig 2E).



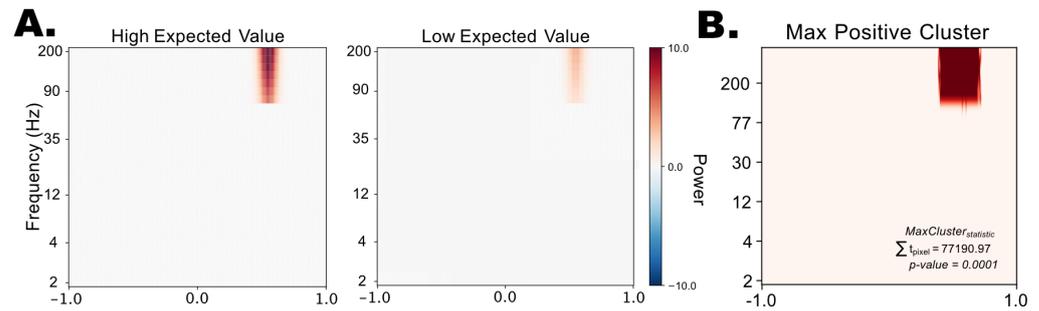
**Figure 2:** NeuroCluster methods. A.  $\beta$  coefficients for continuous predictor of interest (RPE) predicting power in given time-frequency pair (red outline = maximum positive cluster; blue outline = maximum negative cluster). B. T-statistics corresponding with  $\beta_{\text{RPE}}$  coefficients. C. Clusters as determined using t-critical threshold. D. Maximum positive and negative clusters determined by summing t-statistics in identified clusters. E. Null distribution of cluster statistics generated by permuting dataset for predictor of interest (100 permutations; red dashed line = true cluster statistic).

132 **4. Comparison of results to existing methods.**

133 To evaluate the advantages of NeuroCluster, we compared its results to those obtained using  
 134 MNE-Python's two-sample cluster-based permutation test. This approach requires discretizing  
 135 the continuous variable of interest (RPE) into distinct categories, which reduces the resolution  
 136 of the behavioral predictor. Additionally, MNE-Python's implementation does not support  
 137 multivariate analyses, limiting the ability to model multiple behavioral covariates simultaneously.  
 138 When applying the two-sample cluster test to our data, we did not identify any significant  
 139 clusters of increased or decreased activity related to RPE. In contrast, NeuroCluster successfully  
 140 detected significant clusters (Fig. 2), demonstrating its ability to preserve the richness of  
 141 continuous behavioral variables and reduce the likelihood of false negatives. This comparison  
 142 highlights NeuroCluster as a powerful and flexible alternative to existing statistical methods  
 143 for analyzing continuous brain-behavior relationships.

144 **5. Metric validation in synthetic data with known ground truth.**

145 Thus far, we have demonstrated NeuroCluster using biological data. However, because these  
 146 data are experimental, there is no definitive ground truth for the observed neural fluctuations  
 147 associated with behavioral predictors. To validate the NeuroCluster method, we generated  
 148 synthetic TFR data (2-200 Hz, sampling rate = 250, -1 to +1 seconds around "choice", 1  
 149 channel, 100 trials) with a known linear association between power in a specific time-frequency  
 150 region and a continuous behavioral variable—in this case, the expected value of choice. Code for  
 151 simulating these data is provided in the NeuroCluster repository. We then applied NeuroCluster  
 152 to the synthetic dataset and, as expected, successfully identified a significant positive cluster  
 153 corresponding to the known association embedded in the data (Fig. 3). This validation  
 154 confirms the accuracy of NeuroCluster and provides evidence against its susceptibility to false  
 155 positives.



**Figure 3:** NeuroCluster validation in synthetic data. A. Time-frequency representation (TFR) showing power differences between high ( $>0.50$ ) and low ( $<0.50$ ) expected value trials in synthetic data (1 channel, 100 trials, time-locked to “choice”). B. A significant positive cluster identified in the expected time-frequency region, consistent with the predefined association embedded in the synthetic dataset.

## 156 Acknowledgements

157 We acknowledge feedback from Shawn Rhoads and support from Xiaosi Gu and Angela  
158 Radulescu during the genesis of this project.

## 159 References

- 160 Bianchi, B., Shalom, D. E., & Kamienskowski, J. E. (2019). Predicting known sentences: Neural  
161 basis of proverb reading using non-parametric statistical testing and mixed-effects models.  
162 *Frontiers in Human Neuroscience*, *13*, 82. <https://doi.org/10.3389/fnhum.2019.00082>
- 163 Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M.  
164 J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference  
165 between two groups of structural MR images of the brain. *IEEE Transactions on Medical*  
166 *Imaging*, *18*(1), 32–42. <https://doi.org/10.1109/42.750253>
- 167 Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. MIT Press.
- 168 Collins, A. G., & Shenhav, A. (2022). Advances in modeling learning and decision-making  
169 in neuroscience. *Neuropsychopharmacology*, *47*(1), 104–118. <https://doi.org/10.1038/s41386-021-01126-y>
- 170 Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal  
171 response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to  
172 continuous stimuli. *Frontiers in Human Neuroscience*, *10*. <https://doi.org/10.3389/fnhum.2016.00604>
- 173  
174
- 175 Domenech, P., Rheims, S., & Koechlin, E. (2020). Neural mechanisms resolving exploitation-  
176 exploration dilemmas in the medial prefrontal cortex. *Science*, *369*(6507), eabb0184.  
177 <https://doi.org/10.1126/science.abb0184>
- 178 Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T.,  
179 Lara, A. H., Wallis, J. D., Knight, R. T., & others. (2020). Parameterizing neural power  
180 spectra into periodic and aperiodic components. *Nature Neuroscience*, *23*(12), 1655–1665.  
181 <https://doi.org/10.1038/s41593-020-00744-x>
- 182 Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in  
183 functional neuroimaging using the false discovery rate. *NeuroImage*, *15*(4), 870–878.  
184 <https://doi.org/10.1006/nimg.2001.1037>

- 185 Gramfort, A. (2013). MEG and EEG data analysis with MNE-python. *Frontiers in Neuroscience*,  
186 7. <https://doi.org/10.3389/fnins.2013.00267>
- 187 Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related  
188 brain potentials/fields i: A critical tutorial review. *Psychophysiology*, 48(12), 1711–1725.  
189 <https://doi.org/10.1111/j.1469-8986.2011.01273.x>
- 190 Hoy, C. W., Steiner, S. C., & Knight, R. T. (2021). Single-trial modeling separates multiple  
191 overlapping prediction errors during reward processing in human EEG. *Communications*  
192 *Biology*, 4(1), 910. <https://doi.org/10.1038/s42003-021-02426-1>
- 193 Ivanova, A. A., Schrimpf, M., Anzellotti, S., Zaslavsky, N., Fedorenko, E., & Isik, L. (2022).  
194 Beyond linear regression: Mapping models in cognitive neuroscience should align with  
195 research goals. *Neurons, Behavior, Data Analysis, and Theory*, 1. <https://doi.org/10.51628/001c.37507>  
196
- 197 König, S. D., Safo, S., Miller, K., Herman, A. B., & Darrow, D. P. (2024). Flexible multi-step  
198 hypothesis testing of human ECoG data using cluster-based permutation tests with GLMEs.  
199 *NeuroImage*, 290, 120557. <https://doi.org/10.1016/j.neuroimage.2024.120557>
- 200 Kosciessa, J. Q., Grandy, T. H., Garrett, D. D., & Werkle-Bergner, M. (2020). Single-trial  
201 characterization of neural rhythms: Potential and challenges. *NeuroImage*, 206, 116331.  
202 <https://doi.org/10.1016/j.neuroimage.2019.116331>
- 203 Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, 49(4),  
204 549–565. <https://doi.org/10.1111/j.1469-8986.2011.01320.x>
- 205 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.  
206 *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>  
207
- 208 Mathis, M. W., & Mathis, A. (2020). Deep learning tools for the measurement of animal  
209 behavior in neuroscience. *Current Opinion in Neurobiology*, 60, 1–11. <https://doi.org/10.1016/j.conb.2019.10.008>  
210
- 211 Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional  
212 neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1), 1–25. <https://doi.org/10.1002/hbm.1058>  
213
- 214 Rey, H. G., Ahmadi, M., & Quian Quiroga, R. (2015). Single trial analysis of field potentials  
215 in perception, learning and memory. *Current Opinion in Neurobiology*, 31, 148–155.  
216 <https://doi.org/10.1016/j.conb.2014.10.009>
- 217 Saez, I., Lin, J., Stolk, A., Chang, E., Parvizi, J., Schalk, G., Knight, R. T., & Hsu, M.  
218 (2018). Encoding of multiple reward-related computations in transient and sustained  
219 high-frequency activity in human OFC. *Current Biology*, 28(18), 2889–2899.e3. <https://doi.org/10.1016/j.cub.2018.07.045>  
220
- 221 Stokes, M., & Spaak, E. (2016). The importance of single-trial analyses in cognitive neuro-  
222 science. *Trends in Cognitive Sciences*, 20(7), 483–486. <https://doi.org/10.1016/j.tics.2016.05.008>  
223
- 224 Whitten, T. A., Hughes, A. M., Dickson, C. T., & Caplan, J. B. (2011). A better oscillation  
225 detection method robustly extracts EEG rhythms across brain state changes: The human  
226 alpha rhythm as a test case. *NeuroImage*, 54(2), 860–874. <https://doi.org/10.1016/j.neuroimage.2010.08.064>  
227
- 228 Yu, Z., Guindani, M., Grieco, S. F., Chen, L., Holmes, T.-C., & Xu, X. (2022). Beyond t test  
229 and ANOVA: Applications of mixed-effects models for more rigorous statistical analysis in  
230 neuroscience research. *Neuron*, 110(1), 21–35. <https://doi.org/10.1016/j.neuron.2021.10.030>  
231



Icahn  
School of  
Medicine at  
**Mount  
Sinai**

**Christina Maher**  
*Neuroscience PhD Candidate*  
*Friedman Brain Institute*  
*New York City, NY*

# **Empirical Research**

# **Intracranial recordings reveal neural encoding of attention-modulated reinforcement learning in humans**

**Christina Maher (christina.maher@icahn.mssm.edu)**

Friedman Brain Institute  
Icahn School of Medicine at Mount Sinai, New York, NY

**Salman Qasim (salman.qasim@mssm.edu)**

Friedman Brain Institute  
Icahn School of Medicine at Mount Sinai, New York, NY

**Lizbeth Nuñez Martinez (lizbeth.nunezmartinez@mssm.edu)**

Departments of Neuroscience, Neurosurgery, Neurology  
Icahn School of Medicine at Mount Sinai, New York, NY

**Ignacio Saez (ignacio.saez@mssm.edu)**

Departments of Neuroscience, Neurosurgery, Neurology  
Icahn School of Medicine at Mount Sinai, New York, NY

**Angela Radulescu (angela.radulescu@mssm.edu)**

Departments of Psychiatry and Neuroscience  
Icahn School of Medicine at Mount Sinai, New York, NY

**Abstract:** Reinforcement learning (RL) is tractable in multidimensional environments when agents maintain efficient state representations, or mental models of relevant information. Attention supports state representations in service of RL by constraining learning to relevant dimensions. However, the physiological processes supporting value updating and attentional control are unknown. To investigate the neural mechanism supporting these processes we relate attention-modulated RL models to neuronal activity recorded directly from the prefrontal cortex of neurosurgical patients playing a multidimensional decision-making task. These models revealed that participants deploy selective attention during RL. Model-estimated expected value of the chosen stimulus correlated with neuronal activity in the orbitofrontal (OFC) and lateral prefrontal cortex (LPFC), though value signals in the LPFC were additionally biased by model-estimated attention. In sum, these results provide mechanistic insight into the neuronal implementation of the computations involved in attention-modulated RL.

**Keywords:** reinforcement learning; attention; intracranial electrophysiology; human prefrontal cortex

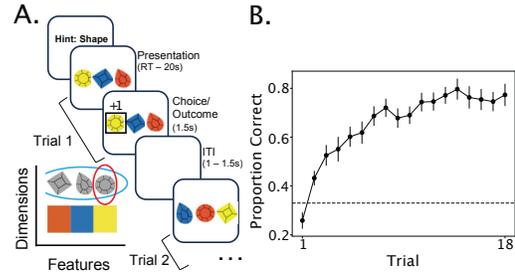
## Introduction

Attention supports real-world RL by constraining available information in multidimensional environments (Niv, 2019). In doing so, attention facilitates the maintenance of state representations, or mental models of the environment which include relevant information in service of RL. Previous work proposes an algorithmic interaction of value-based learning and attention (Leong et al., 2017; Niv et al., 2015; Wilson & Niv, 2012), processes associated with the OFC (Saez et al., 2018) and LPFC (Buschman & Miller, 2007), respectively. However, how these regions interact to support multidimensional learning is not well understood. Combining intracranial electrophysiology (iEEG) and behavioral modeling we hypothesized that: (1) participants deploy selective attention during RL, (2) OFC and LPFC encode attention-modulated expected value, (3) attention biases neural value signals. We reveal a neural mechanism by which model-based computations are implemented in the OFC and LPFC.

## Methods

Neurosurgical epilepsy patients (N=20) completed a multidimensional decision-making task in which they chose between stimuli varying along two dimensions: shape and color (Fig. 1A). In each block, participants were instructed which dimension was relevant (i.e., “shape”). Participants’ selectively attended to the relevant dimension and learned which feature (i.e., “circle”) was most rewarding. All participants performed well (Fig 1B). Gem Hunters captures naturalistic learning dynamics, as in the real-world only a subset of available information is relevant. Instructing participants

of the relevant dimension allowed us to investigate efficient state representation in service of RL.



**Figure 1:** A. Gem Hunters task (6 blocks; 18 trials per block). B. Accuracy increased across trials (N=20). Dashed line = chance. Error bars = SEM.

## RL models

We evaluated two RL models: Uniform Attention (UA) and Attention at Choice and Learning (ACL). Both models are based on Rescorla-Wagner learning rule. UA model implements uniform attention to both dimensions of each stimulus, whereas ACL model implements selective attention to the instructed relevant dimension. We assume participants choose between available stimuli based on their expected value (EV):

$$V_{(t)}(S_j) = \sum_d \phi_d \cdot v_t(d, S_j) \quad (\text{Eq. 1})$$

$V_{(t)}(S_j)$  is the value of stimulus  $j$  on trial  $t$ ,  $\phi$  is the attention weight on dimension  $d$ , and  $v_t(d, S_j)$  denotes the value of the feature in dimension  $d$  of stimulus  $S_j$ . Following feedback, a reward prediction error (RPE) is calculated:

$$\delta_t = r_t - V_t(S_c) \quad (\text{Eq. 2})$$

where  $V_t(S_c)$  is the chosen stimulus’ EV. The RPE updates the chosen stimuli’s associated feature values:

$$v_{t+1}(d, S_c) = v_t(d, S_c) + \eta \cdot \phi_d \cdot \delta_t \quad (\text{Eq. 3})$$

The update is scaled by learning rate  $\eta$ . Choice probability was computed using a softmax action selection rule. The ACL model’s  $\phi_d$  was a free parameter implementing selective attention to favor the relevant dimension (Eq.1/3). The UA model’s  $\phi_d$  was fixed at 0.50 for both dimensions.

