





Mood variations decoded from multi-site intracranial human brain activity

Omid G Sani^{1,6}, Yuxiao Yang^{1,6}, Morgan B Lee²⁻⁴, Heather E Dawes²⁻⁴, Edward F Chang^{2-4,7}
& Maryam M Shanechi^{1,5,7}

The ability to decode mood state over time from neural activity could enable closed-loop systems to treat neuropsychiatric disorders. However, this decoding has not been demonstrated, partly owing to the difficulty of modeling distributed mood-relevant neural dynamics while dealing with the sparsity of mood state measurements. Here we develop a modeling framework to decode mood state variations from multi-site intracranial recordings in seven human subjects with epilepsy who self-reported their mood state intermittently over multiple days. We built dynamic neural encoding models of mood state and corresponding decoders for each individual and demonstrated that mood state variations over time can be decoded from neural activity. Across subjects, the decoders largely recruited neural signals from limbic regions, whose spectro-spatial features were tuned to mood variations. The dynamic models also provided an analytical tool to compute the timescales of the decoded mood state. These results provide an initial line of evidence indicating the feasibility of mood state decoding.

The design of mood state decoders, which requires modeling large-scale neural dynamics, is critical for developing effective therapies for mood disorders, but has been impeded by several challenges. Neurobiological, neuroanatomical and neuroimaging studies suggest that neural circuits underlying mood representation do not reside in a single brain region, but rather involve multiple, distributed cortico-limbic regions¹⁻¹¹. Moreover, tracking mood state variations over time is difficult given the complex nature of mood^{12,13}. Thus, modeling and decoding would require recording neural activity across multiple distributed brain sites while simultaneously obtaining mood measurements. Also, these goals would require new computational techniques that can incorporate data from large-scale distributed brain sites while dealing with the sparsity in available mood state measurements, caused by the difficulty of mood assessment. Therefore, decoding mood state from neural activity has so far remained elusive.

Given these challenges, functional representations of mood in the human brain have been largely studied in experimental settings using noninvasive neuroimaging^{2,4,5,7-11}. These studies have shown regional changes induced by emotional stimuli in healthy subjects^{4,7,10} and identified altered resting-state activity that may be related to neural circuit dysfunctions or treatment effects in mood disorders^{4,5,8,11}. Also, studies using noninvasive electroencephalogram and localized intracranial local field potential (LFP) have shown neurophysiological changes that distinguish between depressed and healthy subjects¹⁴⁻¹⁶. Informed by these findings, seminal work has been done on open-loop deep brain stimulation (DBS) for treatment-resistant depression¹⁷⁻¹⁹.

While open-loop treatments including DBS have been promising¹⁷⁻¹⁹, precisely tailored therapies for mood disorders would require a closed-loop approach in which an objective neural decoding of mood state guides treatment (e.g., electrical stimulation) in real time. However, an obstacle to tailored therapies is the lack of methodology to decode mood. The precise spatiotemporal characteristics of mood-relevant neural activity in the distributed cortico-limbic networks are not well understood^{4,8,9}; also, the tuning of neural activity to mood may be different depending on factors such as the subject's psychiatric conditions (depression vs. anxiety vs. good health)^{2,20-22}. These aspects create a challenging decoding problem. Also, continuous long-term mood tracking may not be practical in current neuroimaging settings, and localized LFP leads may hinder neural recording from the relevant distributed cortico-limbic regions.

Here we continuously recorded large-scale intracranial electrocorticogram (ECoG) signals²³⁻²⁵ and simultaneously collected self-reported mood state measurements²⁶ over multiple days in subjects with epilepsy. We devised a modeling framework that can use the sparse mood state measurements to identify a parsimonious mood-predictive network from the high-dimensional neural recordings in each subject and train a dynamic neural encoding model within the identified network. We used the trained model to develop decoders that could predict mood state over time from neural spectral features in each subject. These mood decoders represent a first step toward facilitating future personalized closed-loop therapies for neuropsychiatric disorders.

¹Ming Hsieh Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, California, USA. ²Department of Neurological Surgery, University of California, San Francisco, California, USA. ³Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California, USA. ⁴Kavli Institute for Fundamental Neuroscience, University of California, San Francisco, San Francisco, California, USA. ⁵Neuroscience Graduate Program, University of Southern California, Los Angeles, California, USA. ⁶These authors contributed equally to this work. ⁷These authors served as co-senior authors, with M.M.S. leading the decoding work and E.F.C. leading the data collection. Correspondence should be addressed to M.M.S. (shanechi@usc.edu) or E.F.C. (edward.chang@ucsf.edu).

Received 27 June 2017; accepted 28 June 2018; published online 10 September 2018; doi:10.1038/nbt.4200

RESULTS

Neural recording and mood state measurement

We continuously recorded large-scale intracranial activity across multiple days from seven human subjects with epilepsy who had intracranial electrodes implanted for seizure localization (see Online Methods and **Supplementary Tables 1** and **2**). Our recordings covered multiple limbic regions, including the orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), insular cortex, amygdala, and hippocampus, as well as other regions in frontal, temporal and parietal cortices.

