Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Distinct neural representations encode psychiatric states across timescales

Nicole Provenza

Nicole.Provenza@bcm.edu

Baylor https://orcid.org/0000-0002-6952-5417

Timon Merk

Baylor College of Medicine https://orcid.org/0000-0003-3011-2612

Grace Bezold

Johns Hopkins Applied Physics Laboratory

Rahul Hingorani

Johns Hopkins Applied Physics Laboratory

Andrew Wiese

Baylor College of Medicine

Brian Robinson

Johns Hopkins Applied Physics Laboratory

Han Yi

Johns Hopkins Applied Physics Laboratory

Jordan Altman

Baylor College of Medicine https://orcid.org/0009-0006-3701-2023

Sameer Rajesh

UT Southwestern

Tori Gobo

Baylor College of Medicine

Matthew Ochoa

Baylor College of Medicine

Thomas Hamre

Baylor College of Medicine

Vinayak Belavadi

Baylor College of Medicine

Corey St. Romain

Baylor College of Medicine https://orcid.org/0000-0002-2495-6037

Michelle Avendano-Ortega

Baylor College of Medicine

Sarah Soubrah

Baylor College of Medicine
Gabriel Reyes
Baylor College of Medicine
Tomasz Fraczek
Baylor College of Medicine https://orcid.org/0009-0008-7871-8588
Eric Storch
Baylor College of Medicine
Wayne Goodman
Baylor College of Medicine
Sameer Sheth
Baylor College of Medicine https://orcid.org/0000-0001-8770-8965

Biological Sciences - Article

Keywords:

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11199139/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. S.A.S. has been a consultant for Boston Scientific, Zimmer Biomet, Koh Young, Sensoria Therapeutics, Varian Medical Systems, Abbott and Neuropace and is co-founder of Motif Neurotech. W.K.G. receives royalties from Nview, LLC and OCDscales, LLC. E.A.S. reports receiving research funding to his institution from the International OCD Foundation, Wellcome Trust, and NIH. He receives direct funding from the International OCD Foundation as well as MHNTI for providing trainings on treating obsessive-compulsive disorder with psychotherapy. Furthermore, Dr. Storch co-founded Rethinking Behavioral Health which provides training and consultation in the treatment of obsessive-compulsive disorder and related conditions. He was a consultant for Brainsway and Biohaven Pharmaceuticals in the past 36 months. He owns stock options less than $5000 in NView (for distribution of the Y-BOCS and CY-BOCS) and receives royalties from OCD Scales LLC (for distribution of the Y-BOCS and CY-BOCS). He receives book royalties from Elsevier, Wiley, Oxford, American Psychological Association, Guildford, Springer, Routledge, and Jessica Kingsley.

Distinct neural representations encode psychiatric states across timescales

Authors:
Timon Merk[1], Grace Bezold[2], Rahul Hingorani[2], Andrew D. Wiese[3], Brian Robinson[2], Han Yi[2], Jordan Altman[1], Sameer V. Rajesh[1,4], Tori Gobo[1], Matthew Ochoa[1], Thomas Hamre[1], Vinayak Belavadi[1], Corey P. St. Romain[1], Michelle Avendano-Ortega[3], Sarah Soubra[1], Gabriel Reyes[1], Tomasz M. Frączek[1], Eric A. Storch[3], Wayne K. Goodman[3], Sameer A. Sheth[1,3,5,6,7,8], Nicole R. Provenza[1,6,7,9]

[1]Department of Neurosurgery, Baylor College of Medicine, Houston, TX, USA
[2]Research and Exploratory Development Department, Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA
[3]Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA
[4]Department of Neurological Surgery, University of Texas Southwestern, Dallas, TX, USA
[5]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
[6]Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA
[7]Neuroengineering Initiative, Rice University, Houston, TX, USA
[8]Gordon and Mary Cain Pediatric Neurology Research Foundation Laboratories, Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA
[9]Department of Bioengineering, Rice University, Houston, TX, USA

Corresponding author: Nicole R. Provenza (Nicole.Provenza@bcm.edu)

Abstract
Psychiatric states manifest over multiple timescales, from transient fluctuations in disorder-related distress to long-term variations in disease severity, yet whether these expressions share common neural representations remains unknown. We addressed this question in obsessive-compulsive disorder (OCD) by comparing neural activity associated with momentary distress evoked by triggers versus that associated with clinically assessed chronic symptom severity. We analyzed intracranial recordings from the ventral capsule and orbitofrontal cortex spanning more than 4595 patient-days in 8 individuals undergoing deep brain stimulation therapy, together with quantitative facial- and speech-derived behavioral measures. Neural signatures of OCD severity were dissociable from and stronger than signatures of momentary distress, indicating distinct underlying brain states. Integrating behavioral data into a predictive computational model improved neural decoding performance of both momentary distress and OCD severity. These findings show that acute manifestations of psychiatric illness are not simply short-timescale readouts of the same neural states that encode longer-term clinical burden. Instead, transient distress and longitudinal disease severity are represented by dissociable neural processes. This distinction has important implications for how psychiatric biomarkers and adaptive neuromodulation strategies should be designed.

Introduction

Psychiatric disorders are dynamic conditions in which clinically relevant states evolve across various temporal scales. Affected individuals may experience transient fluctuations in emotion, mood, or distress that occur within the context of slower changes in overall disease severity, functional impairment, and treatment response. A fundamental unresolved question is whether transient symptom expression reflects a rapidly varying readout of the same underlying neural state that determines longer-term disease burden, or whether these clinical dimensions are encoded by separable neural processes. Resolving this distinction is important both for understanding how psychiatric illness is organized in the brain and for determining which neural states should be targeted by adaptive neuromodulation or brain-computer interface therapies.

Obsessive–compulsive disorder (OCD) provides a particularly well-defined model in which to address this question. OCD is a chronic and debilitating psychiatric condition affecting approximately 2–3% of the global population[1,2]. Although individuals with OCD commonly experience episodes of acute distress elicited by exposure to environmental or internal triggers, the overall burden of the disorder is not clinically measured by these transient experiences but by the persistent inability to control obsessions and resist compulsions in everyday life. Consequently, successful treatment reflects more than reductions in momentary distress; it restores behavioral control, reduces avoidance, and diminishes interference of OCD symptoms with daily functioning. These latter dimensions are captured by the Yale-Brown Obsessive Compulsive Scale (Y-BOCS), the gold-standard measure of OCD symptom severity used to guide clinical decision-making and evaluate treatment efficacy. Accordingly, this scale quantifies these behaviors and effects over timescales of days to weeks[4,5]. Exposure and response prevention (ERP), the most effective behavioral treatment for OCD, exemplifies this distinction by requiring patients to repeatedly experience distress while learning to refrain from compulsive behaviors, thereby reducing long-term symptom severity without necessarily eliminating acute distress during individual exposures[6–8]. These clinical dissociations suggest that acute distress may represent a transient expression occurring within a broader disease state, rather than a direct moment-to-moment proxy for that state.

Prior neurophysiological studies of OCD have predominantly focused on neural activity measured during symptom provocation or task-based paradigms lasting seconds to minutes, with the goal of identifying neural biomarkers of acute distress, obsessive thoughts, or compulsive behaviors[9–12]. More recently, studies using on-device neural recordings have begun to identify neural correlates of longitudinal OCD symptom severity during DBS therapy over long timescales[13,14], demonstrating that intracranial recordings can track disease progression over months to years of treatment. When considered in aggregate, however, this body of work has thus far failed to converge on a set of common neural biomarkers[15]. In contrast, the analogous pursuit in Parkinson's disease has consistently highlighted subcortical and cortical neural features that track cardinal symptoms and can be used for closed-loop DBS control[16].

One possible explanation for the heterogeneity in OCD is that acute distress and longitudinal disease severity represent distinct clinical constructs with different underlying neural representations. In support of this notion, clinical experience demonstrates that episodes of intense distress can occur in individuals with relatively mild OCD, whereas some individuals with severe OCD may experience periods of relatively little acute distress though remaining profoundly impaired by avoidance and diminished behavioral control. Recent advances in chronic DBS sensing provide a unique opportunity to determine whether neural representations of transient distress differ from those associated with longitudinal disease severity by measuring brain activity across the wide range of timescales over which these behaviors and symptoms evolve.

Although chronic neural recordings provide an opportunity to measure brain activity directly, psychiatric diagnosis and treatment remain fundamentally grounded in behavioral observation. Clinicians routinely integrate facial expression, speech, affect, and motor behavior when evaluating symptom severity and treatment response[17]. Advances in computational behavioral analyses now allow these observations to be quantified objectively from video and audio recordings[18–20]. Rather than serving solely as an additional source of clinical information, behavioral measurements may also guide the identification of neural features most relevant to disease state. By using these behavioral measures as supervisory signals, neural representation learning can incorporate the same types of clinically meaningful information that routinely inform psychiatric assessment. Integrating neural and behavioral measurements therefore enables more clinically informed neural decoders of transient distress and longitudinal symptom severity than either modality alone[21].

Here, we investigated whether momentary distress and longitudinal OCD symptom severity are represented by distinct neural signatures. Using chronic intracranial recordings acquired during DBS treatment, we separately characterized neural activity associated with naturalistic momentary distress during remote ERP teletherapy in patients' homes and longitudinal fluctuations in clinical OCD symptom severity measured over hundreds of days. We further investigated whether objective facial and speech-derived behavioral measures could improve neural representations and decoding of these clinically relevant states. By distinguishing neural representations of transient distress from those associated with long-term disease severity, this work establishes a framework for developing neural decoders that are explicitly aligned with the clinical state they are intended to estimate. More practically, it provides a foundation for adaptive neuromodulation strategies that target the neural processes underlying long-term clinical improvement in psychiatric disorders.

Results

Chronic intracranial recordings during naturalistic exposures and longitudinal clinical follow-up