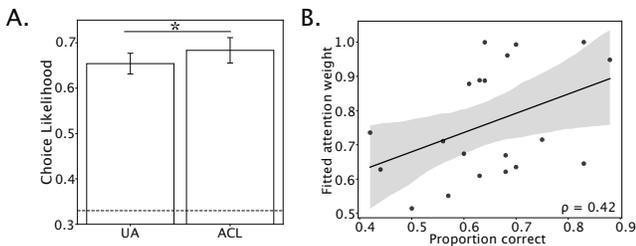
## iEEG

Local field potentials were recorded from OFC (N= 144 electrodes) and LPFC (N=124 electrodes; Fig 3A). We leveraged iEEG’s high spatiotemporal resolution to measure region-specific fluctuations in neuronal activity in response to model-based parameters. As our hypotheses involve local information encoding, we focused analyses on high gamma activity (70-200 Hz; HGA) because this signal captures population-level spatiotemporal dynamics and is correlated with fMRI

BOLD signal and single-unit spiking (Nir et al., 2007). Oscillatory power was z-scored to a baseline ITI.

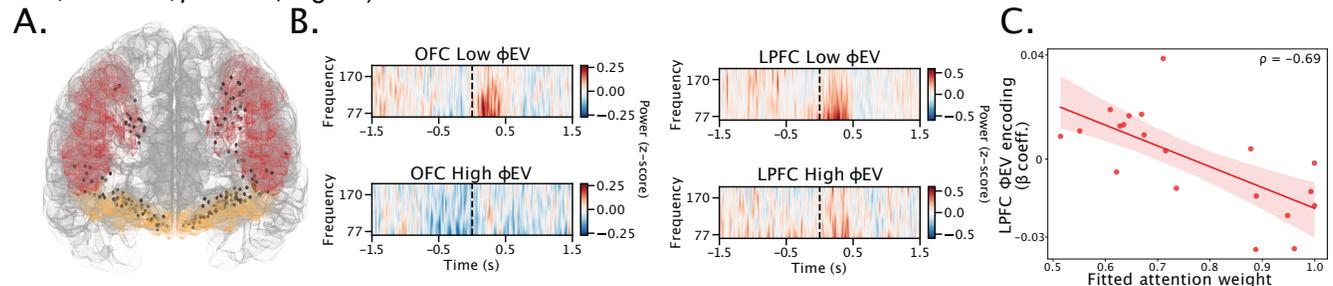
## Results and Discussion

**Selective attention modulates RL.** We used a leave-one-game-out cross validation procedure for maximum likelihood estimation. The ACL model best explained participants' behavior ( $t(19) = 2.32, p < 0.05$ ; Fig 2A). This finding confirms our hypothesis that participants deploy selective attention to maintain efficient representations of relevant information during RL. Participants' fitted attention weight ( $\phi_d$ ) was positively correlated with task performance ( $\rho(18) = 0.42, p < 0.05$ ; Fig 2B), demonstrating that even with instruction, sustained selective attention is necessary for successful RL in multidimensional environments.



**Figure 2: A.** Average choice likelihood per trial shows ACL model predicted behavior significantly better than UA model ( $p < 0.05$ ). **B.** Correlation between fitted attention weight and task performance shows attention is necessary for successful RL ( $p < 0.05$ ).

**OFC and LPFC encode attention-modulated value signals.** As hypothesized, we observed a significant effect of attention-modulated EV for the chosen stimulus ( $\Phi EV$ ; Eq.1) on OFC and LPFC HGA power. A linear mixed effects model nested within subjects was conducted within region to estimate how strongly  $\Phi EV$  was represented in OFC and LPFC HGA power while controlling for reward, chosen features, and relevant dimension.  $\Phi EV$  was represented significantly in both the OFC ( $\beta = -0.01, z = -3.78, p < 0.001$ ) and LPFC ( $\beta = -0.02, z = -3.14, p < 0.01$ ; Fig 3B).



**Figure 3: A.** Electrodes (black) in OFC (orange; 144 electrodes) and LPFC (red; 124 electrodes). **B.** Z-scored HGA power for low/high  $\Phi EV$  in two exemplar patients (OFC=11 electrodes; LPFC=5 electrodes). Dashed line = choice/reward. **C.** Correlation between fitted attention weight and LPFC attention-modulated EV encoding reveals an LPFC-specific interaction of attention and value-learning ( $p < 0.001$ ).

**LPFC value signals are biased by attention.** We found participants' selectively attend to relevant information to guide RL (Fig 2A). Further, individual differences in selective attention were related to performance (Fig 2B). Therefore, we hypothesized neural encoding of value signals will reflect an attentional bias. To test this hypothesis, subject-level estimates of  $\Phi EV$  encoding ( $\beta$  coefficient) within region were correlated with participants' fitted attention weight ( $\phi_d$ ). There was a significant negative correlation between LPFC  $\Phi EV$  encoding and  $\phi_d$  ( $\rho(18) = -0.69, p < 0.001$ ), demonstrating that greater selective attention to the relevant dimension is associated with stronger  $\Phi EV$  encoding in the LPFC. This finding suggests the neural mechanics of attention and RL are overlapping which is supported by findings in nonhuman primates (Chiang et al., 2022; Jahn et al., 2024; Wallis et al., 2001; Wallis & Miller, 2003). This finding was region specific (OFC:  $\rho(17) = -0.13, p = 0.59$ ), suggesting specialized roles for the OFC and LPFC in RL wherein the LPFC directs attention to relevant information while the OFC tracks values for relevant states (Schuck et al., 2016; Wilson et al., 2014).

## Conclusion

We leveraged behavioral modeling's parameterization of latent cognitive processes and access to direct-brain recordings in humans to identify the neural architecture that supports the computational processes underlying adaptive decision-making. By integrating attention and RL, we address the complexity of value-based learning in multidimensional environments and relate this computational solution to a biologically plausible neural mechanism. Our behavioral results suggest humans selectively attend to reward-relevant information, thus maintaining efficient state representations to guide RL. Neural results reveal OFC and LPFC HGA encodes  $\Phi EV$ . This encoding is biased by attention in the LPFC. Together our results provide neurocomputational correlates of flexible learning and decision-making.



# Intracranial substrates of meditation-induced neuromodulation in the amygdala and hippocampus

Christina Maher<sup>a,b</sup>, Lea Tortolero<sup>c</sup>, Soyeon Jun<sup>a,b</sup>, Daniel D. Cummins<sup>c</sup>, Adam Saad<sup>d</sup>, James Young<sup>d,e</sup>, Lizbeth Nunez Martinez<sup>a,b</sup>, Zachary Schulman<sup>c</sup>, Lara Marcuse<sup>d,e</sup>, Allison Waters<sup>a,b,e,f</sup>, Helen S. Mayberg<sup>a,b,c,d,e,f</sup>, Richard J. Davidson<sup>g</sup>, Fedor Panov<sup>c,e</sup>, and Ignacio Saez<sup>a,b,c,d,e,1</sup>

Edited by Michael Goldberg, Columbia University, New York, NY; received May 17, 2024; accepted December 2, 2024

Meditation is an accessible mental practice associated with emotional regulation and well-being. Loving-kindness meditation (LKM), a specific subtype of meditative practice, involves focusing one's attention on thoughts of well-being for oneself and others. Meditation has been proven to be beneficial in a variety of settings, including therapeutic applications, but the neural activity underlying meditative practices and their positive effects are not well understood. It has been difficult to understand the contribution of deep limbic structures given the difficulty of studying neural activity directly in the human brain. Here, we leverage a unique patient population, epilepsy patients chronically implanted with responsive neurostimulation devices that allow chronic, invasive electrophysiology recording to investigate the physiological correlates of LKM in the amygdala and hippocampus of novice meditators. We find that LKM-associated changes in physiological activity were specific to periodic, but not aperiodic, features of neural activity. LKM was associated with an increase in  $\gamma$  (30 to 55 Hz) power and an alternation in the duration of  $\beta$  (13 to 30 Hz) and  $\gamma$  oscillatory bursts in both the amygdala and hippocampus, two regions associated with mood disorders. These findings reveal the nature of LKM-induced modulation of limbic activity in first-time meditators.

loving-kindness meditation | amygdala | hippocampus | intracranial electrophysiology

Meditation is a set of mental techniques aimed at cultivating well-being, which require honing attentional skills related to emotional regulation (1–5). Numerous studies have proven that meditation can improve mental well-being in population-based settings (6), and potentially improve psychiatric diseases such as anxiety and depression (7). In concert with its clinical effects, meditation has been shown to change brain activity assessed through electrophysiology and functional neuroimaging (8, 9). Recent literature has parceled out primary categories of meditation, thus allowing rigorous scientific evaluation of specific practices (10). Loving-kindness meditation (LKM) is a technique within the constructive meditation family, in which practitioners actively focus their attention on cultivating positive thoughts of well-being for oneself and others. Preliminary work has suggested varying forms of meditation may share common effects on brain electrophysiology (11): This remains an important area for potential study. LKM may have therapeutic potential through the cultivation of positive emotion (12), but its underlying neural correlates are not well known, especially in deep brain areas involved in emotional regulation.

Functional and structural MRI studies have demonstrated changes in both the amygdala and the hippocampus from continued LKM practice (13, 14). EEG studies, in addition, have shown increased  $\gamma$  activity during meditation (11, 14), including during LKM in experienced meditators (3). BOLD-fMRI signals have been shown to correlate with  $\gamma$  activity (15), suggesting that these processes are related. Therefore, we first hypothesized that LKM is associated with increased  $\gamma$ -band power and duration of  $\gamma$ -band oscillatory events in both the amygdala and hippocampus (3, 11, 14). Regulation in other frequency bands may accompany gamma changes: FA meditation is linked to decreased long-range neuronal synchrony within and between brain regions.  $\beta$  oscillations are associated with attentional shifts in nonhuman and human primates: When animals are paying attention to external stimuli, beta coherence is increased to support effective information-gathering (16–19). During LKM, participants are turned to turn attention away from external stimuli and inward; therefore, we additionally hypothesized that LKM would be associated with decreased power or duration of  $\beta$  oscillatory events compared to baseline.

Despite this theoretical basis for our hypotheses regarding LKM-associated changes in neuronal activity several additional questions remain. For example, whether LKM is discretely associated with oscillatory processes rather than an overall excitation/inhibition profile

## Significance

We leverage rare chronic, invasive electrophysiology recordings while participants engage in loving-kindness meditation to demonstrate that meditation induces neural changes in beta and gamma activity in the amygdala and hippocampus of novice meditators. These results build on previous findings in experienced meditators and reveal meditation's potential for noninvasive neuromodulation of brain activity associated with emotional regulation and mood disorders.

Author affiliations: <sup>a</sup>Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029; <sup>b</sup>The Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029; <sup>c</sup>Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY 10029; <sup>d</sup>Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; <sup>e</sup>Nash Family Center for Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY 10029; <sup>f</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029; and <sup>g</sup>Departments of Psychology and Psychiatry, University of Wisconsin–Madison, Madison, WI 53706

Preprint server: This paper was submitted to bioRxiv on 5/10/2024 with distribution license: CC BY-NC-ND.

Author contributions: R.J.D. and F.P. designed research; J.Y., A.W., H.S.M., F.P., and I.S. provided/contributed funding; L.T., A.S., and L.N.M. collected the data; L.M., A.W., H.S.M., and I.S. guided data acquisition; C.M. analyzed the data, with assistance from J.Y. and supervision from I.S. and C.M., D.D.C., F.P., and I.S. interpreted the data and wrote the manuscript.

Competing interest statement: F.P. is a paid speaker for NeuroPace. Remaining authors declare no competing interests.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>To whom correspondence may be addressed. Email: ignacio.saez@mssm.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2409423122/-/DCSupplemental>.

Published February 4, 2025.

modulation in associated neural circuits has not been determined. In addition, the involvement of other frequency bands (e.g.,  $\beta = 12$  to 30 Hz) in limbic regions remains unclear, despite their role in emotional regulation and mood processes (20–22). Meditation has been associated with modulating local neuronal activity (2–4), measured by oscillatory power estimates (23), and decreased population synchrony within and between discrete regions (16, 17), measured by the duration of rhythmic, oscillatory episodes (24).

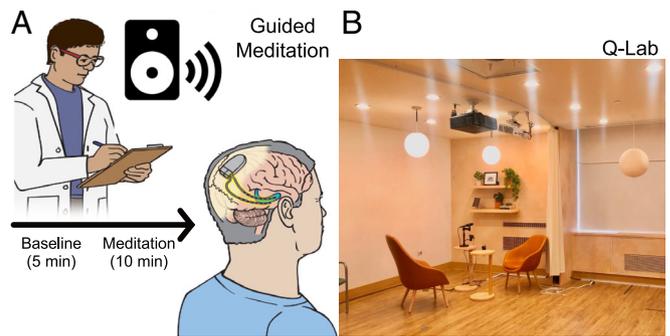
Therefore, despite its importance, the anatomically precise and physiologically detailed neural basis of meditative practice remains to be determined, especially in deep brain areas that are inaccessible to non-invasive electrophysiological recording methods. This, in turn, limits our understanding of the neural changes associated with the positive impacts of meditative practices and the development of generalizable insights that may be useful for therapeutic development. Neurosurgical interventions for the management of epilepsy allow recording from such areas in humans, including recording local field potentials (LFPs) capturing circuit activity directly via intracranial depth leads with great electrophysiological detail and signal to noise ratio (25), and providing a unique opportunity to study the neural basis of human behavior and thought. However, the most common settings for these invasive recordings, during drug-resistant epilepsy (DRE) patients' hospitalization (e.g., in the Epilepsy Monitoring Unit), are not ideal for the study of meditation because of their perioperative nature and the lack of an adequate environment for calm meditative practices. In contrast, responsive neurostimulation (RNS) system (NeuroPace Inc.) allows chronic electrophysiological brain recordings from implanted regions, frequently from the mesial temporal lobe structures of the hippocampus and amygdala (26), during the patients' daily life after surgery. This allows combining intracranial recordings with the practice of meditation in a controlled setting providing adequate environmental conditions for the practice of meditation.