Subjects' mood state was measured at discrete times during multiple days of neural recording using a validated tablet-based self-report questionnaire termed Immediate Mood Scaler (IMS)²⁶ (12 ± 2.4 IMS data points (reports) across an average of 6.0 ± 4.5 d per subject; mean \pm s.d.; Online Methods and **Supplementary Table 3**). IMS provides a momentary assessment of a set of mood states related to depression and anxiety symptoms²⁶ and served as our operational definition of mood for developing interventional decoders (**Supplementary Note 1**). IMS score is based on sum of answers to 24 questions (**Supplementary Table 4**) and lies between -72 and 72 (Online Methods, **Supplementary Note 1**). IMS variations covered 73% of the total possible IMS range across our subjects and 33 \pm 7.2% of this range within each subject (**Supplementary Fig. 1** and **Supplementary Table 3**), reflecting meaningful mood state variations over multiple recording days.

Modeling and decoding framework

We aimed to build a decoder for each subject that generalized to prediction of a new IMS point (**Fig. 1**). As the main neural features, we selected powers in various frequency bands: delta plus theta (1–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), gamma-1 (30–55 Hz) and gamma-2 (65–100 Hz) (Online Methods). Neural features were high-dimensional (224 ± 99 features; **Supplementary Table 2**). Because IMS points are sparse due to difficulties of mood assessment (i.e., mood is not directly observable and cannot be measured frequently), training a generalizable decoder presents a challenging modeling problem. We evaluated our decoders using a rigorous leave-one-out cross-validation in which we left out an IMS point to be predicted (test IMS) and used the rest of the IMS points (training IMS) to train the decoder (**Fig. 1a**); this cross-validated train–test procedure was repeated for all IMS points. The test IMS was never seen by any step of decoder training. Thus, if the decoder overfitted to the sparse training IMS, it would not generalize to the test IMS and would consequently fail for decoding in cross-validation.

To resolve the challenge of modeling distributed neural activity while dealing with IMS sparsity, we decided to restrict the number of model parameters that need to be fitted using the sparse IMS points to be less than the number of available IMS points. We constructed a two-component neural encoding model to relate neural activity to IMS (**Fig. 1b**). Training this model fully specified the decoder. The first component extracted a small number of hidden variables from neural activity by performing dimensionality reduction. The second component linearly regressed only these hidden variables, termed neural predictors, to IMS points. Given that for a given neural encoding model only the regression parameters are fitted using the IMS points, during dimensionality reduction we restricted the number of neural predictors—which is equal to the number of regression parameters—to be smaller than the number of available IMS points (Online Methods). We designed our neural encoding models to be dynamic to describe mood state variations over time in terms of the time variations (i.e., dynamics) of neural activity.

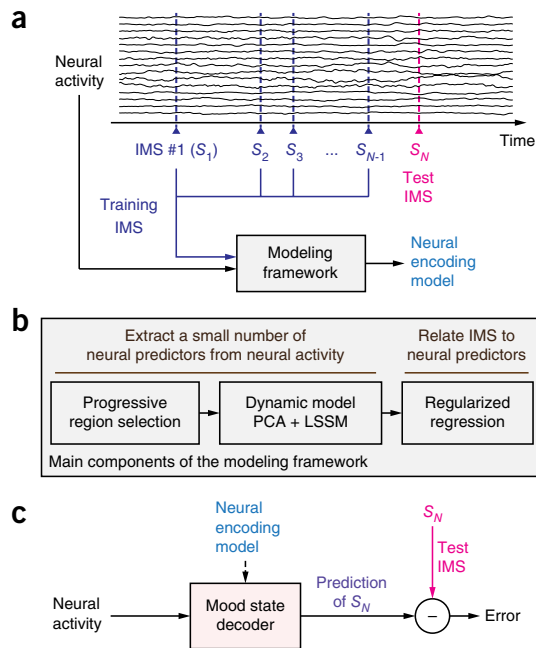


Figure 1 Illustration of decoder training and evaluation using cross-validation. (a) In cross-validation, an IMS point (as an example, S_N) is left out as the test IMS to be predicted. The other IMS points (i.e., training IMS, here S_1 to S_{N-1}) and the associated neural activity are used within the modeling framework to train a neural encoding model (details in **Supplementary Fig. 2**). (b) To enable training generalizable decoders, multiple components within the modeling framework reduce the number of model parameters that need to be fitted using the IMS points (i.e., regression parameters). Progressive region selection selects only a small subset of regions for use in the decoder. Dynamic modeling using principal component analysis (PCA) and LSSM describes the temporal dynamics of neural activity in the selected network through a low-dimensional hidden neural state, which constitutes the neural predictors. A regularized regression model then relates this small number of neural predictors to mood state. Regularization further reduces the effective number of regression parameters. (c) The trained neural encoding model is used to build a decoder that predicts mood state on the basis of the observed neural activity. The decoder is used to predict the test IMS (S_N). We calculate the prediction error as the difference between the test IMS value and its prediction. The above procedure is performed once for each IMS, thus forming leave-one-out cross-validation. Decoding performance is evaluated using the average cross-validation error.