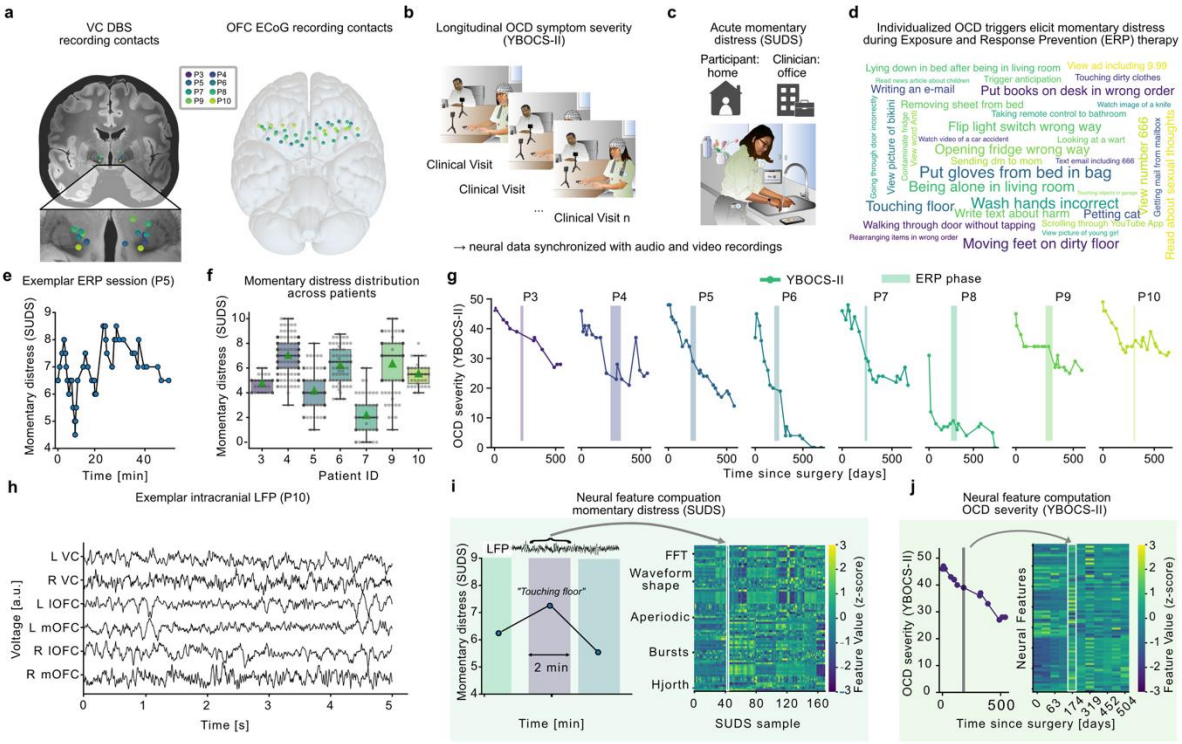


Figure 1. Experimental design for chronic intracranial recordings of longitudinal OCD symptom severity and momentary distress. (a) Bilateral ventral capsule (VC) DBS leads enabled chronic local field potential (LFP) recordings throughout an 18-month clinical trial. In five patients, bilateral orbitofrontal cortex (OFC) electrocorticography (ECoG) strip electrodes provided simultaneous cortical recordings. **(b)** OCD symptom severity was monitored longitudinally throughout the DBS clinical trial using the Yale-Brown Obsessive Compulsive Scale II (YBOCS-II). Figure adapted from[12]. **(c)** Exposure and response prevention (ERP) teletherapy sessions were conducted at patients' homes via video conference. During ERP sessions, patients were exposed to individualized OCD-related triggers and reported momentary distress. Figure adapted from[12]. **(d)** ERP sessions enabled recording of neural activity during naturalistic, patient-specific OCD exposures encountered in everyday life. **(e)** Example ERP session illustrating within-session fluctuations in momentary distress. During ERP, participants were presented with symptom-provoking exposures and reported their Subjective Units of Distress Scale (SUDS) ratings, either when requested or when they experienced a change in distress. These assessments captured fluctuations in momentary distress over minutes within a single ERP session. **(f)** Across ERP sessions, repeated SUDS assessments yielded distributions of momentary distress ratings for each participant, representing short-timescale changes in affective state in contrast to **(g)** longer-timescale changes in OCD symptom severity over many months. **(h)** Neural features were extracted from time-series LFP recordings to identify neural correlates of **(i)** momentary distress and **(j)** longitudinal OCD symptom severity.

As part of a clinical trial to identify intracranial neural biomarkers of OCD (NCT04281134), we implanted eight treatment-resistant OCD patients with DBS leads in the ventral capsule (VC) and connected the leads to an investigational recording-capable DBS device (Supplementary Table 1). In a subset of patients (n=5), we additionally implanted bilateral orbitofrontal cortex (OFC) electrocorticography (ECoG) electrodes for passive neural sensing (Figure 1a). Throughout the 18-month clinical trial, patients underwent repeated assessments of OCD symptom severity during clinic visits (22.5±1.5 assessments per patient; Figure 1b; complete trajectory of YBOCS-II scores including assessments without corresponding neural recordings is shown in Supplementary Figure 1). Patients additionally participated in ERP teletherapy sessions in their homes (6.6±3.4 sessions per patient; Figure 1c). During ERP, patients

were exposed to personalized OCD triggers (Figure 1d) and verbally reported momentary distress from zero to ten using the subjective units of distress rating scale (SUDS; ratings per session: 12.3±8.5; Figure 1e-f). All but one patient achieved a clinically significant response to DBS (responder criteria: ≥35% reduction in YBOCS[26]; Supplementary Table 1), resulting in substantial longitudinal variation in OCD symptom severity throughout the study[25] (Figure 1g).

These complementary paradigms enabled us to examine neural activity associated with two clinically relevant but distinct phenomena: momentary distress experienced during individualized ERP exposures and fluctuations in longitudinal OCD symptom severity (Figure 1h-j). Therefore, we recorded local field potentials (LFP) bilaterally from the VC DBS target (n=8) and the OFC (n=5) during both clinic visits and at-home ERP sessions (Figure 1h). From these recordings, we extracted a comprehensive set of spectral, temporal, waveform, bursting, and connectivity features (Methods; Figure 1i,j)[24,27]. To investigate neural features associated with momentary distress, we analyzed the neural recording surrounding each SUDS rating collected during ERP, and for long-term OCD symptom severity, we computed one neural feature vector per clinical visit and paired it with the corresponding YBOCS-II score obtained at that visit. This experimental design allowed us to directly compare neural representations of transient distress and longitudinal clinical symptom severity within the same individuals.

Distinct neural representations of momentary distress and long-term symptom severity

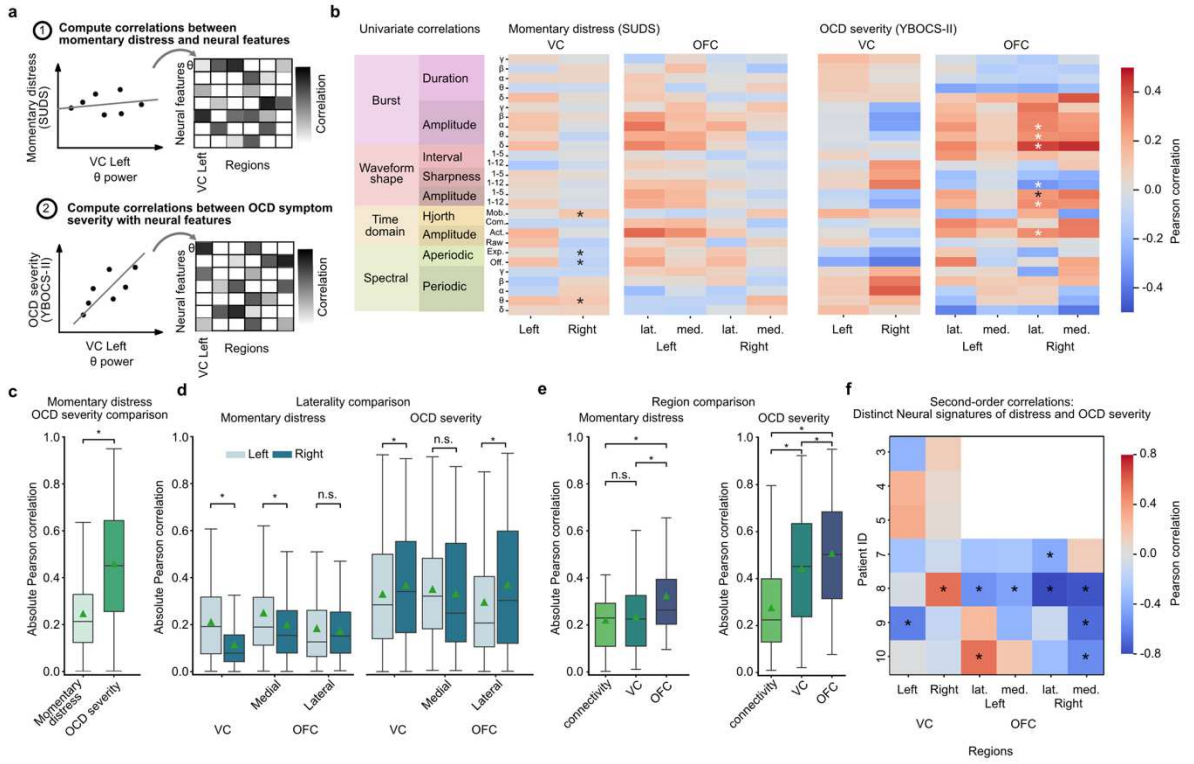


Figure 2. Neural signatures of momentary distress and longitudinal OCD symptom severity. (a) Overview of correlation analyses comparing neural features with momentary distress and OCD symptom severity. **(b)** Mean region- and hemisphere-specific Pearson correlation heatmaps across

patients, including orbitofrontal cortex (OFC) and ventral capsule (VC) recordings in both left and right hemispheres. Significant correlations after multiple-comparison correction are marked with (*). Patient-specific correlations are shown in Supplementary Figure 2. **(c)** Absolute neural feature correlations with OCD symptom severity and momentary distress. **(d)** Comparison of left and right hemisphere correlations for momentary distress and OCD symptom severity separate for VC, lateral and medial OFC. **(e)** Comparison of VC, OFC and connectivity correlation strengths for momentary distress and OCD symptom severity. **(f)** Second-order correlation analysis comparing feature–region correlation maps for momentary distress and OCD symptom severity for VC and OFC recordings.

We first asked whether longitudinal OCD symptom severity and momentary distress exhibit similar neural correlates. To address this question, we computed correlations between a comprehensive set of intracranial neural features and both OCD symptom severity and momentary distress across all recorded brain regions (Figure 2a-b, Supplementary Figure 2). The strongest individual neural feature associated with OCD symptom severity across patients was right lateral OFC delta burst amplitude (|R|=0.45±0.40, multiple comparison corrected p=0.45) and for momentary distress left lateral OFC Hjorth activity (|R|=0.36±0.36, multiple comparison corrected p=0.21). However, these group-wise correlations did not reach significance, likely due to inter-patient heterogeneity. Motivated by previous findings implicating cortico-basal hyperconnectivity in OCD[10,28,29], we tested whether OFC–VC connectivity was associated with OCD symptom severity. In two of five patients, delta coherence across bilateral OFC–VC channels was significantly and positively correlated with symptom severity (Supplementary Figure 2). At the group level, this pre-specified hypothesis test also showed a significant positive association (R=0.14±0.29, p=0.039), although it did not remain significant after correction for multiple comparisons (corrected p=0.13, Supplementary Figure 2).

Across all features, neural activity demonstrated significantly stronger associations with longitudinal OCD symptom severity than with momentary distress ($|R|$=0.46±0.23 vs. $|R|$=0.24±0.17, respectively; relative permutation test $p=2\cdot10^{-4}$, Figure 2c). Furthermore, neural correlates of momentary distress exhibited significant lateralization within the VC and medial OFC, with stronger associations observed in the left hemisphere than the right (relative permutation test VC p=0.003; medial OFC p=0.001, lateral OFC p=0.51, Figure 2d). In contrast, neural correlates of OCD symptom severity showed significantly stronger associations in the right than the left VC and lateral OFC (relative permutation test VC p=0.027; lateral OFC p=0.0006, Figure 2d). OFC demonstrated significantly stronger correlations than VC for both OCD symptom severity (p=0.01) and momentary distress ($p=2\cdot10^{-4}$, Figure 2e). Together, these findings show that the neural representation of symptom severity is more strongly represented than that of momentary distress; that these neural representations are lateralized; and that cortical features correlate with both severity and distress more strongly than do subcortical features.

Because DBS parameters were adjusted throughout the clinical trial, we next asked whether the observed neural correlates of OCD symptom severity could simply reflect acute stimulation effects rather than longitudinal changes in clinical state (Supplementary Figure 3). To control for this effect, we used neural recordings acquired before and after stimulation parameter changes within individual visits to distinguish acute programming-related changes in neural features from longitudinal

changes associated with fluctuations in OCD symptoms. Although DBS programming produced measurable changes in neural features (Supplementary Figure 4), these changes were significantly smaller than those observed longitudinally across visits (Bernoulli test comparing the number of feature changes across versus within visits: $p=1.5\cdot10^{-17}$). These findings support that the observed neural correlates primarily reflected longitudinal changes in OCD symptom severity rather than the amount of stimulation delivered or acute effects of DBS programming.

Finally, we asked whether the neural representations of momentary distress and longitudinal OCD symptom severity captured a common underlying neural organization. To address this question, we computed second-order correlations that quantified the similarity between the feature-region correlation maps of OCD severity and momentary distress. Strikingly, second-order correlations were absent in the VC (R=-0.02±0.29, p=0.82) and negative in the OFC (R=-0.28±0.37, p=0.008), demonstrating that neural representations of momentary distress are dissociable from those underlying long-term OCD symptom severity (Figure 2f).