Patients implanted with the RNS device can move around freely while continuous iEEG activity is recorded as LFPs. Recordings made with the RNS also offer high-quality data from deep brain structures, implicated in emotional regulation, such as the mesial temporal lobe (27). This is a major advantage in contrast to meditation studies using scalp surface EEG, which have significantly lower signal-to-noise ratio and do not allow the high simultaneous spatial and temporal resolution of RNS iEEG recordings (28). Patients implanted with the RNS are thus ideal candidates for investigating the neural correlates of naturalistic behavior, such as meditation (29, 30). In addition, DRE patients often suffer from psychiatric comorbidities including depression and anxiety (31), which provides an opportunity to study the relationship between intracranial activity and comorbid state, as well as the potential modulation during meditation.

We therefore explored changes in neural oscillatory activity associated with LKM within the amygdala and hippocampus using iEEG in DRE patients chronically implanted with an RNS device. Our findings show that first-time LKM modulates frequency-specific power and duration of oscillatory events in the hippocampus and amygdala. The selective nature of these findings in regard to the modulation of periodic, but not aperiodic, features of the neural signal reveals potential biomarkers by which LKM, a readily accessible therapeutic technique, noninvasively modulates physiological processes associated with mood regulation (20–22) even in first-time meditators.

## Results

Participants included eight neurosurgical patients with DRE who were chronically implanted with the NeuroPace Responsive Neurostimulation System (RNS). Participants completed the

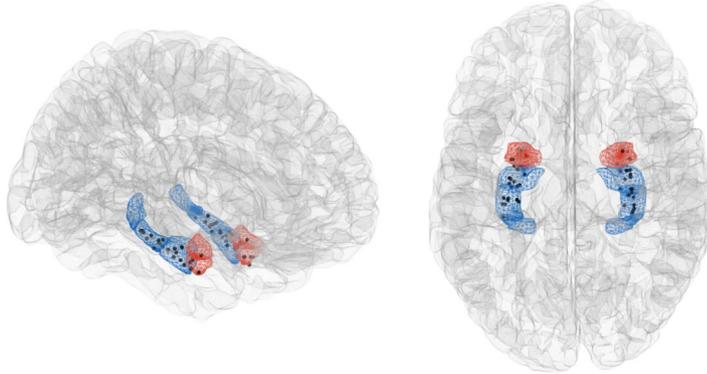


**Fig. 1.** Behavioral methods. (A) Experimental design. Subjects ( $n = 8$ ) completed a loving-kindness meditation (LKM) paradigm consisting of 5 min of audio-guided instruction (baseline) and 10 min of audio-guided LKM. (B) Experimental setting. The experimental paradigm was administered in Mount Sinai West's Q-Lab, a dedicated, immersive research environment designed to provide participants with a restorative space to participate in this experiment.

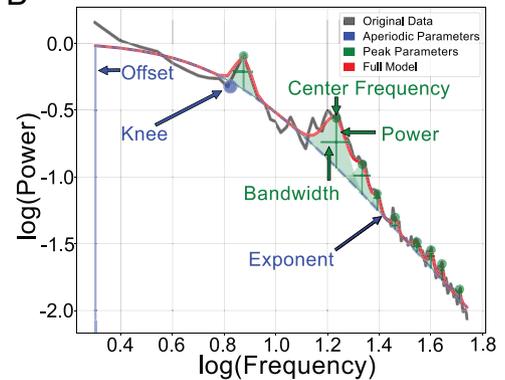
present study in Mount Sinai's Quantitative Biometric Laboratory (Q-Lab), designed to provide patients with a relaxing environment to receive therapeutic treatment free from typical distractors associated with a hospital setting and therefore highly conducive to engaging in meditative practice (Fig. 1B). Participants were self-reported novice meditators prior to the present study and completed a 5-min audio-guided instruction (baseline) followed by 10 min of audio-guided LKM (*Materials and Methods*). To evaluate the LKM induction, participants were asked to self-report their experienced depth of meditation following the session using a 1 to 10 scale (higher score = deeper meditation). On average, participants reported a high degree of meditation (mean = 7.43, SD = 2.50; *SI Appendix, Fig. S1*).

We analyzed LFPs by creating bipolar derivations between the two most anterior contacts (typically located in the amygdala and anterior hippocampus) and the two most posterior contacts (the middle-posterior hippocampus). Therefore, we collected two bipolar channels (one anterior pair and one posterior pair) per hemisphere implanted with RNS. Six patients had bilateral RNS implantations, and two patients had unilateral RNS implantation (left hemisphere). The electrode implant sites were determined solely based on clinical criteria, with the amygdala contacts being bipolar referenced to the anterior hippocampus. For readability, we refer to these bipolar recordings as “amygdala.” In contrast, the electrode pairs referred to as “hippocampus” represent hippocampus–hippocampus bipolar derivatives, enabling us to distinguish activity originating specifically from the hippocampus from signals that include unique contributions from the amygdala. All participants included in the present study had at least one contact in either the amygdala or hippocampus (count: amygdala = 14 electrodes/13 bipolar channels; hippocampus = 36 electrodes/14 bipolar channels; Fig. 2A). Because of hardware limitations on the sampling rate of the recorded electrophysiological data (125 Hz maximum), we restricted our analyses to frequencies up to low gamma (55 Hz). Anatomical localization of electrodes was determined by coregistering high-resolution postoperative CT scans with preoperative MRI (*Materials and Methods*). Although RNS implantation occurs in DRE patients' presumptive seizure onset zone, the leads which are composed of 4 contacts typically span just over 30 mm of tissue (1.5 mm in length for each contact and 10 mm between contact centroids). This results in a proportion of data recorded from likely normative tissue (29, 30). To mitigate the influence of interictal noise in the data, we implemented a data preprocessing approach mentioned in previous publications (32). Briefly, following visual inspection of all channel data, we confirmed the absence of any stimulation artifacts and eliminated

A



B



**Fig. 2.** Neural methods. (A) Anatomical reconstruction showing the hippocampal and amygdala location of RNS contacts. Depicted is the placement of 50 NeuroPace RNS electrodes in the amygdala (blue) and hippocampus (red) across eight participants. Each black dot corresponds to one electrode (amygdala = 14 electrodes/13 bipolar channels; hippocampus = 36 electrodes/14 bipolar channels). (B) FOOOF approach. We characterized LFP activity from bipolar channels recorded from depth electrodes in the amygdala and hippocampus. We used the FOOOF approach to separately characterize aperiodic (i.e.,  $1/f$  background activity) from oscillatory neural activity. Depicted is an example power spectrum from one hippocampal channel (gray trace) overlaid by the FOOOF fitted model (red trace) parameterized to extract aperiodic (blue trace) and periodic (green trace) spectral features between 2 and 55 Hz. The aperiodic components are characterized by the offset, knee, and exponent. Periodic components are assigned to a canonical frequency band ( $\delta = 2\text{--}4$  Hz,  $\theta = 4\text{--}8$  Hz,  $\alpha = 8\text{--}13$  Hz,  $\beta = 13\text{--}30$  Hz,  $\gamma = 30\text{--}55$  Hz) depending on their center frequency, and their power and bandwidth estimated.

any data meeting clinical criterion for interictal discharge (33). This resulted in discarding approximately 6% of the data. The proportion of time across participants in which interictal activity was detected was not significantly different between conditions (baseline and meditation;  $P > 0.05$ , Pearson's chi-square goodness-of-fit test; *SI Appendix, Fig. S2A*). The proportion of omitted data from the entire experimental session ranged from 2 to 16% across participants (*SI Appendix, Fig. S2B*).

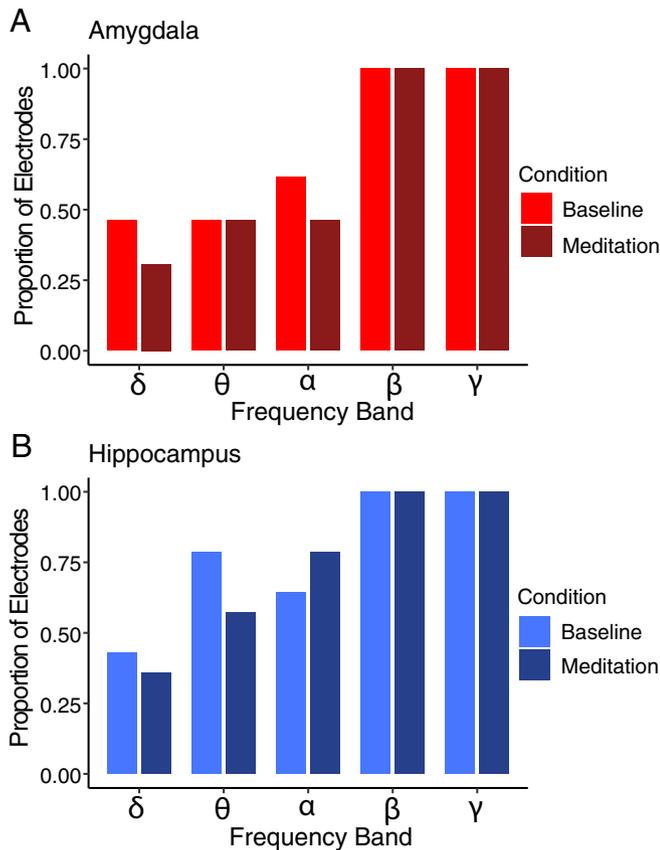
**LKM Was Not Accompanied by Changes in Aperiodic Neural Activity.** We set out to identify whether meditation was accompanied by changes in neural activity by comparing LFP activity patterns between active control (learning about meditation) and LKM epochs. Given the largely temporally unresolved nature of these data, we chose to focus our analysis on the power profile throughout a single epoch during the meditation block, examining both aperiodic ( $1/f$  profile) and periodic (i.e., oscillatory) neural activity on each bipolar channel. We parameterized the aperiodic and periodic features of the power spectra using the 'fitting oscillations and one over  $f$ ' method (FOOOF; see *Materials and Methods* and Fig. 2B) (23). The FOOOF model fit was performed for each channel's data in each condition (baseline and meditation). The aperiodic components include a knee parameter, which accounts for an often-observed bend in the  $1/f$  profile, and offset and exponent, which reflect the y-intercept and rate of decay of the  $1/f$  profile, respectively. Together, these aperiodic components capture broadband shifts in the  $1/f$  profile often ascribed to changes in excitatory/inhibitory balance, different and separate from oscillations in individual frequency bands. In addition, oscillatory components were estimated for predefined frequency bands ( $\delta = 2\text{--}4$  Hz,  $\theta = 4\text{--}8$  Hz,  $\alpha = 8\text{--}13$  Hz,  $\beta = 13\text{--}30$  Hz,  $\gamma = 30\text{--}55$  Hz; see *Materials and Methods*).

We first sought out to identify whether acute LKM induced changes in the aperiodic features of the power spectra by comparing knee frequency, offset, and exponent separately between baseline and LKM epochs (Fig. 4). We did not find significant differences between conditions in knee frequency, offset, or exponent in either the amygdala or hippocampus (all  $P > 0.05$ , two-sided paired-samples Wilcoxon signed rank test; Fig. 3), suggesting that meditation is not accompanied by general changes in the excitatory/inhibitory profile of either amygdala or hippocampus. Further, we found no significant differences in aperiodic

components between the amygdala and hippocampus (all  $P > 0.05$ , two-sided Wilcoxon signed rank test; *SI Appendix, Fig. S3*).

**LKM Was Not Accompanied by Changes in the Proportion of Electrodes Showing Oscillatory Activity.** Next, we evaluated whether LKM was accompanied by changes in oscillatory neural activity by examining power across frequencies in baseline and meditation conditions. We considered a significant oscillation to be present if at least one peak within predefined frequency bins ( $\delta = 2\text{--}4$  Hz,  $\theta = 4\text{--}8$  Hz,  $\alpha = 8\text{--}13$  Hz,  $\beta = 13\text{--}30$  Hz,  $\gamma = 30\text{--}55$  Hz) was detected by the FOOOF model fit. If no peak was found within a given frequency range for a channel, we considered the channel to not contain an oscillation in that band. Therefore, we started by quantifying the proportion of channels containing active oscillations across conditions. We observed differences in the proportion of active channels across frequencies between conditions in both amygdala and hippocampus, with higher frequencies containing a larger proportion of active channels than lower frequencies (Fig. 4). The lowest proportion of active electrodes was in  $\delta$  in both amygdala (30.8%) and hippocampus (35.7%); the highest was in  $\beta$  and  $\gamma$  in both regions (100%). However, we did not find significant differences in the proportion of active electrodes between conditions in any frequency bands (all  $P > 0.05$ , Fisher's exact test). Therefore, LKM was not associated with an increase in the proportion of channels showing significant oscillations. These data indicate that our ability to detect significant oscillations was strongly frequency dependent, and that there was significant  $\beta$  and  $\gamma$  oscillatory activity in both baseline and in meditation conditions in the amygdala and hippocampus.

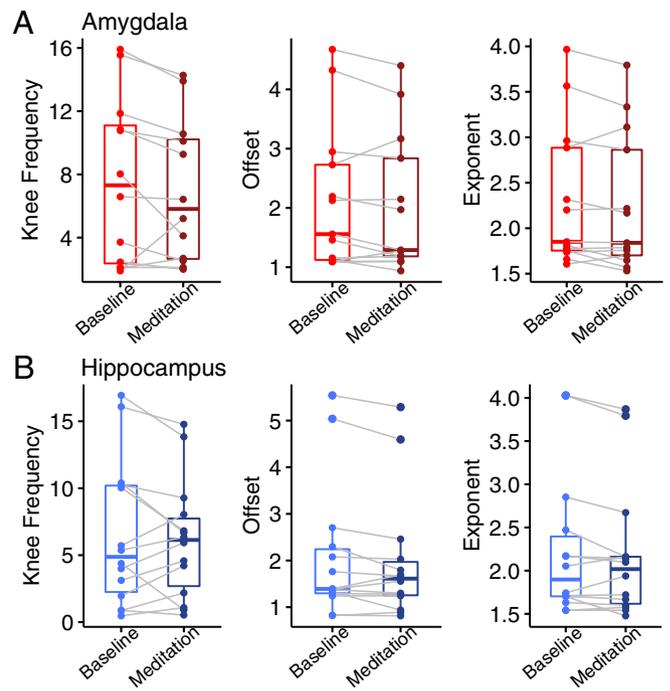
**LKM Was Associated with an Increase in  $\gamma$  Power.** To further investigate whether LKM modulated oscillatory activity, we compared the amplitude, or power, of detected oscillations between conditions in both regions (Fig. 5). To investigate between-condition differences in oscillatory power, we only considered channels in which at least one peak was detected in each frequency band (i.e., containing significant oscillatory activity). If more than one peak was found within a given frequency range, the average power of all detected peaks within the frequency band was computed to determine an average power score. We observed a significant increase in  $\gamma$  power during LKM in both the amygdala ( $P < 0.01$ , one-sided paired-samples Wilcoxon



**Fig. 3.** No difference in aperiodic neural components between baseline and meditation epochs. The aperiodic components of FOOOF model fit (knee frequency, offset, and exponent) was extracted and compared between conditions (FA meditation and baseline). (A) No differences in FOOOF aperiodic parameters in amygdala channels. No significant differences in knee frequency, offset, or exponent of FOOOF model fit for amygdala channels ( $n = 13$  bipolar channels) were observed between conditions (all  $P > 0.05$ , Two-sided paired sample Wilcoxon signed-rank test). (B) No differences in FOOOF aperiodic parameters in hippocampus channels. No significant differences in knee frequency, offset, or exponent of FOOOF model fit for hippocampus channels ( $n = 14$  bipolar channels) were observed between conditions (all  $P > 0.05$ , Two-sided paired sample Wilcoxon signed-rank test).

signed rank test, Fig. 5A) and hippocampus ( $P < 0.001$ , one-sided paired-samples Wilcoxon signed rank test, Fig. 5B). We did not observe concomitant changes in  $\beta$  ( $P > 0.05$ , one-sided paired-samples Wilcoxon signed rank test, Fig. 5) or any other frequency bands (all  $P > 0.05$ , two-sided paired-samples Wilcoxon signed rank test, Fig. 5), indicating that this modulation was specific to the  $\gamma$  frequency band. Overall, these results reveal that LKM is accompanied by changes in oscillatory power in active channels, but not an increase in the proportion of active channels. Furthermore, these changes are specific to high-frequency activity ( $\gamma$ ) and not present in lower frequencies. We did not find differences in baseline oscillatory power or the difference in oscillatory power from baseline to meditation between the amygdala and hippocampus (all  $P > 0.05$ , two-sided Wilcoxon signed rank test; *SI Appendix, Figs. S4A and S5A*).