To reduce dimensionality in the first component, we used several steps (**Fig. 1b** and **Supplementary Fig. 2**). First, since mood is likely represented across multiple distributed brain regions^{1–11}, we hypothesized that it may be sufficient to model and use a small subset of these regions for decoding. Thus, instead of using the large number of neural features, we developed a progressive region selection method to identify and model only the smallest network of brain regions that were sufficient for decoding (Online Methods and **Supplementary Fig. 2a**). This small network was selected by the modeling framework purely on the basis of the training data and was selected to minimize both sensitivity to and prediction error for this training data (Online Methods and **Supplementary Fig. 2c**). Second, even within a small network and unsupervised with respect to IMS points (i.e., only using the neural data consisting of multiple days of ECoG recordings; Online Methods), we trained a dynamic model for neural features to describe their time variations in terms

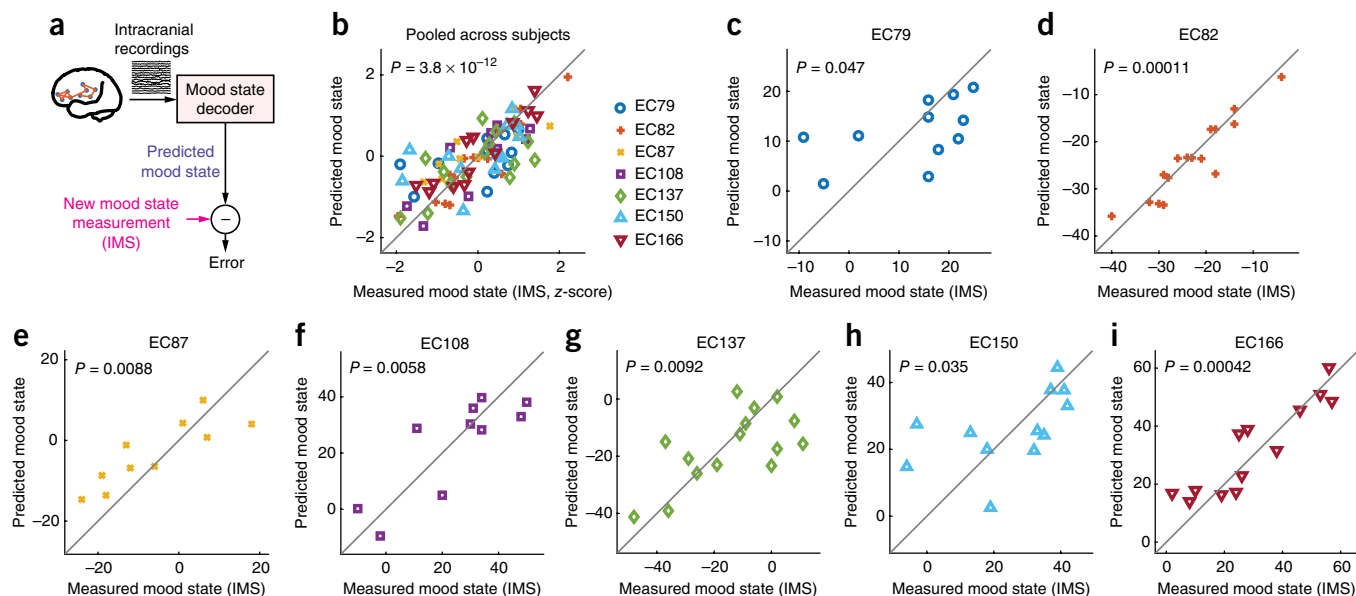


Figure 2 Mood state could be decoded from neural activity in each subject. **(a)** Overview of the decoding process. The decoder can process the intracranial recordings causally over time to predict mood state. At discrete time points when a new mood state measurement is available, the subtraction of the predicted from the measured mood state gives the prediction error. **(b–i)** Cross-validated prediction of the mood state is shown against the true measured mood state, pooled across subjects in **b** and within individuals in **c–i**. Random-test P value is noted on each plot ($P < 0.05$ for all after FDR correction). Number of samples in **b** is 87 and in **c–i** is equal to the number of IMS points for each subject (**Supplementary Table 3**).

of a low-dimensional hidden neural state. The dynamic model first applied principal component analysis²⁷ to neural features and then identified the low-dimensional hidden neural state by training a linear state-space model (LSSM)²⁸ on the retained principal components (Online Methods and **Supplementary Fig. 2b**). This hidden neural state constituted the neural predictors; thus we restricted its dimension to be similar to or smaller than the number of available IMS points. The second component consisted of a regularized linear regression from neural predictors to IMS points²⁷ (Online Methods and **Supplementary Fig. 2b**). Together, these components avoided overfitting to the training IMS and thus enabled generalizable decoding (**Supplementary Note 2**).