Behavioral signatures distinguish momentary distress and OCD symptom severity

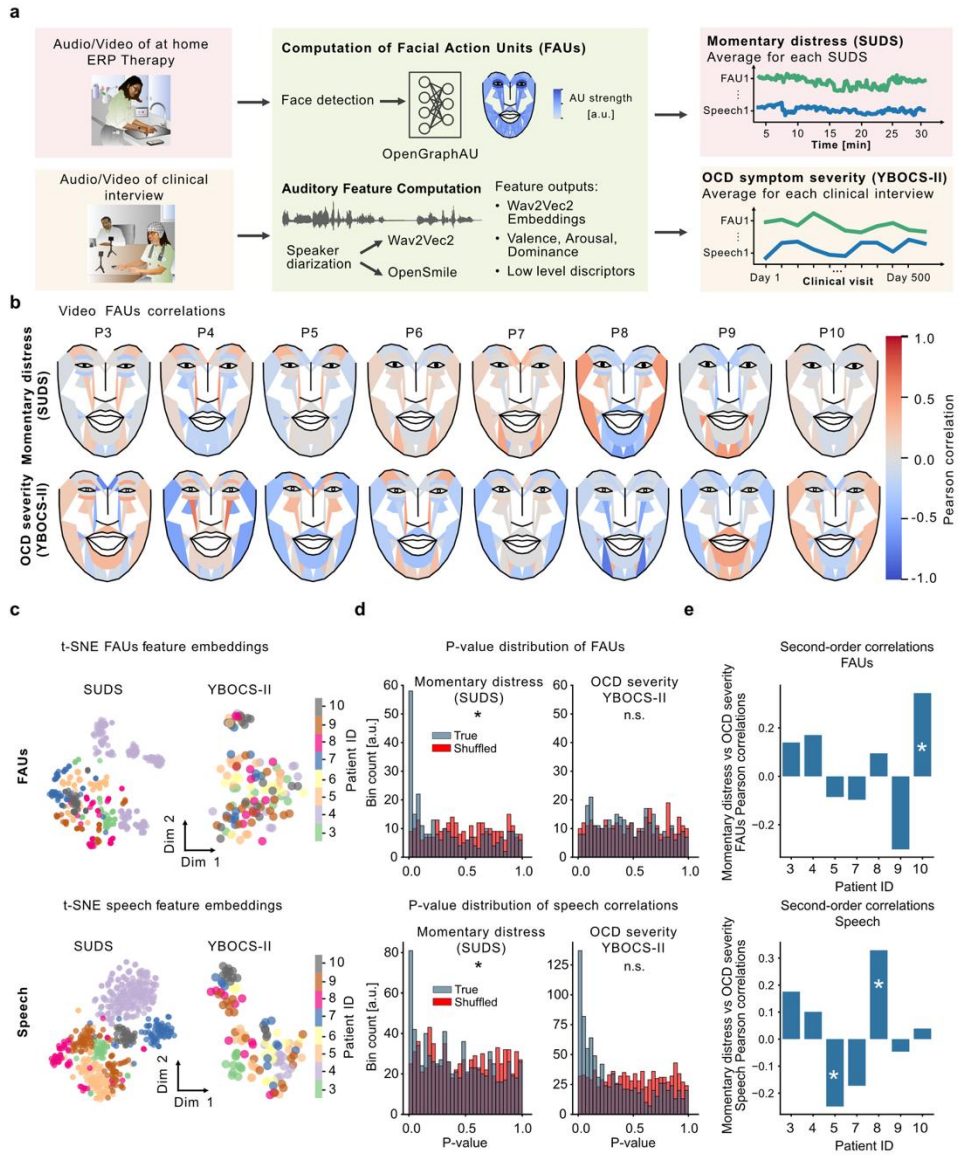


Figure 3. Behavioral representations of momentary distress and OCD symptom severity. (a) Multimodal analysis integrating chronic intracranial neural recordings with facial expression and speech characteristics extracted from video recordings obtained during at-home exposure and response prevention (ERP) therapy and interviews during clinic visits. Facial action units (FAUs) and speech characteristics (OpenSmile derived acoustic features[30], Wav2Vec2 embeddings and valence/arousal/dominance estimates[15,31]) were extracted from video and audio recordings, respectively. For momentary distress, behavioral features were averaged over windows around each distress rating. For OCD symptom severity, behavioral features were averaged within each clinical interview and paired with the neural recording corresponding to each YBOCS-II assessment. Figure adapted from[12]. **(b)** Patient-specific correlations between FAUs and momentary distress (top) or OCD symptom severity (bottom). **(c)** Patient-specific t-SNE visualization of FAUs (top) and speech features (bottom) with momentary distress and OCD symptom severity. **(d)** Permutation test distributions evaluating associations between FAUs (top) and speech features (bottom) with momentary distress and OCD symptom severity. **(e)** Second-order correlation analysis comparing behavioral representations associated with momentary distress and longitudinal OCD symptom severity for FAUs (top) and speech features (bottom).

Having established that momentary distress and longitudinal OCD symptom severity are associated with distinct cortical and subcortical neural signatures, we hypothesized that these clinically relevant states were also associated with distinct

objective behavioral features. To test this hypothesis, we extracted facial action units (FAUs)[32] and patient-derived acoustic speech features (OpenSmile acoustic metrics[30], Wav2Vec2 speech embeddings[33], and valence, arousal, dominance speech estimates[31]) from video and audio recordings obtained during clinic visits and ERP sessions (Figure 3a). For momentary distress, video- and audio-derived behavioral features were averaged surrounding each SUDS rating during ERP sessions. Likewise, for longitudinal OCD symptom severity, behavioral features were averaged across each clinical interview and paired with the corresponding YBOCS-II score. In this way, we objectively quantified patterns of facial expressions and verbal output produced by each patient across various states of momentary distress and of clinical response.

Behavioral correlates of both measures were highly individualized (Figure 3b), resulting in distinct patient-specific clustering of facial and speech representations (Figure 3c). At the group level, both facial action units and speech features exhibited significant associations with momentary distress compared with shuffled-score null distributions after false discovery rate (FDR) correction (FAUs: $p=10^{-4}$; speech: $p=10^{-4}$; Figure 3d). Neither of the behavioral variables remained significantly correlated with OCD symptom severity after FDR correction. We then quantified the similarity between behavioral representations of momentary distress and OCD symptom severity. Consistent with the neural analyses, behavioral representations of momentary distress were not positively correlated with those of longitudinal OCD symptom severity (FAUs: $R=0.04\pm0.21$, permutation test $p=0.65$, speech: $R=0.02\pm0.20$, $p=0.75$; Figure 3e). These results recapitulate the finding that short-term (momentary distress) and long-term (symptom severity) measures have distinct representations, not just neurally but also behaviorally.

Joint neural–behavioral modeling enhances OCD symptom decoding

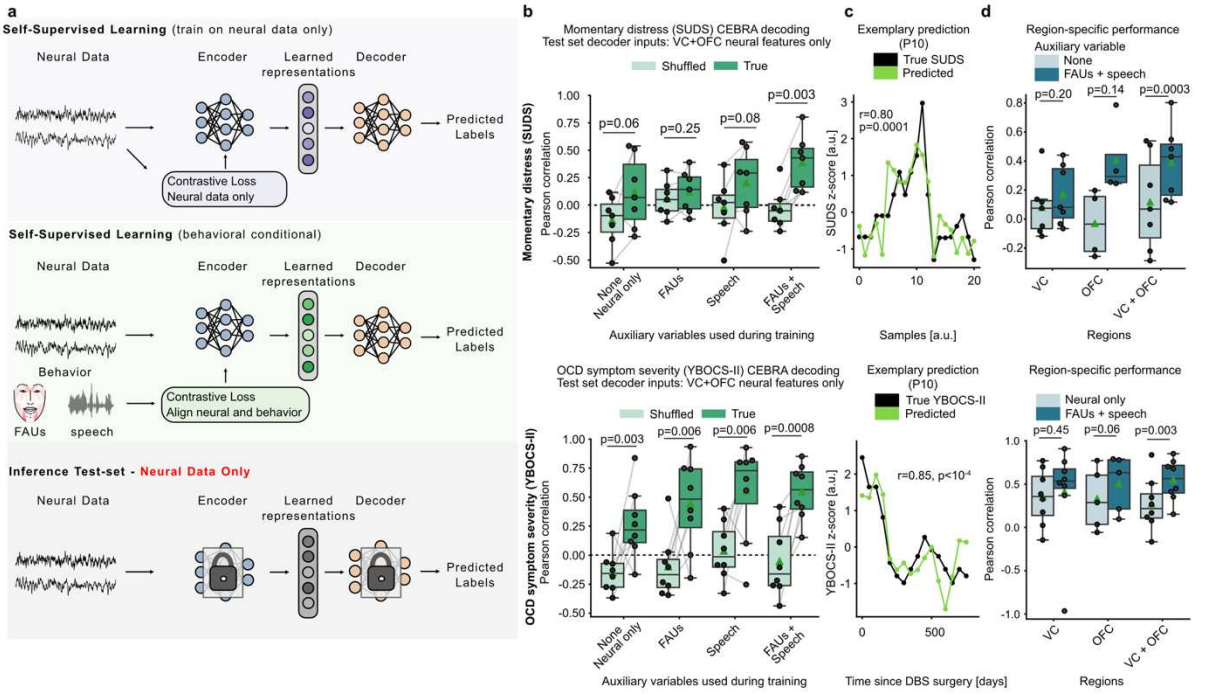


Figure 4. Behavior-guided neural representation learning improves decoding of momentary distress and longitudinal OCD symptom severity. (a) Overview of the CEBRA-based[34] representation learning and decoding framework. Neural embeddings were learned either from neural recordings alone (self-supervised) or jointly with auxiliary behavioral variables (facial action units and speech features). During self-supervised training, neural samples were grouped solely based on temporal proximity (top). In contrast, during behavior-guided training, neural samples with similar behavioral representations were grouped within the latent space (middle). During inference, only neural recordings were provided to the frozen encoder and downstream decoder (bottom). **(b)** Decoding performance for momentary distress (top) and OCD symptom severity (bottom) using neural-only and behavior-guided neural embeddings. Chance-level decoding performance was estimated using shuffled clinical labels. **(c)** Representative predictions of momentary distress (P10; top) and OCD symptom severity (P10; bottom) using neural-only and behavior-guided embeddings. **(d)** Regional decoding performance using ventral capsule (VC)-only, orbitofrontal cortex (OFC)-only, or combined VC-OFC neural features.

Given the significant but distinct neural and behavioral correlates of acute vs. chronic measures, we next tested whether these correlates were clinically relevant and complementary. We assessed clinical relevance by incorporating these variables within a predictive modeling framework to determine their ability to predict, rather than simply correlate with, measured outcome. To assess complementarity, we compared the performance of models utilizing subsets of the neurobehavioral data as predictors. We hypothesized that neural and behavioral data provide independent information regarding clinical state (i.e., acute distress and chronic severity), such that models trained using the full neurobehavioral dataset would outperform models trained on a single modality.

To do so, we leveraged the CEBRA framework[34] to learn latent neural representations using either neural features alone or neural features jointly trained with behavioral auxiliary variables (Figure 4a). Importantly, behavioral variables were used only during latent representation learning to shape the latent neural embedding; all decoding performance was subsequently evaluated using neural recordings alone, such that test set predictions require only neural data and do not depend on continuous monitoring of behavioral markers. In the neural-only model, latent embeddings were learned using the temporal structure of neural activity alone, whereas the behavior-guided model used facial and speech features to organize neural samples with similar behavioral representations. We hypothesized that, if the behavioral data provided information about the outcome measures beyond that captured by the neural data, incorporating these data would increase the separability of the learned representations and thereby improve outcome prediction. We trained separate models to decode momentary distress and longitudinal OCD symptom severity within each participant using leave-one-sample-out cross-validation. For decoding OCD symptom severity, we left out the entire recording from each clinical visit for test evaluation. This framework enabled us to directly compare neural-only embeddings with joint neural-behavioral embeddings and determine whether behavioral information improves downstream neural decoding.