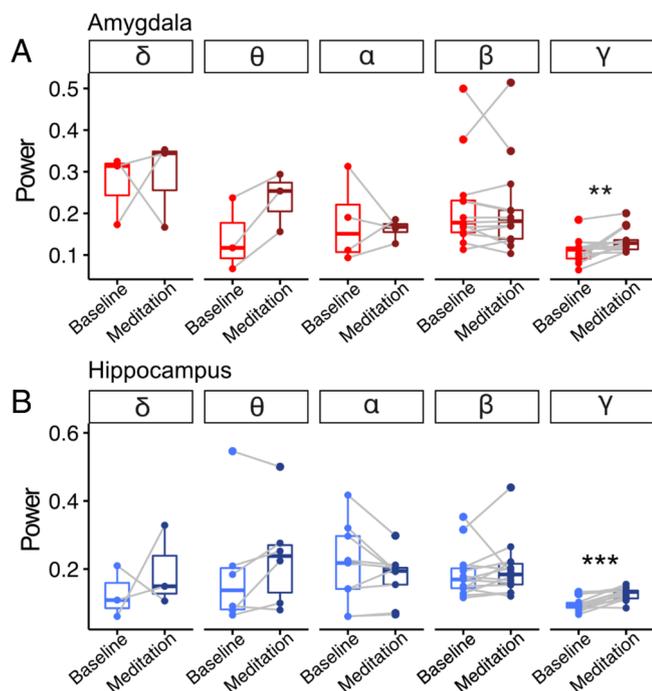
To examine the significance of power modulations at the subject level, we performed subject-level analyses of power differences by leveraging the high temporal resolution of the iEEG data. We focused our model-free analysis on the beta and gamma frequency bands, as our FOOOF model fitting confirmed the presence of both beta and gamma oscillations across all channels in both baseline and LKM conditions. Briefly, we estimated power for 1-s neural recording segments across all channels for each subject.



**Fig. 4.** Proportion of channels showing neural oscillations varied across frequency bands. Plotted is the proportion of channels in which an oscillation was detected during baseline and meditation epochs in each frequency band ( $\delta = 2-4$  Hz,  $\theta = 4-8$  Hz,  $\alpha = 8-13$  Hz,  $\beta = 13-30$  Hz,  $\gamma = 30-55$  Hz) according to FOOOF power spectrum model fit. (A) Proportion of amygdala channels showing significant oscillatory activity across frequencies. The proportion of channels showing significant modulation varied between ~30% in  $\delta$  (2 to 4 Hz) and 100% in  $\gamma$  (30 to 55 Hz), with greater proportions for higher frequencies. There were no differences in the proportion of amygdala channels ( $n = 13$  bipolar channels) in which oscillations were detected between baseline (light blue) and meditation (dark blue) epochs (all  $P > 0.05$ , Fisher's exact test). (B) Proportion of hippocampal channels showing significant oscillatory activity across frequencies. The pattern of activation was like that observed in the amygdala, with the proportion of channels ranging between ~30% in  $\delta$  (2 to 4 Hz) and 100% in  $\gamma$  (30 to 55 Hz), with greater proportions for higher frequencies. There were no differences in the proportion of hippocampal channels ( $n = 14$  bipolar channels) in which oscillations were detected between baseline (light blue) and meditation (dark blue) epochs (all  $P > 0.05$ , Fisher's exact test).

LKM segments were z-scored relative to baseline segments, providing a normalized measure of LKM-induced changes in each frequency band. As anticipated, we observed a significant increase in gamma power during LKM in both the amygdala (Fig. 6A) and hippocampus (Fig. 6A) for most subjects (amygdala: 6/8 participants; 75%; hippocampus: 6/7 participants; 86%). In contrast, there was no significant difference in beta power during LKM compared to baseline in either region for most subjects (both  $P > 0.05$ , Fig. 6B). These results align with the group-level, model-based findings, further supporting a significant LKM-associated increase in gamma power, and provide limited evidence for modulations in beta power in some, but not all, patients.

**LKM Was Accompanied by an Alteration in Duration of  $\beta$  and  $\gamma$  Bursts.** The FOOOF method allowed us to investigate aperiodic and periodic components of brain activity but does not offer a way to quantify quick bursts of neural activity which may be important to facilitate switches in cognitive states (e.g., from baseline to LKM). To investigate this possibility, we applied the extended Better Oscillation detection (eBOSC) method (24, 34), to investigate how meditation modulated the duration of oscillations across frequency bands (*Materials and Methods*). eBOSC allows us to detect temporal windows with significant frequency-specific



**Fig. 5.** Group-level increased  $\gamma$  oscillatory power in meditation compared to baseline. We examined power modulation across conditions (baseline versus LKM) by averaging spectral power estimates (the periodic component of FOOOF model fit) within frequency bands ( $\delta = 2-4$ ,  $\theta = 4-8$ ,  $\alpha = 8-13$ ,  $\beta = 13-30$ ,  $\gamma = 30-55$ ) and compared across conditions. For this analysis, we employed only the subset of channels that showed significant oscillations in both conditions, which varied across frequency bands. (A) Increased  $\gamma$  oscillatory power in amygdala electrodes during meditation. We observed a significant increase in amygdala  $\gamma$  power (30 to 55 Hz) during LKM compared to baseline ( $P < 0.01$ , one-sided paired-samples Wilcoxon signed rank test; all other frequency bands  $P > 0.05$ ). (B) Increased  $\gamma$  oscillatory power in hippocampal channels during meditation. As in the amygdala channels, we observed a significant increase in hippocampus  $\gamma$  power (30 to 55 Hz) during LKM compared to baseline ( $P < 0.001$ , one-sided paired-samples Wilcoxon signed rank test; all other frequency bands  $P > 0.05$ ).

oscillations that surpass power and duration thresholds while accounting for the background  $1/f$  profile of the neural signal (24). Further, recent work has implemented this method with hippocampal iEEG data collected using the RNS device, allowing us to compare our findings with previously published iEEG work (30). Using eBOSC, we calculated the proportion of time within a given experimental condition (baseline or LKM) in which a particular oscillation was detected. We averaged these proportions for each frequency in each of our predefined frequency bands and compared between conditions to identify whether LKM induced changes in the duration of oscillatory events (Fig. 7). We found that LKM was associated with a significant decrease in the duration of  $\beta$  oscillations in both the amygdala ( $P < 0.05$ , one-sided paired-samples Wilcoxon signed rank test, Fig. 7A) and hippocampus ( $P < 0.05$ , one-sided paired-samples Wilcoxon signed rank test, Fig. 7B). Further, we found a significant increase in the duration of  $\gamma$  oscillations during LKM in the amygdala ( $P < 0.01$ , one-sided paired-samples Wilcoxon signed rank test, Fig. 7A). Additionally, the amygdala displayed a significantly greater increase in the duration of  $\gamma$  oscillations during meditation compared to the hippocampus ( $P < 0.05$ , one-sided Wilcoxon signed rank test, *SI Appendix*, Fig. S5B; all other frequency bands  $P > 0.05$ ). Overall, these results demonstrate a decrease in the amount of  $\beta$  events during meditation in both amygdala and hippocampus, and an increase in  $\gamma$  events in the amygdala, suggesting an involvement of fast oscillatory events in meditative states. We did not find differences in baseline oscillatory duration

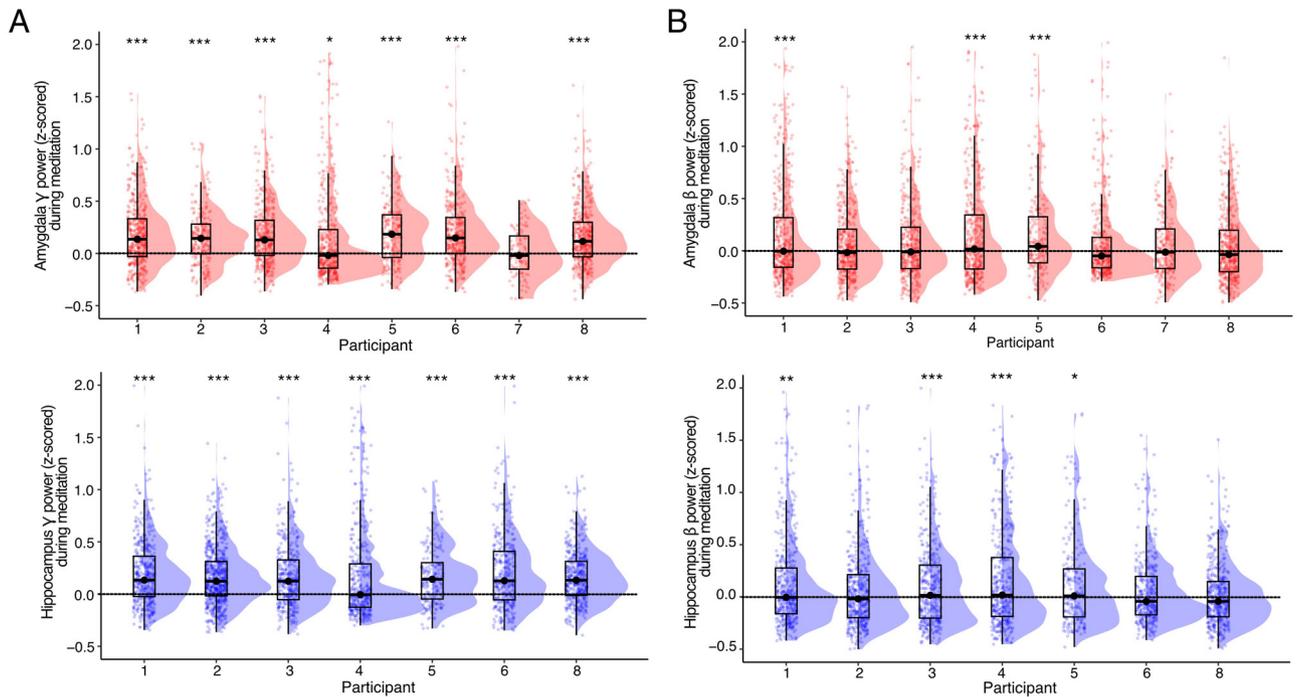
between the amygdala and hippocampus (all  $P > 0.05$ , two-sided Wilcoxon signed rank test; *SI Appendix*, Figs. S4B and S5B).

## Discussion

This study leveraged unique access to chronic ambulatory iEEG recordings in the amygdala and hippocampus during LKM. We directly examined the nature of neurophysiological activity during LKM in the amygdala and hippocampus, finding primary results of increased  $\gamma$  oscillations and decreased  $\beta$  duration during LKM. Further, changes in neural dynamics with meditation occurred in oscillatory activity, rather than in aperiodic neural activity. Research on the effects of meditation on neural dynamics to date has been limited to fMRI and scalp EEG, due to rare access to iEEG in real-world settings. While such approaches have advanced the field of meditation, iEEG research can augment such data with increased spatial and temporal resolution.

Meditative practices have existed for millennia in different traditions (35) and allow humans to cultivate fundamental states of focus and emotional regulation (36). Meditation, perhaps independent of practice type, induces scalp EEG increases in  $\gamma$  activity (11). In addition, meditation induces changes in BOLD-fMRI signal in the hippocampus and amygdala (13, 37), which could correlate with changes in  $\gamma$  oscillatory activity (15). Our study directly addresses the nature of intracranial electrophysiology within the hippocampus and amygdala during meditation. Our focus on LKM comes from previous research that connects this emotionally laden contemplative practice to the amygdala (13). LKM, utilized in this study, stresses finding joy and sharing it with others (10). Such constructive contemplative practices have been shown to induce positive emotions, which may lead to an increased sense of purpose and meaning in life (38). Such training has also been shown to chronically modulate the amygdala, hence of specific interest in our cohort where contacts were situated in the basolateral amygdala in all cases (13). Yet, no experiments to date have attempted to record such changes in novice meditators acutely and intracranially. These findings complement the earlier results on increased  $\gamma$  oscillations in LKM practices in long-term meditators by showing specific increases in these fast oscillations in novice meditators when directly examining the amygdala and hippocampus with iEEG.