Once the neural encoding model was fitted and selected using the training data, we used it to build the decoder that predicted the test IMS (**Fig. 1c** and **Supplementary Fig. 2d**). Optimal decoding involved using a Kalman filter to estimate the hidden neural state at the time of the test IMS and then using the regression model to predict the test IMS from the estimated neural state, all within cross-validation. We obtained an average cross-validated IMS prediction error across all test IMS points (Online Methods). To evaluate significance, we applied the same modeling framework to IMS points randomly drawn from the same range as the true IMS in each subject (random-test) and to time-permuted IMS points (permuted-test; Online Methods). We defined the random or permuted P value as the probability that the random or permuted IMS points will have lower cross-validated prediction error than the true IMS points.

Mood state variations over time can be decoded in each subject

We built neural decoders of mood state for each subject (**Fig. 2**). We found that in each of the seven subjects, the decoders were significantly predictive of the IMS points according to both evaluation tests (**Fig. 2c–i** and **Table 1**, $P \leq 0.05$ for all subjects; all P values were corrected with false discovery rate (FDR) control if needed; Online

Methods). The decoders were also highly predictive of the IMS points at the population level (**Fig. 2b** random-test $P = 3.8 \times 10^{-12}$ and permuted-test $P = 1.1 \times 10^{-15}$; **Supplementary Figs. 3** and **4**; Online Methods). Finally, we confirmed the robustness of our modeling framework to the choice of neural features by decoding IMS using coherence features (**Supplementary Note 3** and **Supplementary Table 5**). These results suggest that the framework can tap into distributed neural representations to decode mood state.

Once trained, the same decoder could be used for mood state prediction across hours and days. The time-delay between the first and the last IMS points was on average 6 d and between consecutive IMS points was on average 13 h (**Supplementary Table 3**). Since an IMS point that is decoded is not used for training the decoder (cross-validation), our results show that the same decoder could be used hours or days apart from the data used for training it. Further, even IMS points that were the only report obtained in a day were significantly predicted across subjects (random-test, $P = 1.4 \times 10^{-3}$; **Supplementary Fig. 5**). Moreover, the time-distance from a test IMS to the closest training IMS was not correlated with its prediction error (Spearman's $P = 0.99$ across the population and $P > 0.15$ in every subject; **Supplementary Note 2**).

The decoders could also generalize across a wide range of IMS. First, in cross-validation, the decoders could predict IMS variations that covered 73% and $33 \pm 7.2\%$ of the total possible IMS range across our subjects and within individuals, respectively. Also, we could decode the minimum and the maximum IMS points in each subject when they were used as the test IMS points and thus were outside the range of the training IMS points (random-test $P = 1.3 \times 10^{-6}$; **Supplementary Fig. 6**). Further, the same cross-validated modeling framework could separately decode the depression and anxiety subscales of IMS (**Supplementary Note 1**) using the same networks that were predictive of the total IMS (**Supplementary Fig. 7**; random-test $P = 1.1 \times 10^{-6}$ and $P = 5.9 \times 10^{-7}$ across subjects, respectively).

Table 1 Summary of cross-validated IMS prediction accuracy of the mood state decoders and the spatial distribution of the best small mood-predictive networks

Subject	NRMSE × 100%	P value [FDR-corrected if different]		Regions in mood-predictive network
		Random-test	Permuted-test	
EC79	78.5%	0.047	0.042	OFC
EC82	35.1%	0.00011	0.00013	OFC
EC87	56.0%	0.0088 [0.035]	0.012 [0.050]	OFC, dACC, vACC
EC108	51.5%	0.0058	0.0059	AMYG
EC137	72.8%	0.0092 [0.037]	0.0075 [0.030]	dACC, HPC, sFG, mFG
EC150	78.5%	0.035	0.045	HPC
EC166	40.3%	0.00042	0.0027	OFC
Population	68.3%	3.8×10^{-12}	1.1×10^{-15}	–

Number of IMS samples for the population is 87; number for each subject is given in **Supplementary Table 3**. NRMSE, normalized root mean square error; OFC, orbitofrontal cortex; dACC, dorsal anterior cingulate cortex; vACC, ventral anterior cingulate cortex; AMYG, amygdala; sFG, superior frontal gyrus; mFG, middle frontal gyrus; HPC, hippocampus.

Finally, as a control, we found that interictal discharge rates around the time of IMS in the selected networks used for decoding were not significantly predictive of mood state in any subject ($P > 0.18$ for random-test and permuted-test in all subjects; Discussion and **Supplementary Note 4**). Taken together, these results suggest that our decoders can predict mood state variations from neural activity across multiple days of recordings in each subject.