Joint neural-behavioral representation learning did indeed enable improved decoding performance compared to neural-only models for both measures (Figure 4b-c). For acute distress (i.e., predicting SUDS ratings), neural-only latent embeddings

performed at chance levels (actual R=0.12±0.31, shuffled R=-0.14±0.02, p=0.06), but the model achieved significant decoding performance with inclusion of behavioral variables (actual R=0.39±0.23, shuffled R=-0.03±0.17, p=0.003). Across the board, decoding of chronic symptom severity outperformed decoding of momentary distress, consistent with the stronger neural correlations for the former (Figure 2c). For chronic severity (i.e., predicting YBOCS-II scores), neural-only latent embeddings performed above chance (actual R=0.27±0.28, shuffled R=-0.15±0.16, p=0.003), but performance improved significantly when facial and speech variables were included (actual R=0.55±0.22, shuffled R=-0.06±0.28, p=0.0008). Notably, behavior-guided representation learning significantly improved OCD symptom severity decoding in all eight patients. These findings demonstrate that face- and speech-derived behavioral features provide independent and complementary predictive information regarding clinical states beyond that provided by cortical and subcortical intracranial neural recordings. Moreover, this contrastive learning approach allows incorporation of the rich behavioral data into the learned low-dimensional representation when they are available during model training but does not require them during inference. Since the neural data are passively recorded on the DBS device and therefore readily available, but the behavioral data require availability of external audio/video equipment, this approach provides enhanced decoding performance even when only neural data are available during the vast majority of the time in real-world settings.

We further investigated the relative model contributions of neural data from the cortex (OFC) vs. the subcortical stimulation site (VC). Based on the correlation analyses above (Figure 2e) and results from other conditions such as Parkinson's disease[35–37], we hypothesized that cortical signals would enhance decoding. We therefore assessed decoding performance by restricting neural features to VC-only, OFC-only, or combined VC-OFC features. OFC-only embeddings yielded the highest decoding performance for acute distress (VC: R=0.17±0.21, OFC: R=0.41±0.26, VC-OFC: R=0.39±0.25), whereas combining VC and OFC produced the highest decoding performance for chronic symptom severity (VC: R=0.41±0.58, OFC: R=0.50±0.33, VC-OFC: R=0.55±0.24; Figure 4d).

Finally, we asked whether decoders optimized to predict one clinical measure generalized to the other. Consistent with the second-order correlation results (Figure 2f), decoders trained to predict acute distress failed to predict chronic symptom severity (R=−0.09±0.32, permutation test p=0.73), and decoders trained on severity likewise failed to predict distress (R=−0.20±0.40, p=0.25; Supplementary Figure 5). These cross-decoding results provide convergent evidence that acute distress and chronic symptom severity are encoded by distinct neural representations.

Discussion

Our findings reveal that an acute manifestation of psychiatric illness is not simply a short-timescale expression of the same neural state that encodes longer-term clinical burden. In OCD, momentary distress and longitudinal symptom severity were associated with dissociable neural and behavioral representations, and neural decoders trained on one state failed to generalize to the other. These results suggest that psychiatric illness may be organized across nested but separable neural states,

with transient symptom expression occurring within broader disease-related dynamics.

This distinction has important implications for the interpretation of psychiatric biomarkers. Momentary distress is inherently transient, context-dependent, and typically assessed through subjective self-report, whereas symptom severity is measured using clinical scales acquired by trained assessors that capture a broader disease state incorporating chronic burden. Notably, this dissociation existed even though the distressing exposures occurred naturalistically in the home environment, indicating that acute distress and clinical severity remain separable even under ethologically relevant symptom-provoking conditions. Because the YBOCS-II assesses the ability to tolerate obsessions, resist compulsions, limit avoidant behaviors, and maintain behavioral control over a timescale of weeks, our findings suggest that longitudinal clinical measures capture aspects of disease state not reflected by isolated distress ratings. More broadly, these findings suggest that clinically relevant psychiatric states are organized across multiple temporal scales, with distinct neural representations encoding different aspects of disease. Accordingly, psychiatric biomarkers should be developed with careful consideration of the clinical state they are intended to estimate.

Our findings further demonstrate that objective behavioral measures provide complementary information that enhances neural representations of clinically relevant psychiatric states. Clinical assessment in psychiatry relies heavily on behavioral observation, including facial expression, speech, affect, and motor behavior, to infer symptom severity and treatment response[17]. Consistent with this clinical practice, previous studies have demonstrated the value of behavioral readouts for quantifying psychiatric disorders[18–20,38,39], while separate work has shown that behavior-guided latent representation learning can improve downstream decoding[34]. Our findings extend these observations by demonstrating that objective behavioral measures can improve neural representations of both short-term distress and long-term symptom severity, suggesting that behavioral information can guide the identification of clinically meaningful neural states.

Importantly, behavioral information was used only during latent representation learning and was not required during model inference. Rather than functioning as additional decoder inputs, behavioral measures improved the organization of the latent neural representation itself, resulting in substantially improved decoding of both clinical measures. One explanation for this improvement is that intracranial neural recordings simultaneously reflect numerous ongoing brain processes, many of which are unrelated to the clinical state of interest. By providing an independent measure of clinically relevant behavior, facial and speech features may guide representation learning toward neural dimensions that preferentially encode disorder-related neurophysiological processes while reducing emphasis on unrelated neural variability.

The improved decoding performance despite modest bivariate correlations between behavioral features and OCD symptom severity (Figure 3d) may reflect nonlinear relationships captured by CEBRA's latent space. By conditioning the latent embedding

on behavioral features during behavioral contrastive learning, CEBRA can reveal neural patterns that become informative for symptom decoding only in specific behavioral contexts or through nonlinear interactions[34], even when either modality shows little linear association with symptom severity in isolation.

Our findings suggest that objective behavioral measures can provide meaningful supervisory information for learning neural representations of psychiatric state while preserving the ability to decode from neural activity alone. This arrangement is critical, as behavioral data are often the bottleneck, whether acquired from external cameras and microphones (which are intrusive and limited in their deployability) or from clinical interactions (which are relatively infrequent due to patient- and provider-related practical limitations). As methods for joint neuro-behavioral pre-training continue to advance[21], this framework will enable predictive models to be pre-trained on large multimodal datasets and subsequently fine-tuned using sparser psychiatric ratings, thereby facilitating robust biomarker discovery and longitudinal prediction of disease state without increasing cost and burden to practitioners and patients.

Our findings also provide insight into the neural circuitry underlying OCD symptom expression and its implications for closed-loop or adaptive DBS. According to the prevailing model of OCD pathophysiology, symptoms arise from dysfunction within cortico-striato-thalamo-cortical circuits that regulate reward-motivated behavior and controlled decision-making[10,13,40]. Consistent with this framework, we found that elevated delta-band coherence between cortical and striatal regions was associated with greater OCD symptom severity (Supplementary Figure 2), suggesting that worsening clinical state is accompanied by alterations in circuit-level communication. Interestingly, univariate cortical signals consistently exhibited stronger associations with longitudinal OCD symptom severity than either subcortical or connectivity measures, consistent with the hypothesis that cortical recordings provide particularly rich information about disease state. Aligned with this hypothesis, cortical signals have shown stronger predictive power for decoding movement and symptom severity in Parkinson's disease[35,36]. These findings suggest that future adaptive DBS systems may benefit from incorporating cortical recordings as biomarker sources[41]. More generally, these findings illustrate the need to separate priorities of optimizing therapeutic stimulation sites (which are often subcortical to engage broadly distributed networks) from neural biomarker recording sites (which may benefit from the specificity of cortical signals) to facilitate the development of more effective closed-loop neuromodulation systems.

The delayed therapeutic effects of DBS for psychiatric disorders provide important insight into the mechanisms underlying therapeutic response in psychiatric neuromodulation. Unlike in Parkinson's disease, where stimulation often produces rapid symptomatic benefit, clinical improvement after DBS for OCD typically emerges only after weeks to months of DBS therapy[42]. Rather than acutely suppressing obsessions, VC stimulation may facilitate a gradual learning process that helps patients progressively tolerate and manage OCD-related distress over time[7,8,13]. This proposed mechanism parallels ERP, in which repeated exposure to symptom-provoking triggers while refraining from engaging in compulsions gradually

strengthens distress tolerance and behavioral control rather than immediately eliminating distress. Our clinical experience similarly suggests that obsessive thoughts rarely disappear during the initial months of DBS treatment. Instead, patients become increasingly able to tolerate obsessive thoughts and resist compulsive urges[25,43]. This gradual trajectory of recovery is consistent with our correlation and prediction findings, which show that longitudinal clinical severity is more robustly represented by neural activity than is transient distress. These longitudinal neural signatures may therefore provide a window into the slowly evolving neural state that changes at a rate similar to therapeutic response after DBS. This observation raises the possibility that psychiatric DBS may exert its therapeutic effects by gradually modifying an underlying disease-related state rather than by acutely suppressing individual episodes of distress.

The concept that neuromodulation exerts its effect by decreasing the likelihood of pathological manifestations of a disorder rather than disrupting individual bouts of symptom expression is exemplified in epilepsy. Episodic seizure events occur within the background of slower-varying neural states that influence seizure susceptibility[44–46]. Notably, open-loop DBS can reduce seizure burden without necessarily interrupting individual seizures as they occur. Recent evidence suggests that even responsive neurostimulation, which is conceptually designed to abort individual seizure events, may exert its therapeutic effect by influencing seizure likelihood over longer timescales. More broadly, these observations support a framework in which transient pathological events do not necessarily constitute direct readouts of the slower neural states that determine susceptibility to those events.
Although episodes of momentary distress in OCD are fundamentally different from epileptic seizures, our results suggest that a similar distinction between background disease state and transient symptom expression may exist. We hypothesize that longitudinal OCD symptom severity reflects a slowly varying neural state that shapes how patients respond to everyday triggers, whereas momentary distress represents a transient behavioral manifestation occurring within that broader disease state[38,47–49]. Under this framework, the brain states captured by chronic neural recordings, which reflect patients' ability to tolerate distress, resist compulsions, and adapt habitual behaviors[7,8] are more saliently represented in cortical and subcortical neural activity than those related to transient episodes of momentary distress. Future studies incorporating continuous neural sensing and longitudinal behavioral monitoring are needed to further test this hypothesis.

Our dataset includes a large and robust collection of intracranial cortical and subcortical time-domain neural recordings acquired in the home environments of OCD patients during naturalistic exposures and across clinical visits, spanning more than 4595 patient-days. Despite this extensive dataset, both the number of participants and the number of independent distress and symptom ratings remain limited relative to the neural data. Given the substantial heterogeneity in symptom expression and neural biomarkers observed across psychiatric disorders[15], we therefore developed within-patient decoders. As larger datasets become available, more complex models may become feasible and enable generalization across patients[24,35]. Similar to neuroimaging research, establishing robust electrophysiological signatures will require multi-cohort validation and large-scale dataset harmonization[50].

An important future implication of this framework is using it to determine whether adaptive neuromodulation should target slowly evolving disease states rather than transient symptom events. If psychiatric disorders are organized across varying neural timescales, future closed-loop therapies may benefit from modulating the underlying clinical state that governs long-term clinical recovery instead of reacting to individual symptom episodes. This control objective differs from that of adaptive DBS in Parkinson's disease, where stimulation can be adjusted in response to neural signals that track motor symptoms over relatively short timescales. Because therapeutic effects in psychiatric DBS emerge more gradually, adaptive systems may instead need to monitor and stabilize slower-varying clinical states while maintaining stimulation sufficient to sustain therapeutic circuit dynamics. If confirmed, it could fundamentally shift the design of adaptive neuromodulation from reactive symptom detection toward modulation of disease state.