**Amygdala and Hippocampus  $\gamma$  Power Increases During LKM.** By leveraging the spatiotemporal resolution afforded by iEEG, our results build on prior research on  $\gamma$  power increases associated with long-term LKM practice (3, 11, 39, 40) extending these findings to first-time meditators and providing insights into the nature of neural changes associated with meditation in an anatomically precise way. We observed an increase in amygdala and hippocampus  $\gamma$  power (Fig. 5), whose correlation to fMRI-BOLD signal is well established (15), suggesting LKM induces heightened activation of local neuronal ensembles within the amygdala and hippocampus. This meditation-induced neural modulation may be related to the known role in processing of emotional information in the amygdala and hippocampus (41), and possibly with related mental processes such as memory and attention. During LKM, participants are encouraged to actively retrieve positive autobiographical memories during LKM which induces amygdala and hippocampus activation (42). The hippocampus plays a critical role emotional regulation and attention (43, 44) alongside its fundamental roles memory consolidation and retrieval (45, 46), suggesting hippocampal activation may be related to the memory aspects of the meditative task. The amygdala, in turn, can direct attention toward emotionally significant stimuli (13, 45, 47), serving a critical function in the



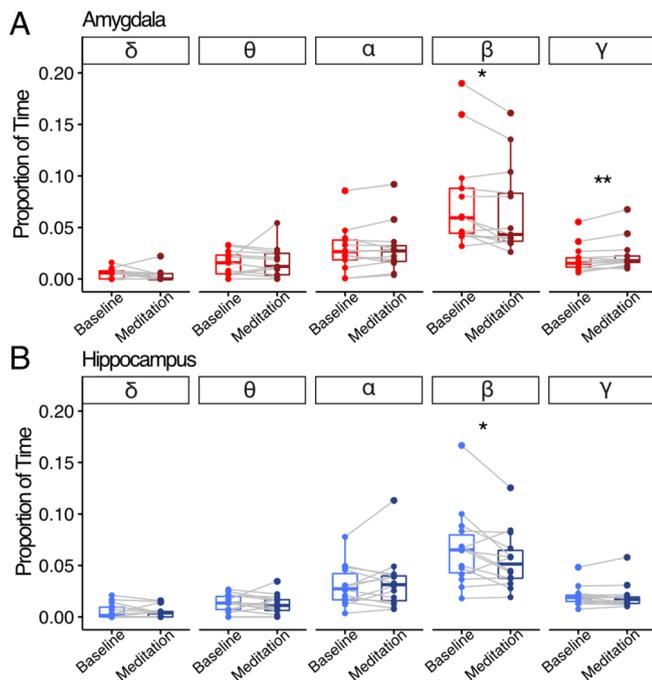
**Fig. 6.** Subject-level increased  $\gamma$  oscillatory power in meditation compared to baseline. We analyzed power modulation within subjects by dividing the LKM session into 1-s epochs and performing frequency decomposition to estimate  $\gamma$  and  $\beta$  power for each epoch. These epochs were then z-scored against the mean and SD of  $\gamma$  and  $\beta$  power during the baseline period (each point represents one 1 s epoch). A positive z-score indicates an increase in power during the LKM session relative to baseline. (A) Generalized gamma-band power increase in the amygdala and hippocampus. Most participants ( $n = 6/8$ , 75%) exhibited significantly greater amygdala  $\gamma$  power during meditation compared to baseline [one-sided one-sample Wilcoxon signed rank test; participant (median,  $P$ -value) = 1 (0.14,  $P < 0.001$ ), 2 (0.14,  $P < 0.001$ ), 3 (0.13,  $P < 0.001$ ), 4 (-0.02,  $P < 0.001$ ), 5 (0.18,  $P < 0.001$ ), 6 (0.15,  $P < 0.001$ ), 7 (-0.02,  $P > 0.05$ ), 8 (0.12,  $P < 0.001$ )]. Most participants ( $n = 6/7$ , 86%) exhibited significantly greater hippocampus  $\gamma$  power during meditation compared to baseline [one-sided one-sample Wilcoxon signed rank test; participant (median,  $P$ -value) = 1 (0.13,  $P > 0.05$ ), 2 (0.12,  $P > 0.05$ ), 3 (0.12,  $P > 0.05$ ), 4 (-0.01,  $P > 0.05$ ), 5 (0.14,  $P > 0.05$ ), 6 (0.13,  $P > 0.05$ ), 8 (0.13,  $P > 0.05$ )]. (B) Beta-band power modulation in the amygdala and hippocampus in a subset of patients. Only a minority of participants exhibit significantly different  $\beta$  power during meditation compared to baseline in the amygdala [3/8 patients; two-sided one-sample Wilcoxon signed rank test; participant (median,  $P$ -value) = 1 (-0.01,  $P < 0.001$ ), 2 (-0.02,  $P > 0.05$ ), 3 (-0.01,  $P > 0.05$ ), 4 (0.02,  $P < 0.001$ ), 5 (0.04,  $P < 0.001$ ), 6 (-0.05,  $P > 0.05$ ), 7 (-0.01,  $P > 0.05$ ), 8 (-0.04,  $P > 0.05$ )] or hippocampus [two-sided one-sample Wilcoxon signed rank test; participant (median,  $P$ -value) = 1 (-0.01,  $P < 0.01$ ), 2 (-0.02,  $P > 0.05$ ), 3 (0.01,  $P < 0.001$ ), 4 (0.02,  $P < 0.001$ ), 5 (0.01,  $P < 0.05$ ), 6 (-0.04,  $P > 0.05$ ), 8 (-0.04,  $P > 0.05$ )].

bottom-up processing of emotionally relevant information (48–50). We further identified an increase in the duration of  $\gamma$  oscillatory events during LKM compared to baseline in the amygdala, but not hippocampus (Fig. 7). This may reflect the specific physiological demands of LKM, wherein the hippocampus exhibits more transient  $\gamma$  bursts associated with memory retrieval (51) while the amygdala exhibits more sustained, rhythmic  $\gamma$  activity to support emotional processing (52).

**Amygdala and Hippocampus  $\beta$  Oscillatory Duration Decreases During LKM.**  $\beta$  oscillations reflect rhythmic phase-locked activity and are thought to facilitate long-range information exchange via cross-regional synchrony. The synchrony afforded by ongoing  $\beta$  oscillations poises neural ensembles to quickly reorganize and integrate incoming information. In human and nonhuman primates, synchrony in lower-frequency (i.e.,  $\beta$ ) oscillations is thought to modulate attention switches (18, 53). This physiological mechanism provides an attentional spotlight that facilitates effective information-gathering from the environment (19). This affordance is advantageous in circumstances which require real-time monitoring and adapting to one's environment. In contrast, one seeks to reduce the influence of external distractions during LKM. Therefore, our present findings of decreased  $\beta$  oscillatory duration during LKM compared to baseline support the notion that  $\beta$  oscillations are a physiological mechanism for long-range temporal synchrony that regulates selective attention in service of LKM. Transient  $\beta$  bursts, instead of sustained rhythmic oscillations, facilitate cognitive processes through functional

inhibition (51).  $\beta$  bursts dictate the spatial and temporal basis of activation relevant for memory- and attention-related processing, while gamma power reflects the processing itself (54). Increased cortical  $\beta$  oscillations have been correlated with negative mood disorders (22, 53, 55). Decreased  $\beta$  oscillations with LKM may also represent a shift from negative emotional states, in turn for more positively salient  $\gamma$  oscillations at the network level. Additionally, given the role of  $\beta$  oscillations in premotor preparation and the fact that LKM can be conceptualized as being action-oriented (56–59), it is possible that our observed  $\beta$  modulation reflects a motor plan in its preparatory stages. These group-level changes in beta oscillatory duration were associated in some patients with modulations in beta power (Fig. 6), suggesting that interpatient heterogeneity may play a role in shaping these neural responses.

**Meditation Does Not Affect Aperiodic Activity or Differentially Impact the Hippocampus versus Amygdala.** Our finding that meditation was not accompanied by changes in aperiodic neural activity suggests that any LKM practice-induced neural fluctuations are specifically related to oscillatory events, not other physiological processes (Fig. 3). We did not find differences in any oscillatory metric (periodic/aperiodic FOOOF metrics or eBOSC) between the amygdala and hippocampus. This may indicate they are similarly modulated, with an important caveat that our amygdala contacts are bipolar referenced to the head of the hippocampus. Such confounds may be overcome in future experiments with adjustments of the recording paradigms of the RNS to monopolar, yet, at this time are beyond the scope of this paper.



**Fig. 7.** Modulation of oscillatory duration across meditation states. We used a rhythm detection method (eBOSC) to determine the duration of rhythmic, oscillatory activity. Estimates of oscillatory duration were averaged within frequency bins ( $\delta = 2-4$ ,  $\theta = 4-8$ ,  $\alpha = 8-13$ ,  $\beta = 13-30$ ,  $\gamma = 30-55$ ) and compared between conditions. (A) Amygdala channels showed a decrease in  $\beta$  and an increase in  $\gamma$  oscillation duration. The duration of amygdala  $\beta$ -range oscillatory activity during LKM decreased compared to baseline ( $P < 0.05$ , one-sided paired-samples Wilcoxon signed rank test), and the duration of amygdala  $\gamma$ -range oscillatory activity during LKM increased compared to baseline ( $P < 0.01$ , one-sided paired-samples Wilcoxon signed rank test; all other frequency bands  $P > 0.05$ ). (B) Hippocampus channels showed a decrease in  $\beta$  oscillation duration. The duration of hippocampal  $\beta$ -range oscillatory activity during LKM decreased compared to baseline ( $P < 0.05$ , one-sided paired-samples Wilcoxon signed rank test; all other frequency bands  $P > 0.05$ ).

**Naturalistic Study on the Neural Correlates of Meditation.** This work on the neural correlates of meditation using RNS has a multitude of advantages over other approaches. Mainly, the RNS allows detailed spatiotemporal study of deep brain structures in a naturalistic setting. While work from fMRI and other neuroimaging studies have gleaned important findings, they are limited in their temporal resolution. Clinically relevant findings from these studies, such as biofeedback for treatment of negative emotions (45), are difficult outside of the naturalistic setting, as with RNS. Our work builds on the uses of RNS to study the neural basis of human brain function in the naturalistic setting, which has been implemented for tasks ranging from spatial navigation (30, 60), to speech (61).

**Study Limitations.** Our experimental design results in several limitations of this study. First, the technical limitations of the RNS device (which is limited to a 250 Hz sampling rate) imposed an upper limit on the frequency bands that could be resolved, and therefore we could not fully estimate the effects of meditation on broadband high-frequency activity (typically 70 to 200 Hz), a marker of local activation (32, 62). We were, however, able to estimate gamma activity (30 to 55 Hz), which also acts as a proxy for local cortical activation (63, 64). Second, because of the limited time available for patient testing, the experimental paradigm did not include a post-meditation active control period, limiting our ability to test for postmeditation baseline changes as reported in previous studies (3). Similarly, the one-shot design limits our ability to draw conclusions regarding

the chronic effects of meditation practice. Our ability to draw conclusions regarding the generalizability of findings is limited by experimental constraints with a clinical sample as baseline period could only consist of a single session at the start of the meditation phase, during which participants listened to audio instructions about meditation, and we did not include multiple baseline periods for further comparison. Whereas it is possible that the observed neuromodulation reflects the influence of factors beyond LKM, the main difference between the active control baseline and the LKM period was the meditation practice itself. Participants maintained the same seated position, received the same quality of audio prompts, and were exposed to the same environmental stimuli in both conditions. Future studies could benefit from incorporating multiple control sessions, a more robust baseline period, counterbalanced order of baseline and meditation periods, and additional forms of meditation or relaxation practice to better isolate the specific contributions of LKM to the observed effects. Third, by including the knee parameter in our FOOOF model fit, we avoid assuming the aperiodic component of the power spectra maintain a single  $1/f$ -like characteristic, as this assumption is often violated across broad frequency ranges ( $\sim$ greater than 40 Hz). While providing a better fit to the data, this model fitting choice may exhibit a bias against lower frequencies, potentially explaining the disproportionate detection of peaks at higher frequencies compared to lower ones in the present analysis. Additionally, the lack of a relationship between self-reported depth of meditation and neurophysiological metrics raises the question of whether participants genuinely engaged in meditation, and the challenges of estimating internal states in a non-intrusive way. More generally, it is possible that our observed neural changes reflected generalized changes in states of arousal or affect. The addition of neurophysiological measures such as heart rate or skin conductance, would help relate amygdala and hippocampal activation to less invasive physiological metrics and help determine the specificity of neural activation versus more general metrics, e.g., arousal. Finally, because of the difficulties of recruiting RNS patients, our study was necessarily limited to a small patient sample ( $n = 8$ ). This is a common limitation of intracranial studies in general, and RNS studies in particular, which often have similar or, frequently, smaller number of patients compared to our study (29, 65). In addition, this concern is alleviated by the significant homogeneity in electrode placement, which limits cross-patient anatomical variability compared to other invasive studies (e.g., sEEG) and by the great signal-to-noise ratio of intracranial recordings, which allowed us to carry out group-level inferences (e.g., Fig. 5).

**Study Strengths.** This study has several key strengths. First, the use of intracranial recordings offers high spatial and temporal precision, allowing for direct measurement of brain activity in subcortical regions, such as the amygdala and hippocampus, which are known for their roles in emotion regulation and memory. These processes are thought to be triggered by LKM practice but specific neuronal activation of these regions during LKM are difficult to measure with non-invasive methods. Additionally, the RNS system enables data collection from fully ambulatory participants, allowing the study to take place in a naturalistic setting. This improves the study's ecological validity by recording direct brain activity while participants behave in conditions that closely resemble a real-world meditation setting. Finally, the statistical methods used in this study account for both periodic and aperiodic neural activity, offering a detailed and mechanistic understanding of how LKM affects the brain.

**Future Directions.** These results are the first to document acute intracranial electrophysiology changes with LKM. Specifically, we see an increase in  $\gamma$  oscillations and a decrease in  $\beta$  burst duration in the amygdala and hippocampus with LKM. While we observe acute changes in amygdala and hippocampal electrophysiology during LKM, we did not assess long-term changes that may occur with LKM practice. The meditation literature has defined “state changes” as temporary modulations of mind (and implicitly brain function) that occur with meditation, versus trait changes as long-term, enduring effects of meditation on the mind (and thus brain) (65). Our research documents brain signals that may contribute to the acute state changes of LKM but does not explore long-term brain changes that may underlie traits unique to long-term meditators. Yet, the RNS system enables long-term tracking of electrophysiology, which may be applied over the course of a novice meditator’s practice as they journey to a higher level of expertise. Further, alternative forms of meditation, such as mindfulness, may be similarly assessed to determine the varying effects of meditation practice on brain state and trait changes. The work presented here focuses on hippocampal and amygdala electrophysiology, as these are the typical targets of RNS implantation for epilepsy. Complementary studies on intracranial EEG recordings in patients implanted with a wider range of contact locations for seizure localization (sEEG) may provide additional insight into the circuit-level activity changes associated with meditation. We feel both approaches are complementary and will improve future investigations via a feedback loop. There thus remains an exciting opportunity for exploration on how meditation practice exerts acute state changes and long-term trait changes through modulation of intracranial electrophysiology in the human brain.

**Summary.** Our findings reveal increased hippocampal and amygdala  $\gamma$  power associated with LKM, and amygdala-specific increases in  $\gamma$  oscillation duration. Further, we identified a global decrease in amygdala and hippocampus  $\beta$  oscillatory duration during LKM. These findings provide anatomically localized and neurophysiologically detailed for the role of these regions in meditation and suggest an association with their known roles in memory and emotional regulation processes. In addition, they confirm the potential of first-time LKM practice to induce transient physiological alterations in brain regions whose maladaptive functioning is implicated in mental health disorders. By identifying physiological mechanisms that are amenable to noninvasive neuromodulation via LKM, we highlight LKM’s potential as a targeted therapeutic intervention.

## Materials and Methods

**RNS Participants.** Eight DRE patients participated in the present study ( $n = 3$  female, mean age =  $44.75 \pm 7.67$  y). All participants had been previously implanted with the NeuroPace RNS System to treat DRE. Electrode placement was solely determined by clinical criterion. A structured clinical interview was conducted based on the NIH Common Data Elements Battery for Epilepsy Patients prior to participation. Participants provided informed consent for participating in the present study which was approved by the Mount Sinai Institutional Review Board.