Recruitment of the limbic regions for mood state prediction

Three analyses demonstrated that the decoder largely recruited the limbic regions. First, in our main analysis above to establish decoding ability, the modeling framework first searched the limbic regions to construct the decoder, given their important role in mood representation^{2,4,5,7–11}. Only if the limbic regions alone were not sufficient for decoding did the method then search all available regions (while using FDR correction; Online Methods). In six of seven subjects, networks within the limbic regions alone were predictive of mood state; in a single subject (EC137), progressive region selection proceeded to search all regions to achieve prediction (**Table 1**).

This main analysis also found the smallest network size that was sufficient for decoding in each subject (Online Methods); next, we identified the best mood-predictive network with that size for each individual (**Table 1**, **Fig. 3**, **Supplementary Fig. 8** and **Supplementary Table 6**). The most commonly recurring region was OFC, which was included in the network in four of seven subjects. OFC alone was selected in EC79, EC82 and EC166, and a distributed network with OFC and ACC was selected in EC87. In EC108 and EC150, amygdala and hippocampus were selected, respectively. Finally, in EC137, the selected network consisted of limbic regions ACC and hippocampus together with an ECoG electrode covering the middle and superior frontal gyri (dorsolateral and dorsomedial prefrontal cortex), which have also been implicated in mood representation^{4,8,9,29}. These results suggest that using a subset of regions that belong to the distributed mood-representation networks is sufficient for decoding^{2,4,5,7–11}. Progressive region selection stops adding more regions as soon as significant prediction is achieved and does not attempt to identify all mood-predictive regions (Online Methods).

Two more analyses further confirmed the role of limbic regions (Online Methods). First, we searched across all electrodes for every subject (**Supplementary Fig. 9**). Even without knowing where these electrodes were, our modeling framework largely selected the same regions for decoding as those selected in the limbic search (**Supplementary Fig. 10** and **Supplementary Table 7**). Second, we found that decoding largely failed when searching within electrodes outside the limbic regions (**Supplementary Table 7** and

Supplementary Fig. 11; subjects included had similar numbers of neural features inside and outside limbic regions). These results further validate our modeling because they are biologically consistent with prior neuroimaging studies^{2,4,5,7–11} that indicate the central role of distributed limbic regions in emotion and mood representation (**Supplementary Note 2**).

Spectro-spatial neural features were tuned to mood variations over time

The spectro-spatial neural features within the identified mood-predictive networks (i.e., powers at different channels and frequency bands) were tuned to mood state variations over time (**Fig. 4**). Certain individual features in all five frequency bands had strong correlations with IMS, some of which were significant after FDR correction (**Fig. 4b,d,f**, $P < 0.05$ after FDR correction). All these significantly tuned features in low frequencies (delta plus theta; alpha; beta) were negatively correlated with IMS. In contrast, the significantly tuned features in high frequencies had both positive and negative correlations with IMS. These results suggest that the decoder predicted mood state by combining information across spectro-spatial features. Consistently, the prediction error when using all five frequency bands was significantly lower than when using single frequency bands alone ($P < 0.05$, one-sided Wilcoxon signed-rank test), suggesting that all frequency bands could contribute to decoding (**Supplementary Fig. 12**).

Calculation of the timescales of decoded mood state

The LSSM together with the regression model facilitates a direct parametric calculation of the power spectral density (PSD) of the decoded mood time-series (**Supplementary Note 5**). The PSD quantifies the relative prominence of different timescales (defined as the inverse of frequency) in the decoded mood state variations. For example, we can compute that more than 70% of the PSD was at timescales of 3.4 ± 2.3 h and slower across subjects (**Supplementary Fig. 13**). We also confirmed that the decoded mood state changed slowly within minutes of a measured IMS point (**Supplementary Fig. 14**). Thus, since our neural encoding models are dynamic, they may provide an analytical tool to study the temporal characteristics of the decoded mood state.

DISCUSSION

Decoding technologies have so far been largely designed for brain states that are well-represented in localized brain areas. For example, in motor brain-machine interfaces (BMIs), given the somatotopic organization of the motor cortex³⁰, motor intentions can be decoded from localized regions in motor cortical areas^{31–45}. Moreover, the