Methods

Clinical Study Design

We conducted an early feasibility trial of adaptive DBS biomarker discovery in adults with severe, treatment-refractory OCD (NCT04281134 and NCT03457675). We enrolled ten adults with a primary diagnosis of severe and intractable OCD, including a previously reported subgroup of five patients[12], which described preliminary clinical outcomes and neural analyses. A limited subset of data from these patients has been reported previously in studies addressing distinct, unrelated research questions[41,51,52], and clinical outcomes were previously reported in a separate publication[25]. Each participant gave fully informed consent according to study sponsor guidelines, and all procedures were approved by the local institutional review board at Baylor College of Medicine (H-40255 and H-44941 to Baylor College of Medicine) and the US Food and Drug Administration (FDA) Center for Devices and Radiological Health (G170268). Participants received a stipend after each study visit via a ClinCard. All participants had a disease duration exceeding five years and had either not responded to or could not tolerate trials of multiple pharmacological treatments (including selective serotonin reuptake inhibitors (SSRIs), clomipramine, and SSRI augmentation with antipsychotics), as well as an expert course of exposure and response prevention (ERP) therapy. We implanted the investigational Activa PC+S system (Medtronic, Inc.) in the first two participants[25]. Data from these participants were excluded from analyses due to the limited streaming capabilities available on the Activa PC+S device in the home environment of participants. We implanted the investigational Summit RC + S system (Medtronic, Inc.) in the following eight participants including bilateral DBS leads (Model 3387, Medtronic, Inc.; 1.5-mm contact length and 1.5-mm intercontact spacing) within the ventral capsule (VC), targeted using preoperative MRI[25]. In five of these patients, we additionally placed bilateral four-contact orbitofrontal cortex (OFC) electrocorticography (ECoG) strip electrodes (model 0913025; 4-mm contact diameter and 10-mm intercontact spacing)[25] to test the hypothesis that cortical signals would improve decodability. In these patients, we implanted two devices per patient, with each system recording from one hemisphere VC and OFC electrodes. We connected the leads to extension cables, tunneled them

subcutaneously along the neck, and attached them to the Summit RC + S pulse generator positioned in the upper chest. Throughout the clinical trial, we assessed OCD symptom severity using the Yale-Brown Obsessive Compulsive Scale, Second Edition (YBOCS-II)[5] during both in-person clinical visits and remote assessments (mean number of YBOCS-II assessments per patient: 22.5±1.5). For the present analysis, we paired YBOCS-II scores with neural recordings acquired during the in-clinic visits. We excluded assessments when neural data was not recorded or contained substantial artifacts (mean number of excluded YBOCS-II assessments per patient: 7.9±5.2).

Electrode localization

We acquired a preoperative clinical computed tomography (CT) scan, T1-weighted (T1w) MRI scan, and a postoperative clinical CT scan to verify electrode placement. We co-registered the postoperative CT to the preoperative T1w MRI and used the aligned images to determine electrode contact locations. We performed automated cortical reconstruction on the preoperative T1w MRI using FreeSurfer v7.1.1[53], and incorporated the T2-weighted (T2w) MRI to refine pial surface reconstruction. We aligned the postoperative CT to the preoperative T1w MRI using the Linear Image Registration Tool (v6.0) from the FMRIB Software Library[54,55]. We manually identified OFC and VC electrode coordinates from the co-registered CT data in BioImage Suite v3.5b1[56] and mapped them into native MRI space. Finally, we visualized OFC cortical electrode coordinates using Surfice (https://www.nitrc.org/projects/surfice/) and VC coordinates using Lead-DBS[57] both in MNI152 standard space.

Electrophysiological recordings

The Summit RC + S system enables simultaneous therapeutic neurostimulation and continuous intracranial local field potential (LFP) recording[58]. We used the Clinician Telemetry Module to establish bidirectional communication between the implanted device and a custom-built application developed with the Research Development Kit, which was connected via Bluetooth to a separate host computer. In both clinical and home settings, we deployed custom clinician- and patient-facing applications on Microsoft Surface Pro and Surface Go tablets. These applications interfaced with the implant and synchronized data to a secure cloud-based storage platform (https://www.box.com/), thereby providing a flexible infrastructure for longitudinal data acquisition and storage. We recorded LFPs at sampling rates of 250, 500, or 1000 Hz during clinic visits and at 250 Hz in the home environment. We configured recordings in a bipolar sensing mode by selecting pairs of adjacent contacts on each lead, referencing the signal from one contact to the other. Supplementary Figure 3 summarizes stimulation contact configurations for each participant. Whenever feasible, we positioned the sensing contact pair to flank the active stimulation contact to minimize stimulation-related artifacts. In addition, we set both stage 1 and stage 2 low-pass filter cutoffs to 100 Hz to further reduce contamination of the recorded signals by stimulation artifacts.

DBS stimulation parameter programming

At each clinical visit, we obtained two resting-state recordings: one prior to DBS programming and one following parameter adjustments (Supplementary Figure 4). The first resting-state recording reflected the stimulation settings programmed during the previous clinical visit. The second resting-state recording captured the updated stimulation parameters when changes were implemented. We optimized DBS parameters exclusively based on clinical judgment, incorporating standardized clinical assessments (e.g., YBOCS-II), direct interaction with participants, and their subjective reports of mood, anxiety, and alertness during acute programming sessions. To identify optimal stimulation settings, the clinician systematically modified a single parameter at a time, either stimulation amplitude or pulse width. DBS frequency was set to 150.6 Hz for all patients throughout the study.

Exposure and response prevention therapy recordings

During ERP teletherapy sessions (Figure 1d–f), we instructed participants to initiate a neural recording at the start of each appointment. We saved video recordings locally on the clinician's computer. Prior to each ERP session, a psychiatrist discussed and defined individualized exposure tasks with the participant. Throughout the session, participants repeatedly rated their distress using the Subjective Units of Distress Scale (SUDS), ranging from 0 to 10.

Audio and video recordings

During clinical visits, we recorded video and audio using a GoPro Hero 6 at 25 frames per second and audio using a Zoom H4n Pro 4-Track portable recorder at 44.1 kHz. We synchronized the audio and GoPro videos by cross-correlating the video's onboard audio track with the external audio recording. After alignment, we removed the original GoPro audio and replaced it with the high-quality external audio track. We recorded videos during ERP sessions directly on the Surface Pro tablet and saved the videos with embedded timestamps on each frame. We used the time-of-day information extracted from the video to align the recordings with the corresponding LFP Unix timestamps.

Electrophysiological signal processing

We used the analysis-rcs-data toolbox to preprocess neural data, account for packet loss, and enable synchronization between LFP and other data streams[59]. For patients with OFC implants, we merged left and right implant recordings based on UNIX timestamps. Within ERP recordings, we aligned SUDS ratings derived from video annotations by converting video time to UNIX timestamps. We annotated and rejected artifacts, including the recently described dual-device beat frequency artifact[51,60], through manual annotation in 5-second segments. We then computed time-resolved features using the py-neuromodulation feature extraction toolbox[24]. Specifically, we calculated band-power using the scipy Welch function (segment length of 250 samples)[61]. We derived log-transformed band-power across multiple frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–15 Hz), beta (15–30 Hz), and gamma (30–55 Hz). We used specparam (default parameters, version 1.1.1)[62] to estimate aperiodic exponent and offset parameters within a fitting range of 10–50 Hz. In addition, we computed bursting features for each of the above-defined frequency bands. We applied the MNE-Python band-pass filter[63], resampled the signal to 50 Hz, calculated

the absolute Hilbert transform, and determined mean burst duration and amplitude using a 75th percentile threshold, as implemented in py-neuromodulation[24,27]. We also extracted temporal waveform-shape features for different filter ranges (1–5 Hz and 1–12 Hz). Waveform-shape features have previously been introduced as complementary LFP characteristics to dissociate physiological and symptomatic states[64–66]. We computed the following temporal waveform shape features in a batch-wise manner for both signal-detected troughs and peaks: prominence, sharpness and interval. Prominence quantifies the mean amplitude difference between a trough and its surrounding peaks (or a peak and its surrounding troughs). Sharpness measures the voltage deflection at each trough or peak relative to voltage values 5 ms before and 5 ms after the detected troughs or peaks, and interval computes the time difference between consecutive troughs or peaks. Because signal polarity can be ambiguous, we computed each waveform-shape feature separately for troughs and peaks and averaged the results. Merk et al. [24] described in further detail the computed waveform shape features. Furthermore, we extracted raw LFP voltage amplitudes as well as Hjorth activity, mobility, and complexity parameters, as originally described[67]. We also computed connectivity measures between recording sites. First, we estimated one-second time-resolved aperiodic exponent and offset parameters (see computation method above) and then calculated Pearson correlations between the aperiodic exponent or offset time series of different recording regions. In addition, we computed coherence using scipy[61] and averaged coherence values within each of the above predefined frequency bands. For ERP recordings, we calculated the mean feature representation surrounding each momentary distress SUDS rating. For in-clinic OCD symptom severity recordings, we computed the average feature representation within each resting-state recording.

Audio and video signal processing

For both momentary distress and OCD symptom severity, we utilized the same audio and video analysis pipeline. We computed a two-minute average for each SUDS report for momentary distress, and the whole-recording average for OCD symptom severity in resting-state recordings (Figure 3a). To extract Facial Action Units (FAUs) from video recordings, we first converted the images to grayscale using the OpenCV computer vision library[68]. We then applied the detectMultiScale algorithm (haarcascade_frontalface_default; scale factor = 1.1; min_neighbors = 5) using OpenCV[68]. Next, we extracted FAUs using OpenGraphAU[32] with a pre-trained ResNet-18 model. Specifically, we extracted the following FAUs: 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 32, 38, 39, L1, R1, L2, R2, L4, R4, L6, R6, L10, R10, L12, R12, L14, and R14. To compute speech features, we used WhisperX[69] with the “large-v2” model for transcription. We then applied the default diarization pipeline (version 3.4.2), limiting the output to two speaker identities. We manually corrected all transcription and diarization segments to ensure that only patient speech was included for analysis, and restricted utterances to durations between 2 and 10 seconds. For each patient utterance, we computed a set of speech features. These included embeddings from the speech foundation model wav2vec2[33] as well as additional signal-processing–based speech features. We hypothesized that these complementary representations would capture speech characteristics relevant for tracking momentary distress and longitudinal symptom severity. From the wav2vec2-large-robust-12-ft-emotion-msp-dim model[31], we extracted 1,024-dimensional embeddings as well as valence, arousal, and dominance predictions.

Using the openSMILE toolbox, we extracted eGeMAPSv02 speech functionals[70] (n=87) and low-level descriptors[30] (n=25). A complete list of speech features is available in the openSMILE documentation (https://audeering.github.io/, version 2.6.0).

Neural decoding pipeline

We separately investigated decoding of momentary distress and OCD symptom severity. Specifically, we examined models trained using neural features alone and evaluated performance improvements when incorporating behavioral variables to learn a joint latent embedding space. We applied the CEBRA framework, which enables both hypothesis-free (neural-only) and hypothesis-guided training approaches[34]. CEBRA is a contrastive learning framework that constructs a low-dimensional latent representation of neural activity by bringing positive pairs closer together and repelling negative pairs according to a supervisory objective.

The framework uses a nonlinear neural network encoder to transform high-dimensional neural features into a latent embedding optimized through contrastive learning. Unlike conventional supervised learning approaches that directly optimize training loss, CEBRA first learns a latent neural representation before fitting a downstream decoder.

In this study, supervisory information consisted of either temporal proximity alone (self-supervised training) or behavioral similarity derived from facial action units and speech features (behavior-guided training). We trained separate CEBRA models for momentary distress and longitudinal OCD symptom severity. Neural features served as inputs to the encoder for all models. During self-supervised training, neural samples that were close in time were encouraged to occupy nearby regions of the latent space. During behavior-guided training, facial and speech features served as auxiliary variables such that neural samples with similar behavioral representations were mapped nearby in the latent space. After training, the encoder weights were frozen, and only neural recordings were used to generate latent embeddings for downstream decoding. Behavioral variables were used exclusively during representation learning and were not provided during model inference or decoding.

We evaluated all decoding analyses using leave-one-sample-out cross-validation. For each fold, a CEBRA encoder was trained exclusively on the training samples, without any temporal overlap to the corresponding test samples, to learn latent neural representations. Separate encoders were trained for momentary distress and longitudinal OCD symptom severity using the hyperparameters described below. After training, the encoder weights were frozen, and latent embeddings were generated for both the training and held-out test samples. Ordinary least squares linear regression models implemented in scikit-learn[71] were then fit using the training embeddings and corresponding SUDS or YBOCS-II scores and evaluated on the held-out sample. We repeated the retraining process for both the encoder and decoder for each cross-validation procedure. We additionally repeated the training procedure under two conditions: first, by separating behavioral modalities (speech only, FAUs only, or speech and FAUs combined); and second, with different neural feature sets (VC only, OFC only, or both brain regions combined). We evaluated each decoding task using leave-one-sample-out cross-validation. We applied z-score normalization to each sample, fitting the normalization parameters limited to the training data. We reported

the Pearson correlation coefficient across all test samples and performed control analyses using shuffled labels.