**RNS Data Acquisition.** RNS is an FDA-approved chronically implanted invasive intervention to treat patients with DRE by continuously monitoring for abnormal electrical activity and delivering electrical stimulation to circumvent seizure onset (Fig. 1A). RNS devices provide access to intracranial LFP activity from the stimulating electrode in postsurgical, chronic conditions, which provides an opportunity to record intracranial LFP data from deep brain regions. iEEG storage in the RNS Neurostimulator was manually triggered by the experimenter to mark the start and end of the experimental session. During the experimental session, the RNS Neurostimulator continuously recorded iEEG activity in two (unilateral

implantation) or four (bilateral implantation) bipolar channels depending on each patient’s electrode placement. Bipolar derivations were carried out between the two most anterior and the two most posterior channels, typically implanted along the anterior–posterior axis of the amygdala–hippocampus (Fig. 2A). iEEG activity was continuously recorded from all available bipolar channels across two depth electrode leads in each patient at a sampling frequency of 250 Hz.

**Behavioral Task.** The experimental session consisted of a 15-min audio-guided meditation paradigm. All experimental sessions occurred in Mount Sinai’s Quantitative Biometrics Laboratory (Q-Lab) which was designed to emulate Japanese teahouses and gardens, affording patients a restorative environment for participation (Fig. 1B). The present study’s LKM paradigm was an audio-only prerecording to ensure standardization across participants. The recording utilized material from the Healthy Minds Program and Mindfulness-Based Stress Reduction course through Palouse Mindfulness. Both programs provide open-source, empirically evidenced LKM resources (10). The recording was performed by a neuropsychologist with specialist training Acceptance and Commitment Therapy, a psychological intervention that utilizes mindfulness, and experience in facilitating meditation sessions for epilepsy patients. An experimenter was present for each experimental session to monitor potential seizure activity. The meditation paradigm began with 5 min of audio-guided instruction during which participants passively listened to discussion of meditation’s objective and were guided in best practices for engaging effectively in LKM. This period was considered baseline in further analysis. Instruction was followed by 10 min of LKM during which participants were guided to wish well to oneself and then others by first imagining that “you are filled and overflowing with warmth and love” and then repeating phrases silently, such as “may you live with ease, may you be happy, may you be free from pain.” The meditation builds from receiving loving kindness directed at oneself followed by sending loving kindness to loved ones, neutral people, and all living beings (see *SI Appendix* for complete transcript). Through active attention regulation, practitioners to exert top-down attention on a target object while disengaging attentional resources from irrelevant objects (66). The goal-directed modulation of attention and its neural underpinnings has been well characterized in humans (67). Until now, significant technical limitations have prevented the direct recording of neural activity with the spatial and temporal resolution necessary to potentially detect neurophysiological fluctuations induced by meditative states, especially in novice meditators. RNS patients present a unique opportunity to record intracranial electrophysiology (iEEG) in freely behaving, postoperative humans. For these reasons, we chose to focus the present iEEG study on LKM. Participants remained seated for the duration of the experimental session. For analysis purposes, the first 2 min of instruction were selected as the baseline epoch, and 2 min in the middle of meditation during each patient’s experimental session were selected as the meditation epoch for comparison. These conditions were referred to as baseline and meditation, respectively.

**Analysis.** Analyses were performed with custom MATLAB (electrode localization), Python (iEEG preprocessing/quantification, visualization, and statistics), and R (statistics and visualization).

**Electrode Localization and iEEG Preprocessing.** Electrode localization was performed with MATLAB LeGUI. We coregistered each participant’s postoperative CT image to their preoperative whole brain MRI. All electrodes determined to be in the amygdala or hippocampus following localization were included in further analysis (amygdala: electrode, bipolar channels; hippocampus: electrodes, bipolar channels). The anatomical location of each electrode was determined using the Yale Brain Atlas (68), a whole-brain atlas of the human cortex, hippocampus, amygdala created using 3,866 electrodes across 25 iEEG subjects, and verified through visual examination. Epileptic or noisy activity was manually removed following visual inspection of each channel similar to previously published methods (32). Similar to previous results (29, 30, 32), we excluded ~6% of the data from further analyses due to the presence of epileptic or noisy activity. The proportion of excluded data across amygdala and hippocampus channels was not significantly different between conditions ( $P > 0.05$  for all channels, Pearson’s chi-square goodness-of-fit test).

**Quantifying Oscillatory Prevalence, Power, and Duration.** We computed the power spectral density (PSD) of the iEEG signal between 0 and 125 Hz using the Welch method (2 s-segments, 12.4% overlap).

**Fitting Oscillations and 1/f (FOOF) Method.** PSDs for each channel across conditions were analyzed using the FOOF method (23) which parameterizes aperiodic and periodic features of the power spectrum allowing us to determine whether FA meditation induces true oscillatory changes in neural activity or impacts other neurophysiological processes. Consistent with prior work leveraging the FOOF method to assess human electrophysiological data, we visually inspected each PSD to determine the appropriate method of aperiodic fitting (linear versus nonlinear). Across most channels ( $n = 26$ ), PSDs depicted a bend in higher frequencies ( $>40$  Hz), therefore spectral parameterization with FOOF was performed using a knee to appropriately capture these power spectra's nonlinear dynamics (Fig. 2B). This was consistent with existing human electrophysiology literature, especially when evaluating higher frequency domains ( $>40$  Hz). In one channel, a bend was not observed in the PSD. Spectral parameterization for this channel was performed without a knee. Using fitted exponent and knee parameters for each channel, we calculated the knee frequency to evaluate group differences. The knee frequency reflects the frequency in which the aperiodic slope of the PSD changes. Further, peak width was restricted to 1 to 8 Hz to minimize overfitting. Model fit was evaluated according to R-squared and error metrics (Amygdala: median  $R^2$ (median error) = baseline (0.99(0.03)), meditation (0.99(0.04)); Hippocampus: baseline (0.99(0.03)), meditation (0.99(0.03))). R-squared for both regions was in accordance with existing literature (68) and not significantly different between conditions suggesting good fit for both the amygdala ( $P > 0.05$ , two-sided paired-samples Wilcoxon signed rank test) and hippocampus ( $P > 0.05$ , two-sided paired-samples Wilcoxon signed rank test).

Using the FOOF method, we fit the aperiodic (offset, knee, exponent) and periodic (power, bandwidth, center frequency) components of each iEEG channel's power spectrum between 2 and 55 Hz for each condition (Fig. 2B). This allowed us to separate true oscillatory features of the PSD versus background (1/f) neural activity, characterized by the aperiodic components (69). The aperiodic offset, or broadband intercept, reflects the up/down translation of the spectrum while the aperiodic exponent reflects the overall curve of the aperiodic component with a smaller exponent reflecting a shallower power spectrum (Fig. 2B). Oscillations are reflected as narrowband peaks in power above the aperiodic component of the PSD. Parameterizing both the periodic and aperiodic components of the PSD allows us to compare differences in the presence/absence of oscillations at individual frequency bands between regions and conditions, as well as the periodic features of those oscillations while accounting for the background 1/f aperiodic profile which reflects other physiological processes. Periodic components of the power spectra refer to activity with a characteristic frequency. Of interest for the present study is frequency-specific power, which quantifies the amount of energy contained in the iEEG signal at a particular frequency band. Aperiodic components of the power spectra refer to recorded activity with no characteristic frequency. We assessed LKM-associated differences in aperiodic offset, which measures the overall up/down translation of the power spectra, exponent, which reflects the slope of the aperiodic component of the power spectra, and the knee frequency, which quantifies the aperiodic component's "bend" or transition point wherein the dominant characteristics of the neural signal change (Fig. 2B; 23). For each channel, we estimated power in putative frequency bands by taking the average power of peaks above the aperiodic component falling within a priori designated ranges according to their center frequency ( $\delta = 2-4$  Hz,  $\theta = 4-8$  Hz,  $\alpha = 8-13$  Hz,  $\beta = 13-30$  Hz,  $\gamma = 30-55$  Hz).

**Quantifying Subject-Level Beta and Gamma Power Modulations.** Data from each condition (baseline and LKM) were segmented into 1-s non-overlapping epochs. Time-frequency representations of power were computed for each epoch using Morlet wavelet convolution. For each frequency (in a logarithmically spaced range between 1 and 60 Hz), the number of cycles used in the wavelet transformation was defined as half of the frequency value. We extracted power estimates

corresponding to the beta (13 to 30 Hz) and gamma (30 to 55 Hz) frequency range separately for each frequency band. Power values within each range were averaged across the frequency dimension for each epoch and channel, providing a measure of broadband beta and gamma power, respectively, for further analysis. To assess changes in beta power during meditation relative to baseline, z-scores were computed for the LKM epochs. First, the mean and SD of baseline power values were calculated across epochs and time points for each channel and used to z-score each LKM epoch across channels for each subject, yielding normalized values that reflect relative deviations from baseline (Fig. 6). For each subject, the median z-scored beta power was computed across all epochs and channels.

**eBOSC Method.** To assess FA meditation-induced temporal differences in oscillatory activity, we estimated the duration of oscillations by applying the eBOSC method (24, 70). By calculating both the strength (i.e., amplitude) and duration of a given oscillation, eBOSC provides a comprehensive characterization of oscillatory events. eBOSC allowed us to estimate the duration of oscillations by identifying significant oscillations that surpass a statistical power threshold, extracting the number of detected cycles in each significant oscillation, and converting the number of detected cycles to duration in a frequency-specific manner. eBOSC first performs time-frequency decomposition on continuous data. The resulting power spectra are fit linearly using robust regression to estimate the relationship between  $\log(\text{frequency})$  and  $\log(\text{power})$ . A power threshold for each frequency was set at the 95% of a  $\chi^2$ -distribution of power values centered around the fitted estimate of background power at a given frequency based on the linear model fit. This establishes a statistical power threshold for determining a significant oscillation at a given frequency. Using the derived power threshold, oscillations and their associated time intervals were extracted. The duration threshold was set to zero, allowing us to compare both transient and sustained oscillatory episodes. The proportion of time spent oscillating at each frequency is calculated by summing detected time intervals and normalizing by total duration of a given iEEG recording.

#### Statistics.

**Data quality checks.** We used a nonparametric Pearson's chi-square goodness-of-fit test to determine whether there was a significant difference in the proportion of data within each condition following preprocessing. We used a nonparametric Wilcoxon rank-sum test to check whether FOOF model fit ( $R^2$ ) differed significantly between conditions.

**Hypothesis testing.** We used a nonparametric Fisher's exact test to assess whether a nonrandom association exists between experimental condition (baseline versus meditation) and categorical variables (presence versus absence of an oscillation within a given frequency band). To assess significant differences between conditions for linear variables (oscillatory power and duration), we used nonparametric Wilcoxon rank-sum tests. One-sided Wilcoxon rank-sum tests were implemented where applicable based on empirical literature regarding meditation-induced modulation of  $\beta$  and  $\gamma$  oscillations (3, 16, 17).

**Data, Materials, and Software Availability.** Anonymized data from participants who consented to data sharing are available upon request with the approval of NeuroPace. Analysis scripts can be found at the following GitHub repository: [https://github.com/christinamaher/iEEG\\_LKM](https://github.com/christinamaher/iEEG_LKM) (71).

**ACKNOWLEDGMENTS.** We thank Jill Gregory, Master of Fine Arts, Certified Medical Interpreter, Associate Director of Instructional Technology at the Icahn School of Medicine at Mount Sinai, for illustrating Fig. 1A. We would like to thank the patient and research and surgical staff at the recording site for their support and dedication. This project was supported by a Nash Family Research Scholars Award to J.Y., A.W., H.S.M., and F.P.

1. R. Chambers, E. Gullone, N. B. Allen, Mindful emotion regulation: An integrative review. *Clin. Psychol. Rev.* **29**, 560-572 (2009).
2. A. Lutz, J. Brefczynski-Lewis, T. Johnstone, R. J. Davidson, Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise. *PLoS One* **3**, e1897 (2008).
3. A. Lutz, L. L. Greischar, N. B. Rawlings, M. Ricard, R. J. Davidson, Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16369-16373 (2004).
4. A. Lutz, H. A. Slagter, J. D. Dunne, R. J. Davidson, Attention regulation and monitoring in meditation. *Trends Cogn. Sci.* **12**, 163-169 (2008).

5. J. A. Brefczynski-Lewis *et al.*, Neural correlates of attentional expertise in long-term meditation practitioners. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11483-11488 (2007).
6. S. L. Shapiro, G. E. Schwartz, G. Bonner, Effects of mindfulness-based stress reduction on medical and premedical students. *J. Behav. Med.* **21**, 581-599 (1998).
7. J. Wielgosz *et al.*, Mindfulness meditation and psychopathology. *Annu. Rev. Clin. Psychol.* **15**, 285-316 (2019).
8. M. Boccia, L. Piccardi, P. Guariglia, The meditative mind: A comprehensive meta-analysis of MRI studies. *Biomed Res. Int.* **2015**, 419808 (2015).
9. C. Kaur, P. Singh, EEG derived neuronal dynamics during meditation: Progress and challenges. *Adv. Prev. Med.* **2015**, 614723 (2015).