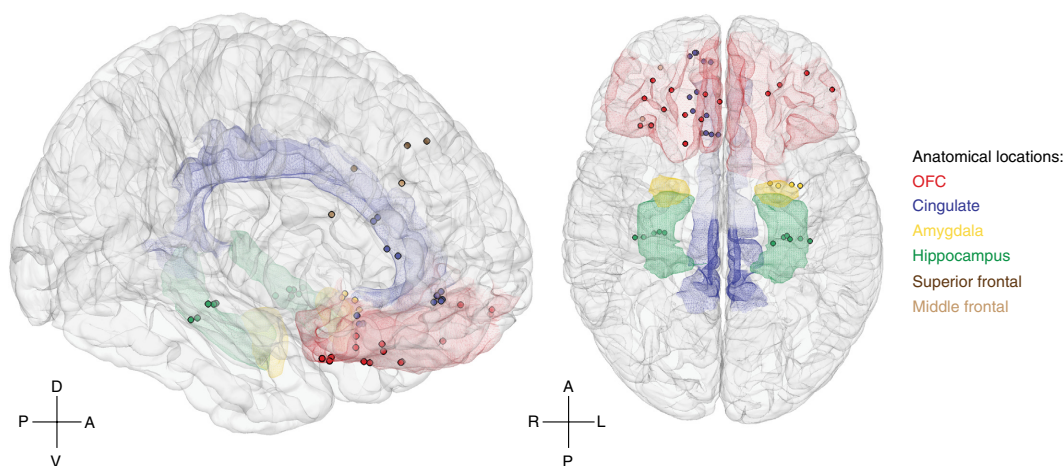


Figure 3 Mood-predictive networks selected for decoding were largely within the limbic regions in all subjects. The location of all recording channels included in the best small mood-predictive network for decoding in each subject are shown on the standard Montreal Neurological Institute template brain. Anterior (A), posterior (P), dorsal (D), ventral (V), left (L) and right (R) directions are indicated. The areas of key limbic regions based on labeling from the FreeSurfer software are shaded with color (Online Methods). Each recording channel is also marked with the color of the anatomical region to which it is assigned. All cingulate recording channels were in ACC. **Supplementary Figure 8** shows the networks selected for each subject.

behavioral measurements required for training these decoders are not sparse, because movement, for example, can be measured or inferred continuously in time. These aspects have facilitated personalized motor decoding in real-time BMIs using neural spiking and LFP^{31–33,36–43}, as well as ECoG^{34,35,44,45}.

Decoding mood poses new modeling challenges as a result of many differences. Mood representation involves multiple distributed brain sites rather than localized regions^{1–11}, and the functional organization of mood-predictive networks is not as well understood^{4,8,9}. Furthermore, mood cannot be behaviorally measured frequently, resulting in sparse measurements at discrete times (e.g., 12 ± 2.4 IMS points per subject here). We resolved the new modeling challenges by designing several components.

To train generalizable decoders with sparse mood state measurements, we restricted the number of parameters that need to be fitted using the IMS points to be smaller than the number of available IMS points (**Supplementary Table 6**). Initially, we applied dimensionality reduction using progressive region selection and dynamic modeling to identify a small number of hidden neural predictors. Then we related only these predictors to IMS with regularized linear regression. Several analyses supported our approach. First, decoding the true IMS was substantially more accurate than decoding an arbitrary random or permuted IMS, which provided the chance level (**Fig. 2**). Second, decoding failed using electrodes outside limbic regions, but succeeded with a similar number of limbic electrodes. Third, misaligning the IMS and neural activity (e.g., by 2 h) caused decoding to fail (**Supplementary Note 2**) despite the same autocorrelations being present, and further, decoding was not more accurate for test IMS points that were closer to training IMS points.

To show the biological consistency and robustness of our modeling framework, we examined the selected regions in the decoders. We found that the decoders recruited largely the same limbic networks in a given subject regardless of the search space, and they largely failed without the limbic regions. Also, all the selected limbic regions in our decoders (i.e., OFC^{2,4,5,8,9}, ACC^{4,5,7,11}, amygdala^{4,5,7,9} and hippocampus^{2,7,9,11}) have been implicated in mood and emotion representation—for example, for population-level mood state classification^{5,7,11}. While some limbic regions were selected in common across some

subjects—for example, OFC in four of the seven subjects—there was considerable heterogeneity in the limbic regions selected across subjects. There are several possible reasons for this. First, across subjects, there were differences in where the electrodes were implanted (**Supplementary Table 2**) and in the mood state ranges, and several neuroimaging studies have suggested that different limbic regions may have differential roles in positive and negative mood state representation^{46,47}. Second, to address the modeling challenge caused by the sparsity of mood state measurements, a key idea in our methodology was to identify and model only the smallest subset of regions that was sufficient for decoding in each subject. This allowed us to obtain a less complex model that could be robustly fitted with the available training data (progressive region selection; Online Methods). It is important to note that we did not attempt to find all mood-predictive regions. Instead, given the evidence that mood is represented across multiple distributed brain regions^{1–11}, we hypothesized that decoding may still be possible even if we only model and use a subset of these regions. For the above reasons, there likely exist additional mood-predictive regions in a given subject. Thus, the regions selected by the modeling framework should be interpreted as the best small networks sufficient for decoding in a subject (given their available electrode coverage), not as the only mood-predictive regions that encode mood in that subject. Future chronic studies may help advance the understanding of brain regions that encode mood and enable their exhaustive investigation across different subjects by obtaining more mood measurements. Such studies may further guide the choice of recording sites and electrode placement in future closed-loop therapies. Finally, just as we trained our mood decoders in each subject, traditional motor BMI decoders^{31–45} also need to be trained in each subject separately and do not apply from one subject to another.