To predict momentary distress, we used the offset-1 model with the following hyperparameters: learning rate=3e-3, temperature=1.0, conditional="time_delta", cosine distance, embedding dimension=50, and hidden units=128. For decoding OCD symptom severity, we used the same offset-1 architecture but adjusted the hyperparameters to account for the smaller sample size: output dimension=8, hidden units=32, conditional="delta", and distance="cosine". For hypothesis-free training, we did not include behavioral labels and instead used the default "time" conditional.

Statistical Analysis

All statistical tests were computed using non-parametric Monte Carlo sampled permutation tests with a significance threshold of α=0.05 and n=10000 iterations across participants unless otherwise indicated. Permutation tests were used to compare paired differences in absolute correlation strength across participants between momentary distress and OCD symptom severity and corresponding brain regions (Figure 2c-e). Correlation values and associated p-values were reported using Pearson correlation coefficients computed using scipy (scipy.stats.pearsonr, version 1.16)[61]. P-values for Pearson correlations were computed using a two-sided test against the null hypothesis of zero correlation, based on the exact beta distribution of the sample correlation coefficient, as implemented in scipy. For analyses of correlations between neural features and either momentary distress or OCD symptom severity, p-values were corrected for multiple comparisons across all neural features within each region using the Benjamini–Hochberg method to control the false discovery rate (statsmodels.stats.multitest.multipletests version 0.14.5)[72]. For assessing the significance of the number of neural features across TEED stimulation parameter changes, Bernoulli tests were used to compare the number of significant neural feature changes across conditions: ΔTEED=0 within clinical visits, ΔTEED=0 across clinical visits and ΔTEED≠0 within clinical visits. For the number of significant behavioral correlations with momentary distress and OCD symptom severity, the above-described false discovery rate correction was applied and compared against a distribution of significant correlations within n=10000 shuffled correlations. All decoding analyses were performed using within-subject leave-one-sample-out cross-validation analysis. All CEBRA encoders and least-squares regression decoders were thus retrained for each fold to predict left-out samples. Decoding performance was compared against shuffled label performances, using a defined random seed. Significance of decoding performances using "time" against behavior-guided learning was assessed using relative permutation tests. Statistical comparisons within regions were assessed using relative permutation tests using behavior-guided (FAUs + speech) against "time" self-supervised distributions. All data are reported as mean ± standard deviation unless otherwise noted. All analyses were conducted in Python 3.12 and MATLAB R2024b. Software versions are accessible in the pyproject.toml version control files within the publicly available code repository.

Code and data availability

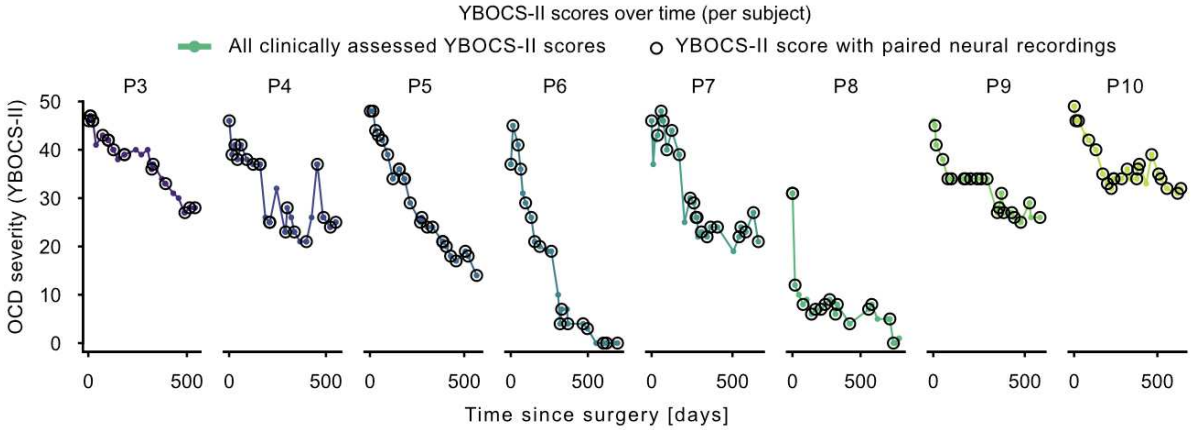
The data that support the findings of this study are available from the corresponding author upon reasonable request. All code is publicly available on GitHub: https://github.com/ProvenzaLab/multiscale-psychiatric-neural-state-representations

Supplementary Tables

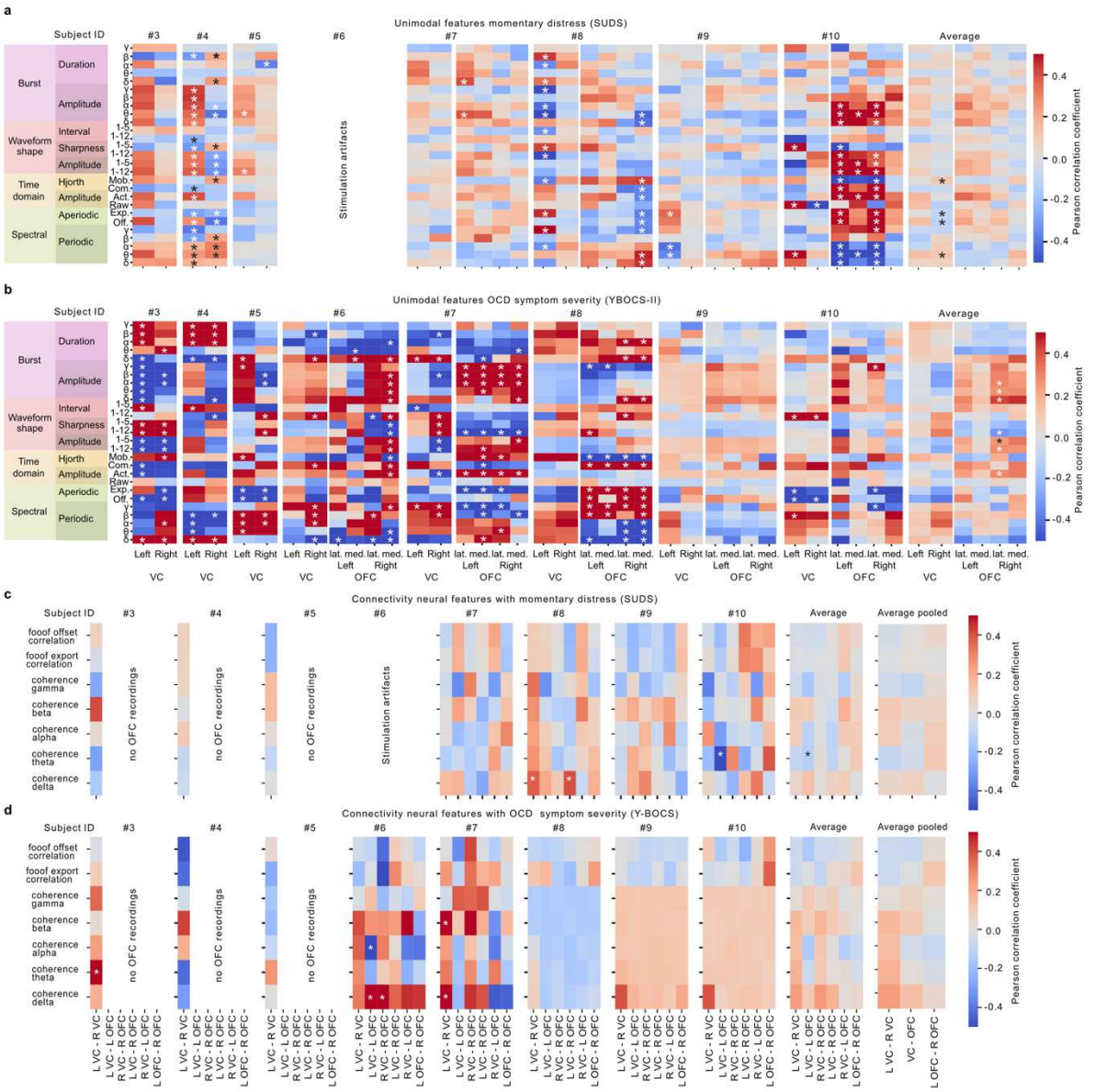
Patient	Gender	Age	Race	Ethnicity	Comorbid diagnoses	OCD subtypes	Age at OCD onset	Initial YBOCS-II score	Final YBOCS-II score
P003	F	38	White	Hispanic/ Latino	PTSD, MDD, TS	Harm, Not Just Right Experiences	12	49	28
P004	M	40	White	Not Hispanic/ Latino	MDD, Bipolar Disorder II	Checking, Contamination, Harm, Scrupulosity, Symmetry and Ordering	16	42	25
P005	F	31	White	Not Hispanic/ Latino	MDD	Contamination, Harm	8	47	14
P006	M	38	Asian	Not Hispanic/ Latino	GAD	Checking, Contamination	30	37	0
P007	M	36	White	Not Hispanic/ Latino	MDD	Harm, Not Just Right Experiences	12	45	21
P008	F	40	White	Not Hispanic/ Latino	MDD, GAD	Checking, Contamination, Harm, Not Just Right Experiences	9	37	1
P009	F	27	White	Hispanic/ Latino	MDD, GAD	Contamination, Harm, Not Just Right Experiences, Scrupulosity	13	45	26
P010	M	33	White	Not Hispanic/ Latino	MDD	Checking, Contamination, Not Just Right Experiences	13	46	32

Supplementary Table 1: Patient demographics and clinical characteristics. YBOCS = Yale-Brown Obsessive-Compulsive Scale; MDD = major depressive disorder; PTSD = post-traumatic stress disorder; GAD = generalized anxiety disorder; TS = Tourette Syndrome; M = male; F = female.