10. C. J. Dahl, A. Lutz, R. J. Davidson, Reconstructing and deconstructing the self: Cognitive mechanisms in meditation practice. *Trends Cogn. Sci.* **19**, 515–523 (2015).
11. C. Braboszcz *et al.*, Increased gamma brainwave amplitude compared to control in three different meditation traditions. *PLoS One* **12**, e0170647 (2017).
12. J. Petrovic *et al.*, The effects of loving-kindness interventions on positive and negative mental health outcomes: A systematic review and meta-analysis. *Clin. Psychol. Rev.* **110**, 102433 (2024).
13. G. Desbordes *et al.*, Effects of mindful-attention and compassion meditation training on amygdala response to emotional stimuli in an ordinary, non-meditative state. *Front. Hum. Neurosci.* **6**, 292 (2012).
14. M.-K. Leung *et al.*, Increased gray matter volume in the right angular and posterior parahippocampal gyri in loving-kindness meditators. *Soc. Cogn. Affect. Neurosci.* **8**, 34–39 (2013).
15. J. Lachaux *et al.*, Relationship between task-related gamma oscillations and BOLD signal: New insights from combined fMRI and intracranial EEG. *Hum. Brain Mapp.* **28**, 1368–1375 (2007).
16. M. Irmischer *et al.*, Controlling the temporal structure of brain oscillations by focused attention meditation. *Hum. Brain Mapp.* **39**, 1825–1838 (2018).
17. M. Irmischer *et al.*, Strong long-range temporal correlations of beta/gamma oscillations are associated with poor sustained visual attention performance. *Eur. J. Neurosci.* **48**, 2674–2683 (2018).
18. E. K. Miller, T. J. Buschman, Cortical circuits for the control of attention. *Curr. Opin. Neurobiol.* **23**, 216–222 (2013).
19. T. J. Buschman, E. K. Miller, Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* **315**, 1860–1862 (2007).
20. S. Alagapan *et al.*, Cingulate dynamics track depression recovery with deep brain stimulation. *Nature* **622**, 130–138 (2023).
21. O. G. Sani *et al.*, Mood variations decoded from multi-site intracranial human brain activity. *Nat. Biotechnol.* **36**, 954–961 (2018).
22. J. J. Young *et al.*, Elevated phase amplitude coupling as a depression biomarker in epilepsy. *Epilepsy Behav.* **152**, 109659 (2024).
23. T. Donoghue *et al.*, Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **23**, 1655–1665 (2020).
24. J. Q. Kosciessa *et al.*, Single-trial characterization of neural rhythms: Potential and challenges. *Neuroimage* **206**, 116331 (2020).
25. N. von Ellenrieder *et al.*, Electrode and brain modeling in stereo-EEG. *Clin. Neurophysiol.* **123**, 1745–1754 (2012).
26. S. Arcot Desai *et al.*, Quantitative electrocorticographic biomarkers of clinical outcomes in mesial temporal lobe epileptic patients treated with the RNS® system. *Clin. Neurophysiol.* **130**, 1364–1374 (2019).
27. S. E. Qasim *et al.*, Neuronal activity in the human amygdala and hippocampus enhances emotional memory encoding. *Nat. Hum. Behav.* **7**, 754–764 (2023).
28. J. Parvizi, S. Kastner, Human intracranial EEG: Promises and limitations. *Nat. Neurosci.* **21**, 474–483 (2018).
29. S. L. L. Maoz *et al.*, Dynamic neural representations of memory and space during human ambulatory navigation. *Nat. Commun.* **14**, 6643 (2023).
30. Z. M. Aghajan *et al.*, Theta oscillations in the human medial temporal lobe during real-world ambulatory movement. *Curr. Biol.* **27**, 3743–3751.e3 (2017).
31. C. B. Josephson *et al.*, Association of depression and treated depression with epilepsy and seizure outcomes: A multicohort analysis. *JAMA Neurol.* **74**, 533–539 (2017).
32. I. Saez *et al.*, Encoding of multiple reward-related computations in transient and sustained high-frequency activity in human OFC. *Curr. Biol.* **28**, 2889–2899.e3 (2018).
33. L. Marcuse, M. Fields, J. Yoo, *Rowan's Primer of EEG* (Elsevier, ed. 2, 2015).
34. T. A. Whitten *et al.*, A better oscillation detection method robustly extracts EEG rhythms across brain state changes: The human alpha rhythm as a test case. *Neuroimage* **54**, 860–874 (2011).
35. A. Wynne, *The Origin of Buddhist Meditation* (Routledge, 2007).
36. A. Manna *et al.*, Neural correlates of focused attention and cognitive monitoring in meditation. *Brain Res. Bull.* **82**, 46–56 (2010).
37. M. Engström *et al.*, Functional magnetic resonance imaging of hippocampal activation during silent mantra meditation. *J. Altern. Complement. Med.* **16**, 1253–1258 (2010).
38. B. L. Fredrickson *et al.*, Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. *J. Pers. Soc. Psychol.* **95**, 1045–1062 (2008).
39. F. Ferrarelli *et al.*, Experienced mindfulness meditators exhibit higher parietal-occipital EEG gamma activity during NREM sleep. *PLoS One* **8**, e73417 (2013).
40. K. A. Garrison *et al.*, Meditation leads to reduced default mode network activity beyond an active task. *Cogn. Affect. Behav. Neurosci.* **15**, 712–720 (2015).
41. D. Roberts-Wolfe *et al.*, Mindfulness training alters emotional memory recall compared to active controls: Support for an emotional information processing model of mindfulness. *Front. Hum. Neurosci.* **6**, 15 (2012).
42. C. C. van Schie *et al.*, When I relive a positive me: Vivid autobiographical memories facilitate autoneurotic brain activation and enhance mood. *Hum. Brain Mapp.* **40**, 4859–4871 (2019).
43. C. Poskanzer, M. Aly, Switching between external and internal attention in hippocampal networks. *J. Neurosci.* **43**, 6538–6552 (2023).
44. Y. Zhu *et al.*, Emotion regulation of hippocampus using real-time fMRI neurofeedback in healthy human. *Front. Hum. Neurosci.* **13**, 242 (2019).
45. M. F. Carr *et al.*, Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nat. Neurosci.* **14**, 147–153 (2011).
46. B. J. Wilting *et al.*, The hippocampus plays a selective role in the retrieval of detailed context memories. *Curr. Biol.* **20**, 1336–1344 (2010).
47. R. H. A. H. Jacobs *et al.*, The amygdala, top-down effects, and selective attention to features. *Neurosci. Biobehav. Rev.* **36**, 2069–2084 (2012).
48. C. Klinge *et al.*, Increased amygdala activation to emotional auditory stimuli in the blind. *Brain* **133**, 1729–1736 (2010).
49. C. J. Peck, C. D. Salzman, Amygdala neural activity reflects spatial attention towards stimuli promising reward or threatening punishment. *Elife* **3**, e04478 (2014).
50. L. Pessoa, R. Adolphs, Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nat. Rev. Neurosci.* **11**, 773–783 (2010).
51. M. Lundqvist *et al.*, Beta: Bursts of cognition. *Trends Cogn. Sci.* **28**, 662–676 (2024).
52. M. Bocchio *et al.*, Synaptic plasticity, engrams, and network oscillations in amygdala circuits for storage and retrieval of emotional memories. *Neuron* **94**, 731–743 (2017).
53. Y. Li *et al.*, Beta oscillations in major depression—Signalling a new cortical circuit for central executive function. *Sci. Rep.* **7**, 18021 (2017).
54. S. L. Brincat *et al.*, Interhemispheric transfer of working memories. *Neuron* **109**, 1055–1066.e4 (2021).
55. S. Kim *et al.*, Alteration of cortical functional networks in mood disorders with resting-state electroencephalography. *Sci. Rep.* **12**, 5920 (2022).
56. J. N. Sanes, J. P. Donoghue, Oscillations in local field potentials of the primate motor cortex during voluntary movement. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4470–4474 (1993).
57. F. Torrecillos *et al.*, Distinct modulations in sensorimotor postmovement and foreperiod  $\beta$ -band activities related to error salience processing and sensorimotor adaptation. *J. Neurosci.* **35**, 12753–12765 (2015).
58. P. Khanna, J. M. Carmena, Beta band oscillations in motor cortex reflect neural population signals that delay movement onset. *Elife* **6**, e24573 (2017).
59. H. T. Darch *et al.*, Pre-movement changes in sensorimotor beta oscillations predict motor adaptation drive. *Sci. Rep.* **10**, 17946 (2020).
60. S. Meisenhelter *et al.*, Cognitive tasks and human ambulatory electrocorticography using the RNS system. *J. Neurosci. Methods* **311**, 408–417 (2019).
61. V. R. Rao *et al.*, Chronic ambulatory electrocorticography from human speech cortex. *Neuroimage* **153**, 273–282 (2017).
62. E. L. Rich, J. D. Wallis, Spatiotemporal dynamics of information encoding revealed in orbitofrontal high-gamma. *Nat. Commun.* **8**, 1139 (2017).
63. P. Fries, Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* **32**, 209–224 (2009).
64. P. Fries *et al.*, The gamma cycle. *Trends Neurosci.* **30**, 309–316 (2007).
65. M. Seeber *et al.*, Human neural dynamics of real-world and imagined navigation. bioRxiv [Preprint] (2024). <https://doi.org/10.1101/2024.05.23.595237> (Accessed 23 May 2024).
66. C. J. May *et al.*, Short-term training in loving-kindness meditation produces a state, but not a trait, alteration of attention. *Mindfulness* **2**, 143–153 (2011).
67. Y. C. Leong *et al.*, Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron* **93**, 451–463 (2017).
68. H. McGrath *et al.*, High-resolution cortical parcellation based on conserved brain landmarks for localization of multimodal data to the nearest centimeter. *Sci. Rep.* **12**, 18778 (2022).
69. M. Kopčanová *et al.*, Resting-state EEG signatures of Alzheimer's disease are driven by periodic but not aperiodic changes. *Neurobiol. Dis.* **190**, 106380 (2024).
70. L. Hu *et al.*, Single-trial time-frequency analysis of electrocortical signals: Baseline correction and beyond. *Neuroimage* **84**, 876–887 (2014).
71. C. Maher, iEEG\_LKM. GitHub. [https://github.com/christinamaher/iEEG\\_LKM](https://github.com/christinamaher/iEEG_LKM). Accessed 10 January 2025.

---

# Decoding Latent Attention Across Cognitive Models

---

**Christina Maher**

Friedman Brain Institute  
Icahn School of Medicine at Mount Sinai  
New York, NY 10029  
christina.maher@icahn.mssm.edu

**Ignacio Saez**

Depts. Neuroscience, Neurology, Neurosurgery  
Icahn School of Medicine at Mount Sinai  
New York, NY 10029  
Ignacio.saez@mssm.edu

**Angela Radulescu**

Department of Psychiatry  
Icahn School of Medicine at Mount Sinai  
New York, NY 10029  
angela.radulescu@mssm.edu

## Abstract

Selective attention plays a critical role in representation learning [1], with two competing computational cognitive models proposing distinct mechanisms by which attention arises. Feature-based reinforcement learning (FRL) posits selective attention arises through retrospective value learning, while serial hypothesis testing (SHT) suggests selective attention arises via prospective hypothesis testing. Here, we apply LaseNet [2], a novel neural network method for directly inferring latent variables from cognitive models with both synthetic and human data, to decode trial-by-trial attention as agents learn in a multidimensional environment. Networks trained on data generated from SHT models outperformed networks trained on data from FRL models in predicting attention in a labeled human dataset. SHT networks also showed limited generalizability to unseen data across model classes, reflecting distinct mechanisms for attention allocation under FRL and SHT models. This work utilizes a cutting-edge approach to infer attention dynamics from cognitive models, significantly enhancing their evaluation and providing deeper insights into the attention mechanisms that drive human representation learning. By leveraging this method, we can uncover latent attention processes underlying human representation learning, ultimately informing model-based neural analyses.

**Keywords:** Representation learning, multi-dimensional reinforcement learning, selective attention, serial hypothesis testing, LaseNet

## 1 Introduction

Previous research highlights the role of selective attention in state representation learning [1]. Two main classes of computational cognitive models, each supported by choice data and neural evidence, propose competing mechanisms for implementing selective attention in multidimensional environments: feature-based reinforcement learning (FRL) models suggest that selective attention emerges retrospectively through value learning of stimulus features [3]. In contrast, serial hypothesis testing (SHT) models propose a prospective process, where selective attention is dynamically allocated by iteratively testing hypotheses about which features of the task are relevant [4], [5]. Hybrid models that incorporate elements of both FRL and SHT have also been proposed [6].

Traditional model evaluation methods based on maximum likelihood estimation (MLE) are limited to models for which analytically tractable likelihoods can be derived and computed [7]. SHT models, and sampling models more generally, are challenging analytically, and typically intractable due to combinatorial explosions. This has constrained cognitive models of attention to be biased towards FRL models. However, mounting evidence suggests that hypothesis sampling plays a key role in attention allocation [4], [5], [6]. Here, we use LaseNet, a novel method for direct inference of latent cognitive variables [2], to directly decode attention generated from different models in the FRL and SHT family. Our goal is to arbitrate between these model classes, and to generate precise predictions of attention allocation for model-based neural analysis [8].

We hypothesized that: (1) networks trained on SHT models would label human participants' trial-by-trial hypotheses more accurately than those trained on FRL models, and (2) networks trained using a given class of models would not generalize well to data generated by models from a different model class. This would indicate that the two model classes capture distinct cognitive processes which the network is able to discern. To test these hypotheses, we trained LaseNet neural estimators to infer trial-by-trial attention allocation (i.e., their hypothesis about the most relevant feature) using synthetic data generated from each cognitive model. We evaluated the networks on held-out synthetic test data, and on a self-labeled human-generated dataset. All networks performed above chance, with SHT models outperforming FRL models in labeling the human dataset. Additionally, networks did not generalize well across model classes. Findings highlight the distinct mechanisms underlying representation learning proposed by FRL and SHT and provide a foundation for adjudicating between their competing theories.

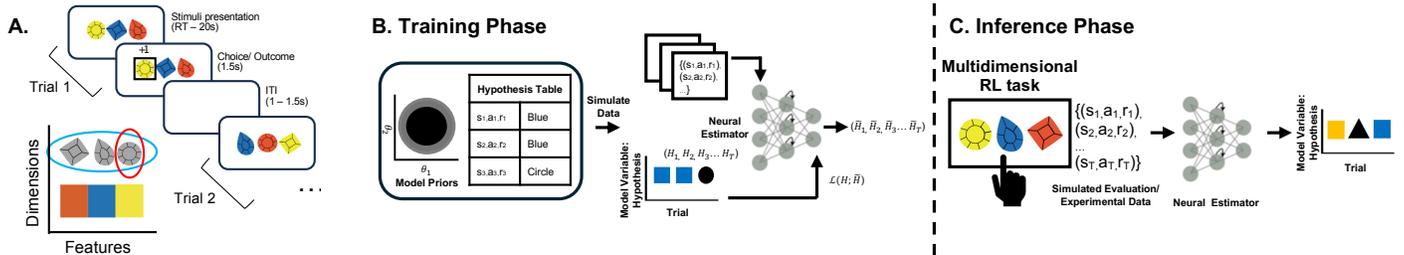
## 2 Task

To train neural networks, we used a multidimensional RL task adapted from prior work [3], [9]. In the human experiment, participants ( $N=21$  neurosurgical patients) completed six 18-trial games in which they made repeated choices between three stimuli varying in shape (square, oval, circle) and color (orange, yellow, blue) (**Fig 1A**). Each game had one relevant dimension (shape or color) and a target feature (e.g., square). After each choice, participants were rewarded with 80% probability if they selected the stimulus containing the target feature. To maximize reward, participants had to learn the target feature via trial-and-error. Changes in the relevant dimension were explicitly signaled between games. And participants were aware of the generative structure of the task (i.e. one target feature being relevant, and the exact reward probabilities).