The choice of the most appropriate mood state measurement scale for effective closed-loop stimulation will be critical to study in future work. Here we chose IMS because it could be useful for developing decoders that guide clinical intervention for alleviating depression and anxiety symptoms (**Supplementary Note 1**). First, as IMS provides a momentary assessment, training decoders on IMS could enable tracking acute momentary mood symptom changes that can happen within a short period of minutes in DBS intervention^{17–19}.

Second, given the high comorbidity of depression and anxiety⁴⁸, to guide clinical intervention, it could be useful to decode an overall mood state related to both of these symptoms as measured by IMS. Third, IMS allows reporting both negative and positive mood states to indicate both the alleviation of symptoms and the emergence of a positive mood state due to intervention. However, we emphasize that our modeling framework can be applied to investigate the decoding of other mood measures as well. Mood is a complex construct^{12,13}, and many self-reports, including IMS, can be decomposed into multiple components^{12,26}. Thus, a change in the total IMS score may at times be driven by one component (i.e., subscale) more than the other. We found that the IMS depression and anxiety subscales could also be separately decoded using the same modeling framework (Supplementary Fig. 7), thus potentially enabling their future use in closed-loop DBS.

Because IMS provides a momentary assessment of anxiety and depression symptoms, a clinical intervention aimed at changing the decoded IMS could be beneficial for symptom control. A critical question is whether such an intervention could also shift the baseline level of mood. Mood often refers to a pervasive and sustained emotional state subject to transient fluctuations in affective state over time^{49,50}. Prior validation²⁶ indicates that IMS is also correlated with measures of sustained depression and anxiety symptoms (PHQ-9⁵¹ and GAD-7⁵²; Supplementary Note 1). An area for future investigation is whether and how such sustained measures can be estimated from IMS (e.g., through averaging).

We performed our study in subjects with epilepsy because semi-chronic intracranial recordings are in most cases only possible in this population. It will be important to examine whether our decoding results generalize to other populations. Given the strong comorbidity between epilepsy and mood disorders⁵³, this population provides a good context and opportunity for conducting studies relevant to mood disorders. For example, vagus nerve stimulation was first found effective in alleviating depressive symptoms in patients with epilepsy⁵⁴ and was later generalized as a therapy for treatment-resistant depression⁵⁵. However, a potential confounding factor to control for is the possibility that epilepsy-specific neural activity may be predictive of mood state. In six of seven subjects, the best small mood-predictive network did not have any overlap with the seizure foci, and in the other subject there was only partial overlap between the two (Table 1 and Supplementary Table 1). Nevertheless, the epileptic interictal discharge rate in the selected networks could not predict mood in any subject.

This study has limitations. The rare opportunity and difficulty of simultaneously collecting both intracranial neural recordings and mood state measurements in human subjects over several days resulted in a limited number of subjects. The only subjects who could participate in this study were patients with epilepsy who had intracranial electrodes implanted for their clinical needs, had these electrodes for at least several days, filled the IMS questionnaire at least several times, and experienced a meaningful range of IMS variation during neural recordings. Meeting all the above criteria leaves rare opportunities for subject inclusion. Moreover, given the difficulties of mood assessment in general and the limited time that patients with epilepsy have the intracranial electrodes, the number of mood measurements per subject was limited. This created a challenging modeling problem. Our findings will need to be corroborated in future chronic studies and clinical trials with higher numbers of subjects that are dedicated to studying mood disorders and thus have much longer implantation periods and consequently can have more mood measurements.