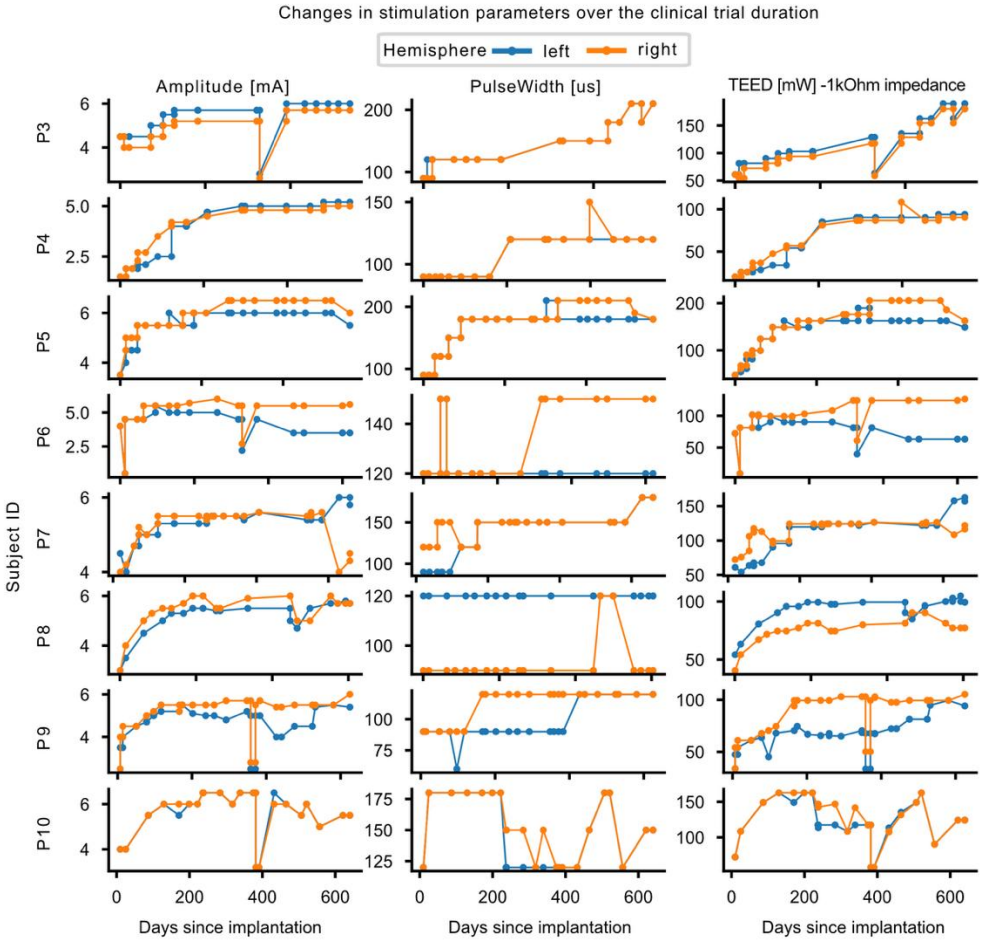
Supplementary Figures



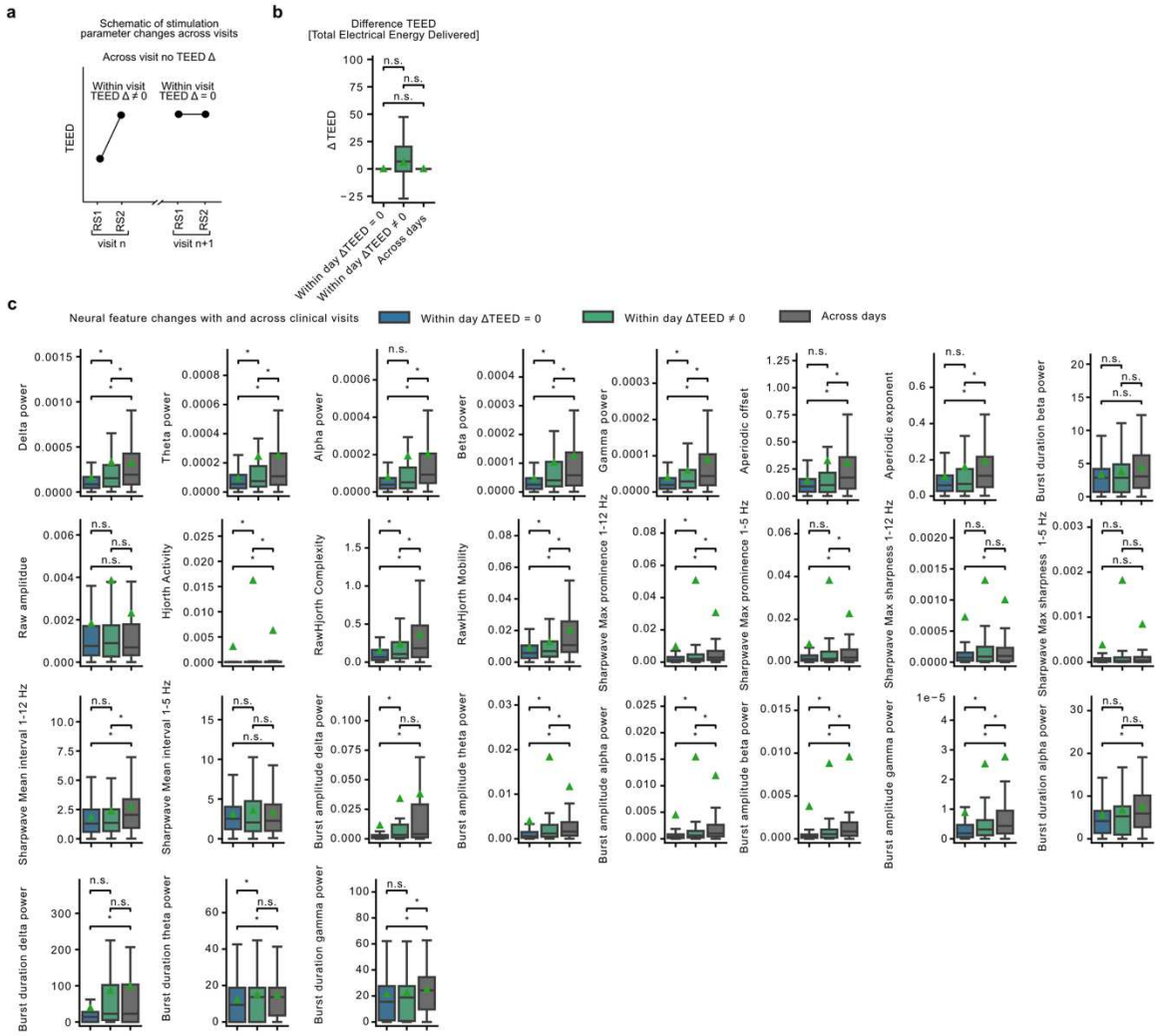
Supplementary Figure 1. OCD symptom severity fluctuations across patients since surgery. All clinically recorded YBOCS-II ratings are displayed, with timepoints included in this analysis indicated by circles. Not all OCD symptom severity assessments were paired with neural recordings, as some were conducted remotely without neural recordings or were excluded due to recording artifacts.



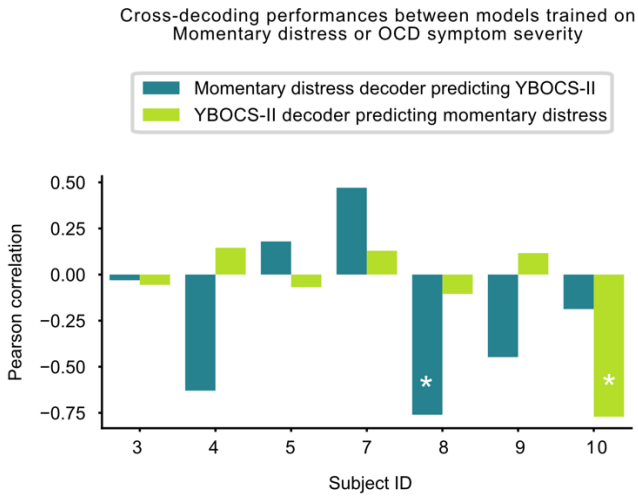
Supplementary Figure 2. Patient-specific neural feature correlation maps. (a,b) Univariate neural feature correlations with momentary distress (a) and longitudinal OCD symptom severity (b). (c,d) Connectivity feature correlations with momentary distress (c) and longitudinal OCD symptom severity (d). Significant correlations after multiple-comparison correction are indicated by (*).



Supplementary Figure 3. Patient-specific DBS parameter changes across the clinical trial. Stimulation amplitude, pulse width, and total electrical energy delivered (TEED; assuming a constant 1-kΩ impedance) are shown separately for the left and right hemispheres across the study duration for each patient.



Supplementary Figure 4. Neural feature changes within and across clinical visits. (a) At each clinic visit, resting-state (RS) recordings were acquired immediately before and after DBS programming, creating a within-visit control that isolated the acute effects of stimulation parameter adjustments from longitudinal changes in clinical state. Changes in neural features observed within visits ($RS1_t$ and $RS2_t$) were then compared with those observed across consecutive visits ($RS2_{t-1}$ and $RS1_t$). (b) Because stimulation parameters were adjusted only during clinical visits, recordings acquired across consecutive visits were obtained under identical stimulation settings, with total electrical energy delivered (TEED) remaining unchanged (ΔTEED=0), whereas within-visit comparisons captured neural activity both before and after DBS programming changes. (c) Distributions of changes in individual neural features shown for three comparisons: within-visit recordings before and after DBS programming when stimulation parameters changed (blue), within-visit recordings without stimulation parameter adjustments (green), and across consecutive clinic visits acquired under identical stimulation settings (gray). Although DBS programming produced measurable changes in neural features, longitudinal changes observed across visits were significantly greater than within-visit changes regardless of whether TEED changed (ΔTEED=0: Bernoulli test number of significant changes across features $p=1.7\cdot10^{-26}$; ΔTEED≠0: $p=1.5\cdot10^{-17}$).



Supplementary Figure 5. Cross-decoding analysis between momentary distress and longitudinal OCD symptom severity. Participant-specific cross-decoding performances for models trained to predict momentary distress and tested on longitudinal OCD symptom severity (blue), and trained on longitudinal OCD symptom severity and tested on momentary distress (green).

Competing interests

S.A.S. has been a consultant for Boston Scientific, Zimmer Biomet, Koh Young, Sensoria Therapeutics, Varian Medical Systems, Abbott and Neuropace and is co-founder of Motif Neurotech. W.K.G. receives royalties from Nview, LLC and OCDscales, LLC. E.A.S. reports receiving research funding to his institution from the International OCD Foundation, Wellcome Trust, and NIH. He receives direct funding from the International OCD Foundation as well as MHNTI for providing trainings on treating obsessive-compulsive disorder with psychotherapy. Furthermore, Dr. Storch co-founded Rethinking Behavioral Health which provides training and consultation in the treatment of obsessive-compulsive disorder and related conditions. He was a consultant for Brainsway and Biohaven Pharmaceuticals in the past 36 months. He owns stock options less than $5000 in NView (for distribution of the Y-BOCS and CY-BOCS) and receives royalties from OCD Scales LLC (for distribution of the Y-BOCS and CY-BOCS). He receives book royalties from Elsevier, Wiley, Oxford, American Psychological Association, Guildford, Springer, Routledge, and Jessica Kingsley.

Acknowledgements

This work relied heavily on the community expertise and resources made available by the Open Mind Consortium (https://openmind-consortium.github.io/). Summit RC + S systems were donated by Medtronic, Inc. as part of the BRAIN Initiative Public-Private Partnership Program. This research was supported by the National Institutes of Health (NIH) National Institute of Neurological Disorders and Stroke BRAIN Initiative via contract UH3NS100549 (to S.A.S., W.K.G., N.R.P., and E.A.S.), the National Institutes of Health (NIH) National Institute of Mental Health (NIMH) R01MH139889 (N.R.P.) and NIH BRAIN Initiative via UH3NS136631 (W.K.G., N.R.P., S.A.S.), the McNair Foundation (S.A.S., N.R.P.), the Gordon and Mary Cain Pediatric Neurology Research Foundation (S.A.S.) and Brain and Behavior Research Foundation Young Investigator Award (N.R.P).

Contributions

T.M., N.R.P. and S.A.S. conceived the study. T.M. conceptualized data analysis procedures, performed data analysis, interpreted data and prepared figures and results, with support from N.R.P., S.A.S., G.B., R.H., B.R., H.Y., J.A., M.O., T.H., V.B., S.S., G.R., S.V.R., and T.F.. A.D.W. performed data collection during ERP therapy. M.A.-O., W.K.G., N.R.P., and S.V.R. performed data collection in the clinic. T.M. wrote the first draft of the manuscript, and all authors contributed to the writing and revision of the manuscript. W.K.G., S.A.S., and E.A.S. performed the clinical care aspects of the study. S.A.S. performed the study surgical procedures. N.R.P. oversaw data collection, analysis and manuscript completion.