## 3 LaseNet Estimators

We decoded trial-by-trial attention in the multidimensional RL task using five cognitive models. We used LaseNet [2], a novel technique based on neural Bayes estimation which directly maps choice data to latent variable space by using recurrent neural networks trained on synthetic data generated by a cognitive model. Unlike traditional maximum likelihood estimation, which requires predefined parameter space comparisons, LaseNet directly infers latent variables from behavior, enabling the use of models with analytically intractable or computationally intensive likelihoods. During the training phase (**Fig. 1B**, adapted from [2]), we create a synthetic dataset by simulating the desired cognitive model. LaseNet is trained using model-simulated observable data  $Y$  as input and a series of model-derived latent variables  $Z$  as output. During the inference phase (**Fig. 1C**), the trained LaseNet takes the observable experimental data as input to infer a sequence of unobservable latent variables. In this project, our goal was to infer participants' attention allocation by decoding their trial-by-trial hypotheses  $H$  about the target feature. For each LaseNet estimator, we simulated 20000  $(Z, Y)$  pairs with 720 trials in each pair as training data. In effect, this means that each individual

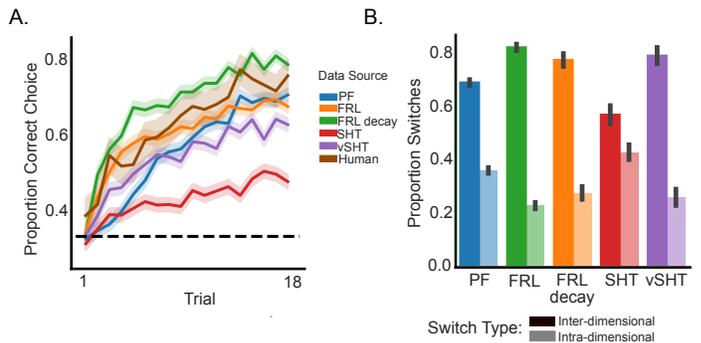
parameter setting (“agent”) generated 40 games, each representing an instance of dynamic attention allocation with that parameter setting. While FRL models would produce fixed attention trajectories with the same parameter setting, SHT models change attention stochastically, so each game represents a possible trajectory through hypothesis space. We simulated an additional unseen 20 ( $Z, Y$ ) pairs with 720 trials each as testing data. Each pair was generated with uniform priors on model parameters  $\theta$ . Although the hypotheses participants are testing are latent in an experimental context, we collected one exploratory dataset (6 games, 18 trials/game, 108 trials in total) in which the participant was instructed to self-label their hypothesis on each trial. Although we have access to these self-reported hypotheses, we do not have access to the generative model that produced their choices. This discrepancy allows us to assess the ability of LaseNet Estimator trained on different cognitive models to infer the participants’ latent hypotheses.



**Fig 1. LaseNet method applied to multidimensional RL.** **A.** Multidimensional RL task. **B.** Network is trained to predict latent variables from a cognitive model (i.e., target feature hypothesis) using simulated data. Input includes trial-wise observable data (stimuli, actions, rewards). **C.** Trained networks predict latent variable for experimental data with unknown ground truth. Schematic adapted from [2].

## 4 Cognitive Models

Two cognitive model classes have emerged to capture attention during state representation learning: FRL [3], [9] and SHT [4], [5], [6]. Models within these respective classes seek to explain how attentional mechanisms support efficient state representation learning by prioritizing relevant information. FRL models propose agents learn to assign values to individual features of stimuli, which are then integrated to guide their choices. Attention is dynamically allocated to specific features based on their perceived value for maximizing future rewards. FRL is an extension of traditional RL, where the focus shifts from learning the value of entire stimuli to learning the value of discrete features within those stimuli. In this context, value learning over time drives attentional allocation towards the features that most reliably predict reward. We trained LaseNet Estimators to decode latent hypotheses using two FRL models: FRL and FRL with decay [9]. The FRL model maintains that participants learn and update values for each feature in the environment. To account for human’s limited working memory capacity, the FRL with decay model decays the value of nonchosen features. At each timestep, the hypothesis  $H_t$  is taken to be the feature with the highest value.



**Fig 2. Cognitive Model Validation.** **A.** Proportion of correct choice increases across trials in simulated ( $N=20$ ; 40 games, 18 trial/game) and human ( $N=21$ ; 6 games, 18 trials/game; shading = SEM; dashed line = chance) data. **B.** Proportion of inter- vs intra-dimensional hypothesis switches by model ( $N=20$  simulated agents, 40 games, 18 trials/game; error bars = SEM).

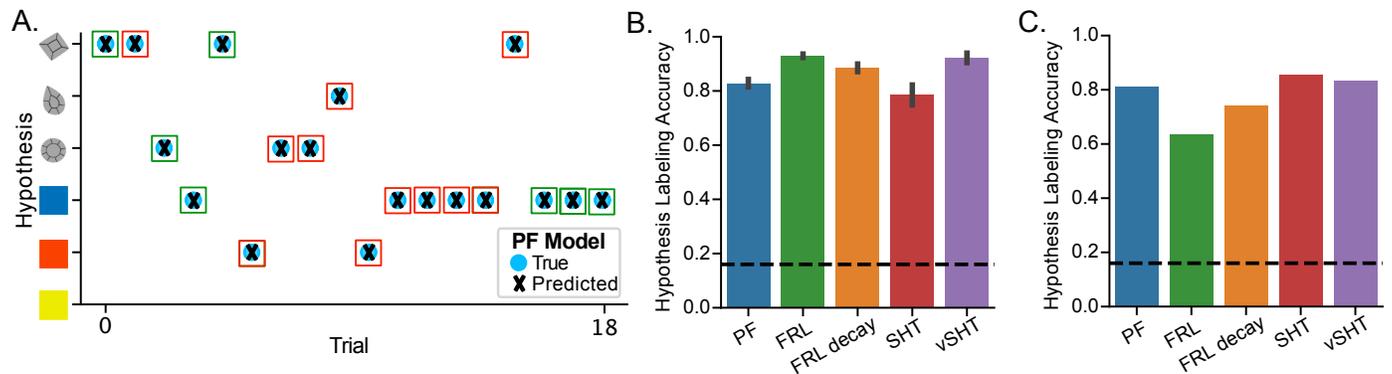
In contrast, SHT models suggest that efficient state representations arise through the evaluation of competing hypotheses regarding the most relevant environmental features. These models can be thought of as a tractable and computationally efficient approximations of full Bayesian inference. According to this framework, representation learning involves maintaining a single hypothesis  $H_t$  about the relevant feature and iteratively updating it as new information is received. Attention is directed toward the feature that the agent hypothesizes is most rewarding. To decode latent hypotheses, we trained LaseNet Estimators using three variants of hypothesis testing models: Serial Hypothesis Testing (SHT) [5], Value-based Serial Hypothesis Testing (vSHT) [6], and a Memory-Augmented Particle Filter (PF) [4].

## 5 Results

Before training LaseNet, we validated the cognitive models by comparing the performance of synthetic test data to human data on the same task. All models successfully learned the task (Fig. 2A) and exhibited attention-switching behavior indicative of adaptive learning, with more inter-dimensional switching (Fig. 2B).

Next, we trained five separate LaseNet Estimators for each of our cognitive models. To evaluate training, we assessed each network's ability to label hypotheses using a held-out test set ( $N = 20$  agents; Fig. 3A). All networks performed above chance (Fig 3B). Then, we evaluated each network's ability to label trial-by-trial hypotheses in this self-labeled human dataset, in which the generative cognitive model is unknown. As we hypothesized, the SHT models performed better than the FRL models, indicating that the behavior in this dataset was more accurately captured by this model class (Fig 3C).

Finally, we assessed the generalizability of the PF- and vSHT-trained networks to determine whether they are more effective at identifying hypothesis testing behavior specific to their own generative models. To do this, we evaluated these networks' ability to label hypotheses using the synthetic test datasets generated by each cognitive model (Fig. 4). As hypothesized, model type had a significant effect on the network's labeling accuracy (PF:  $F=7.2$ ,  $p<0.0001$ , Fig 4A; vSHT:  $F=17.2$ ,  $p<0.0001$ , Fig 4B). Both PF and vSHT networks performed best when labeling test data generated by the model it was trained on, with stronger performance within its model class and reduced accuracy when generalizing to FRL models. These results suggest that the networks capture model-specific cognitive mechanisms that are not generalizable across model classes.



**Fig 3. Evaluation of LaseNet Estimators on simulated and human data.** **A.** PF-trained network's hypothesis labeling accuracy for one example game in which it correctly labeled every trial (blue circle = true hypothesis, black cross = predicted hypothesis, green = reward, red = no reward; target feature = yellow). **B.** Hypothesis labeling accuracy (represented by the proportion of correctly labeled trials, error bars = SEM) for LaseNet Estimators evaluated on synthetic test data ( $N = 20$  agents; 40 games, 18 trials per game) for each cognitive model. All networks perform well above chance (dashed line). **C.** Hypothesis labeling accuracy (represented by the proportion of correctly labeled trials) for LaseNet Estimators evaluated on self-labeled human dataset ( $N = 1$  participant; 6 games, 18 trials per game). All models perform significantly above chance (dashed line), however the hypothesis testing models (PF, SHT, vSHT) outperform the FRL models (FRL, FRL decay).

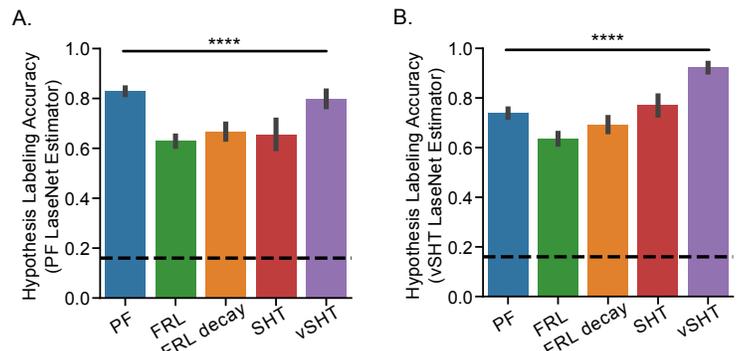
## 6 Discussion

We trained neural networks using LaseNet to infer latent hypothesis testing under two distinct cognitive model classes, FRL and SHT, which capture attentional mechanisms underlying state representation learning. This innovative approach to cognitive dynamics inference circumvents the limitations of traditional model fitting and comparison methods, enabling us to infer latent cognitive variables with high predictive accuracy, and allowing for direct comparison of inferences across different model assumptions.

Our results demonstrated that all networks performed well above chance in inferring hypotheses on simulated data (Fig. 3B). As hypothesized, the SHT models outperformed the FRL models in labeling human hypotheses from a self-labeled dataset with known ground truth (Fig. 3C). These findings suggest that the SHT models are more effective than FRL in capturing the cognitive processes underlying human state representation learning. Furthermore, the networks demonstrated low out-of-class generalizability (Fig. 4), which underscores their ability to isolate behavioral dynamics specific to each cognitive model.

Several limitations warrant consideration. One potential limitation is the use of a uniform prior during the training phase for all models, which may not accurately reflect the true parameter ranges observed in human data. This discrepancy could explain the relatively lower performance of the SHT and vSHT models (Fig. 2A). Although the uniform prior avoided potential training biases, incorporating more accurate priors, derived from previous studies, could help constrain the parameter space more effectively and improve the network's ability to capture realistic patterns in the data. Additionally, performing hyperparameter tuning on unseen simulated data prior to training could further optimize model performance.

An interesting aspect of our results lies within the SHT model class, which performed better at labeling human hypotheses (Fig. 3C) than FRL. Within this class, the SHT, vSHT, and PF models make different assumptions about how the proposal distribution is maintained. PF model captures state inference while accounting for working memory constraints, potentially making it a more biologically plausible model for human representation learning. Although we lack ground truth for participant behavior in experimental data, these competing models can be tested against neural data to identify which model's inferences is most reflective of observed neural activity. We are poised to investigate these inferences in our cohort of 21 neurosurgical participants who completed the task and have corresponding cortical/subcortical local field potential recordings. These data will help us assess the assumptions of different models and identify which most closely aligns with neuronal activity, shedding light on the biological basis of attention in state representation learning.



**Fig 4. LaseNet Estimator generalization.** **A.** PF-trained network achieves the highest hypothesis labeling accuracy on PF model data and performs better overall on SHT models compared to FRL models (N=20 agents, 40 games, 18 trials/game; error bars = SEM). **B.** vSHT-trained network achieves the highest hypothesis labeling accuracy on vSHT model data and performs better overall on SHT models compared to FRL models (N=20 agents, 40 games, 18 trials/game; error bars = SEM).

## 6 References

- [1] A. Radulescu, Y. Niv, and I. Ballard, "Holistic Reinforcement Learning: The Role of Structure and Attention," *Trends Cogn. Sci.*, vol. 23, no. 4, pp. 278–292, Apr. 2019, doi: 10.1016/j.tics.2019.01.010.
- [2] T.-F. Pan, J.-J. Li, B. Thompson, and A. Collins, "Latent Variable Sequence Identification for Cognitive Models with Neural Network Estimators," Dec. 19, 2024, *arXiv*: arXiv:2406.14742. doi: 10.48550/arXiv.2406.14742.
- [3] Y. C. Leong, A. Radulescu, R. Daniel, V. DeWoskin, and Y. Niv, "Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments," *Neuron*, vol. 93, no. 2, pp. 451–463, Jan. 2017, doi: 10.1016/j.neuron.2016.12.040.
- [4] A. Radulescu, Y. Niv, and N. Daw, "A particle filtering account of selective attention during learning," in *2019 Conference on Cognitive Computational Neuroscience*, Berlin, Germany: Cognitive Computational Neuroscience, 2019. doi: 10.32470/CCN.2019.1338-0.
- [5] R. C. Wilson and Y. Niv, "Inferring Relevance in a Changing World," *Front. Hum. Neurosci.*, vol. 5, 2012, doi: 10.3389/fnhum.2011.00189.
- [6] M. Song, P. A. Baah, M. B. Cai, and Y. Niv, "Humans combine value learning and hypothesis testing strategically in multi-dimensional probabilistic reward learning," *PLOS Comput. Biol.*, vol. 18, no. 11, p. e1010699, Nov. 2022, doi: 10.1371/journal.pcbi.1010699.
- [7] B. Van Opheusden, L. Acerbi, and W. J. Ma, "Unbiased and efficient log-likelihood estimation with inverse binomial sampling," *PLOS Comput. Biol.*, vol. 16, no. 12, p. e1008483, Dec. 2020, doi: 10.1371/journal.pcbi.1008483.
- [8] C. Maher, S. Qasim, L. N. Martinez, I. Saez, and A. Radulescu, "Intracranial recordings reveal neural encoding of attention-modulated reinforcement learning in humans," in *Computational Cognitive Neuroscience*, 2024. [https://2024.ccneuro.org/pdf/548\\_Paper\\_authored\\_CCN2024\\_Maheret al.pdf](https://2024.ccneuro.org/pdf/548_Paper_authored_CCN2024_Maheret al.pdf).
- [9] Y. Niv *et al.*, "Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms," *J. Neurosci.*, vol. 35, no. 21, pp. 8145–8157, May 2015, doi: 10.1523/JNEUROSCI.2978-14.2015.