References

1. Abramowitz, J. S., Taylor, S. & McKay, D. Obsessive-compulsive disorder. *The Lancet* **374**, 491–499 (2009).
2. Ruscio, A. M., Stein, D. J., Chiu, W. T. & Kessler, R. C. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Mol Psychiatry* **15**, 53–63 (2010).
3. Stein, D. J. *et al.* Obsessive–compulsive disorder. *Nat Rev Dis Primers* **5**, 52 (2019).
4. Goodman, W. K. *et al.* The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry* **46**, 1006–1011 (1989).
5. Vogt, G. S., Avendaño-Ortega, M., Schneider, S. C., Goodman, W. K. & Storch, E. A. Optimizing Obsessive-Compulsive Symptom Measurement with the Yale-Brown Obsessive Compulsive Scales, Second Edition. *J Psychiatr Pract* **28**, 294–309 (2022).
6. Foa, E. B. & McLean, C. P. The Efficacy of Exposure Therapy for Anxiety-Related Disorders and Its Underlying Mechanisms: The Case of OCD and PTSD. *Annual Review of Clinical Psychology* **12**, 1–28 (2016).
7. Gillan, C. M. & Robbins, T. W. Goal-directed learning and obsessive-compulsive disorder. *Philos Trans R Soc Lond B Biol Sci* **369**, 20130475 (2014).
8. Graybiel, A. M. Habits, Rituals, and the Evaluative Brain. *Annual Review of Neuroscience* **31**, 359–387 (2008).
9. Arbab, T. *et al.* Intracranial electrophysiological biomarkers of compulsivity in obsessive–compulsive disorder. *Nat. Mental Health* **3**, 889–898 (2025).

10. Figee, M. *et al.* Deep brain stimulation restores frontostriatal network activity in obsessive-compulsive disorder. *Nat Neurosci* **16**, 386–387 (2013).
11. Nho, Y.-H. *et al.* Human orbitofrontal neural activity is linked to obsessive-compulsive behavioral dynamics. *Cell* **189**, 739-747.e8 (2026).
12. Provenza, N. R. *et al.* Long-term ecological assessment of intracranial electrophysiology synchronized to behavioral markers in obsessive-compulsive disorder. *Nat Med* **27**, 2154–2164 (2021).
13. Provenza, N. R. *et al.* Disruption of neural periodicity predicts clinical response after deep brain stimulation for obsessive-compulsive disorder. *Nat Med* **30**, 3004–3014 (2024).
14. Vissani, M. *et al.* Pallidal alpha activity is an electrophysiological biomarker of symptom severity in obsessive-compulsive disorder. Preprint at https://doi.org/10.21203/rs.3.rs-9316385/v1 (2026).
15. Kabotyanski, K. E., Provenza, N. R. & Sheth, S. A. Intracranial Neural Biomarkers of Psychiatric Symptoms and Their Utility for Guiding Neuromodulation Therapy: A Systematic Review. *Biol Psychiatry* S0006-3223(25)01525–2 (2025) doi:10.1016/j.biopsych.2025.10.005.
16. Stanslaski, S. *et al.* Sensing data and methodology from the Adaptive DBS Algorithm for Personalized Therapy in Parkinson's Disease (ADAPT-PD) clinical trial. *npj Parkinsons Dis.* **10**, 1–11 (2024).
17. Fontenelle, L. F. Mental Status Examination in the Twenty-First Century Psychiatry. *Psychiatric Clinics* **0**, (2026).

18. Bersani, G. *et al.* Comparison of facial expression in patients with obsessive-compulsive disorder and schizophrenia using the Facial Action Coding System: a preliminary study. *Neuropsychiatr Dis Treat* **8**, 537–547 (2012).
19. Clemmensen, L. K. H. *et al.* Associations Between the Severity of Obsessive-Compulsive Disorder and Vocal Features in Children and Adolescents: Protocol for a Statistical and Machine Learning Analysis. *JMIR Res Protoc* **11**, e39613 (2022).
20. Hinduja, S. *et al.* Multimodal Prediction of Obsessive-Compulsive Disorder and Comorbid Depression Severity and Energy Delivered by Deep Brain Electrodes. *IEEE Trans Affect Comput* **15**, 2025–2041 (2024).
21. Mathis, M. W. & Mathis, A. Joint modelling of brain and behaviour dynamics with artificial intelligence. *Nat. Rev. Neurosci.* **27**, 87–100 (2026).
22. Gadot, R. *et al.* Efficacy of deep brain stimulation for treatment-resistant obsessive-compulsive disorder: systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* **93**, 1166–1173 (2022).
23. Snyderman, D. & Rovner, B. W. Mental Status Examination in Primary Care: A Review. *AFP* **80**, 809–814 (2009).
24. Merk, T. *et al.* Invasive neurophysiology and whole brain connectomics for neural decoding in patients with brain implants. *Nat Biomed Eng* https://doi.org/10.1038/s41551-025-01467-9 (2025) doi:10.1038/s41551-025-01467-9.
25. Sheth, S. A. *et al.* Early Feasibility Study of Sensing-enabled Ventral Capsule Deep Brain Stimulation in 10 Participants with Intractable Obsessive-Compulsive Disorder. 2026.04.24.26351318 Preprint at https://doi.org/10.64898/2026.04.24.26351318 (2026).

26. Mataix-Cols, D. *et al*. Towards an international expert consensus for defining treatment response, remission, recovery and relapse in obsessive-compulsive disorder. *World Psychiatry* **15**, 80–81 (2016).
27. Merk, T. *et al*. py_neuromodulation: Signal processing and decoding for neural electrophysiological recordings. *Journal of Open Source Software* **10**, 8258 (2025).
28. Harrison, B. J. *et al*. Altered Corticostriatal Functional Connectivity in Obsessive-compulsive Disorder. *Arch Gen Psychiatry* **66**, 1189–1200 (2009).
29. Smith, E. E. *et al*. A brief demonstration of frontostriatal connectivity in OCD patients with intracranial electrodes. *NeuroImage* **220**, 117138 (2020).
30. Eyben, F., Wöllmer, M. & Schuller, B. Opensmile: the munich versatile and fast open-source audio feature extractor. in *Proceedings of the 18th ACM international conference on Multimedia* 1459–1462 (Association for Computing Machinery, New York, NY, USA, 2010). doi:10.1145/1873951.1874246.
31. Wagner, J. *et al*. Dawn of the transformer era in speech emotion recognition: closing the valence gap. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 10745–10759 (2023).
32. Luo, C., Song, S., Xie, W., Shen, L. & Gunes, H. Learning Multi-dimensional Edge Feature-based AU Relation Graph for Facial Action Unit Recognition. in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence* 1239–1246 (International Joint Conferences on Artificial Intelligence Organization, Vienna, Austria, 2022). doi:10.24963/ijcai.2022/173.
33. Baevski, A., Zhou, Y., Mohamed, A. & Auli, M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. in *Advances in Neural*

Information Processing Systems vol. 33 12449–12460 (Curran Associates, Inc., 2020).

34. Schneider, S., Lee, J. H. & Mathis, M. W. Learnable latent embeddings for joint behavioural and neural analysis. *Nature* **617**, 360–368 (2023).
35. Merk, T. *et al.* Multitarget brain implants enable generalized decoding of Parkinson's disease symptoms from chronic home recordings. Preprint at https://doi.org/10.21203/rs.3.rs-9125364/v1 (2026).
36. Merk, T. *et al.* Electrocorticography is superior to subthalamic local field potentials for movement decoding in Parkinson's disease. *eLife* **11**, e75126 (2022).
37. Oehrn, C. R. *et al.* Chronic adaptive deep brain stimulation versus conventional stimulation in Parkinson's disease: a blinded randomized feasibility trial. *Nat Med* 1–12 (2024) doi:10.1038/s41591-024-03196-z.
38. Kabotyanski, K. E. *et al.* Towards objective, temporally resolved neurobehavioral predictors of emotional state. *Brain Stimul* **17**, 1208–1212 (2024).
39. Webb, C. A. *et al.* Personalized prediction of negative affect in individuals with serious mental illness followed using long-term multimodal mobile phenotyping. *Transl Psychiatry* **15**, 174 (2025).
40. Alexander, G. E., DeLong, M. R. & Strick, P. L. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* **9**, 357–381 (1986).
41. Herron, J. *et al.* Challenges and opportunities of acquiring cortical recordings for chronic adaptive deep brain stimulation. *Nat. Biomed. Eng* **9**, 606–617 (2025).
42. Westen, M. van *et al.* Optimizing Deep Brain Stimulation Parameters in Obsessive–Compulsive Disorder. *Neuromodulation* **24**, 307–315 (2021).

43. Sheth, S. A., Rolston, J. D., Goodman, W. K. & Provenza, N. R. Is Psychiatry Ready for Closed-Loop Invasive Neuromodulation? *JAMA Psychiatry* **82**, 1163–1164 (2025).

44. Anderson, D. N. *et al.* Closed-loop stimulation in periods with less epileptiform activity drives improved epilepsy outcomes. *Brain* **147**, 521–531 (2023).

45. Baud, M. O. *et al.* Multi-day rhythms modulate seizure risk in epilepsy. *Nat Commun* **9**, 88 (2018).

46. Khambhati, A. N., Shafi, A., Rao, V. R. & Chang, E. F. Long-term brain network reorganization predicts responsive neurostimulation outcomes for focal epilepsy. *Sci Transl Med* **13**, eabf6588 (2021).

47. Alagapan, S. *et al.* Cingulate dynamics track depression recovery with deep brain stimulation. *Nature* **622**, 130–138 (2023).

48. Scangos, K. W. *et al.* Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nat Med* **27**, 1696–1700 (2021).

49. Xiao, J. *et al.* Decoding Depression Severity from Intracranial Neural Activity. *Biol Psychiatry* S0006-3223(23)00048–3 (2023) doi:10.1016/j.biopsych.2023.01.020.

50. Hollunder, B. *et al.* A Multi-Symptom Circuit Architecture of Obsessive–Compulsive Disorder. 2025.12.14.25342121 Preprint at https://doi.org/10.64898/2025.12.14.25342121 (2025).

51. Alarie, M. E. *et al.* Artifact characterization and mitigation techniques during concurrent sensing and stimulation using bidirectional deep brain stimulation platforms. *Front Hum Neurosci* **16**, 1016379 (2022).

52. Provenza, N. R. *et al.* High Beta Power in the Ventrolateral Prefrontal Cortex Indexes Human Approach Behavior: A Case Study. *J. Neurosci.* **45**, (2025).

53. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).

54. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* **17**, 825–841 (2002).

55. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* **5**, 143–156 (2001).

56. Joshi, A. *et al.* Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics* **9**, 69–84 (2011).

57. Neudorfer, C. *et al.* Lead-DBS v3.0: Mapping deep brain stimulation effects to local anatomy and global networks. *NeuroImage* **268**, 119862 (2023).

58. Stanslaski, S. *et al.* A Chronically-Implantable Neural Coprocessor for Investigating the Treatment of Neurological Disorders. *IEEE Trans Biomed Circuits Syst* **12**, 1230–1245 (2018).

59. Sellers, K. K. *et al.* Analysis-rcs-data: Open-Source Toolbox for the Ingestion, Time-Alignment, and Visualization of Sense and Stimulation Data From the Medtronic Summit RC+S System. *Front. Hum. Neurosci.* **15**, (2021).

60. Diab, N. *et al.* Characterizing Deep Brain Stimulation Dual Device Beat Frequency Artifacts. 2025.10.11.25337803 Preprint at https://doi.org/10.1101/2025.10.11.25337803 (2025).

61. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).

62. Donoghue, T. *et al.* Parameterizing neural power spectra into periodic and aperiodic components. *Nat Neurosci* **23**, 1655–1665 (2020).

63. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front Neurosci* **7**, 267 (2013).

64. Cole, S. & Voytek, B. Cycle-by-cycle analysis of neural oscillations. *J Neurophysiol* **122**, 849–861 (2019).

65. Cole, S. R. & Voytek, B. Brain Oscillations and the Importance of Waveform Shape. *Trends Cogn Sci* **21**, 137–149 (2017).

66. Jackson, N., Cole, S. R., Voytek, B. & Swann, N. C. Characteristics of Waveform Shape in Parkinson's Disease Detected with Scalp Electroencephalography. *eNeuro* **6**, ENEURO.0151-19.2019 (2019).

67. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology* https://doi.org/10.1016/0013-4694(70)90143-4 (1970) doi:10.1016/0013-4694(70)90143-4.

68. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* **25**, 120–123 (2000).

69. Bain, M., Huh, J., Han, T. & Zisserman, A. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. Preprint at https://doi.org/10.48550/arXiv.2303.00747 (2023).

70. Eyben, F. *et al.* The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing* **7**, 190–202 (2016).

71. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

72. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference* **2010**, (2010).