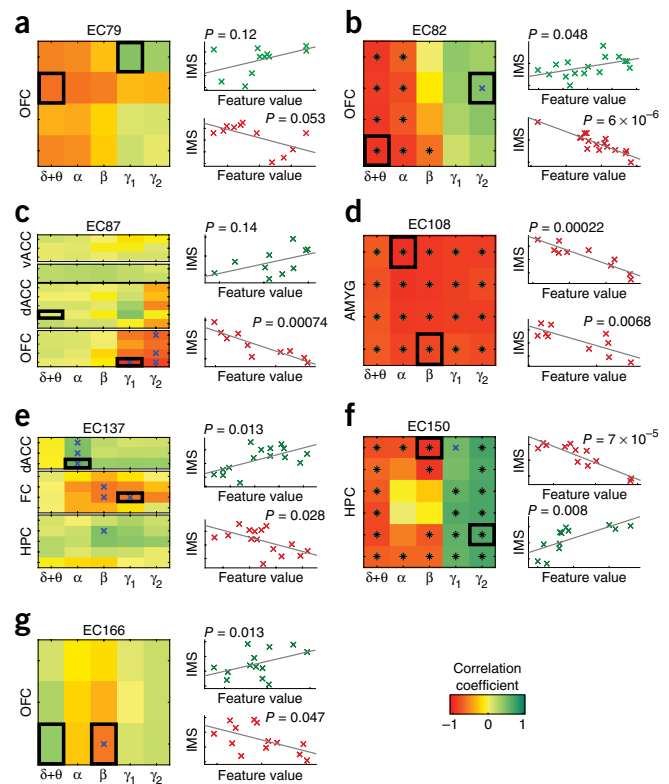


Figure 4 Spectro-spatial features within the mood-predictive networks were tuned to mood state variations over time. (a–g) Correlation of all individual features included in the best small mood-predictive network for decoding in each subject. Horizontal axis shows the frequency bands. Vertical axis shows the recording channels (different regions are shown on separate plots). Color indicates the correlation coefficient value. A positive correlation (green) indicates that feature value increases with an increase in the IMS (better mood) and a negative correlation (red) indicates that the feature value increases with a decrease in the IMS (worse mood). Features with Pearson's $P < 0.05$ are marked with a black asterisk if they remain significant after FDR correction controlled at 0.05 and with a blue cross otherwise. For each subject, two samples (boxed) of tuned feature variations with IMS are plotted on the right (color scheme consistent with the correlation plot). Pearson's correlation P value is noted for each plot. For each subject, number of samples is equal to the number of IMS points (Supplementary Table 3).

Like those related to mood state, neural processes related to neuropsychiatric states implicated in disorders such as chronic pain, addiction or post-traumatic stress disorder are not anatomically localized, but rather span a distributed network of brain regions^{56–58}. Moreover, these neuropsychiatric states are also complex and difficult to track^{56–58}, resulting in sparse behavioral measurements. Given these similar challenges, our modeling framework may generalize to decoding across these neuropsychiatric conditions for future closed-loop treatments.

Future closed-loop electrical stimulation therapies for depression and anxiety may be possible by using the decoded mood state as feedback. To realize such systems, neural encoding models such as the ones we developed here may be extended to predict the effect of stimulation on mood state⁵⁹, thus enabling model-based closed-loop control. Also, future chronic recordings with more mood measurements may allow for modeling larger mood-predictive networks to improve decoding accuracy. Finally, as more data become

available, models may be recalibrated⁶⁰ to maintain accurate prediction over time.

Taken as a whole, our study provides a demonstration that mood state variations in individual subjects can be decoded from large-scale intracranial recordings across multiple days. In the future, advanced personalized closed-loop therapies for neuropsychiatric disorders such as depression and anxiety may be possible.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This research was partially funded by the Defense Advanced Research Projects Agency (DARPA) under Cooperative Agreement Number W911NF-14-2-0043 (to M.M.S. and E.F.C.), issued by the Army Research Office contracting office in support of DARPA's SUBNETS program. The views, opinions and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government.

AUTHOR CONTRIBUTIONS

M.M.S. and E.F.C. supervised the project. M.M.S. conceived the dynamic modeling and decoding framework, and M.M.S., O.G.S. and Y.Y. developed it. E.F.C. and M.B.L. developed and implemented the continuous mood testing protocol and coordinated data collection for all the intracranial recording and imaging data. O.G.S. and Y.Y. implemented and performed the modeling and analyses. M.M.S. supervised all the modeling and analyses work. O.G.S., Y.Y. and M.M.S. wrote the manuscript with input from E.F.C. and H.E.D.

COMPETING INTERESTS

The authors declare no competing interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Calhoun, G.G. & Tye, K.M. Resolving the neural circuits of anxiety. *Nat. Neurosci.* **18**, 1394–1404 (2015).
- Dan, R. *et al.* Separate neural representations of depression, anxiety and apathy in Parkinson's disease. *Sci. Rep.* **7**, 12164 (2017).
- Dmochowski, J.P., Sajda, P., Dias, J. & Parra, L.C. Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Front. Hum. Neurosci.* **6**, 112 (2012).
- Drevets, W.C. Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders. *Curr. Opin. Neurobiol.* **11**, 240–249 (2001).
- Drysdale, A.T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38 (2017).
- Haroush, K. & Williams, Z.M. Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell* **160**, 1233–1245 (2015).
- Kragel, P.A., Knodt, A.R., Hariri, A.R. & LaBar, K.S. Decoding spontaneous emotional states in the human brain. *PLoS Biol.* **14**, e2000106 (2016).
- Kupfer, D.J., Frank, E. & Phillips, M.L. Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *Lancet* **379**, 1045–1055 (2012).
- Mayberg, H.S. Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. *Br. Med. Bull.* **65**, 193–207 (2003).
- Mayberg, H.S. *et al.* Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am. J. Psychiatry* **156**, 675–682 (1999).
- Zeng, L.-L. *et al.* Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain* **135**, 1498–1507 (2012).
- Clark, L.A. & Watson, D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J. Abnorm. Psychol.* **100**, 316–336 (1991).
- Posner, J., Russell, J.A. & Peterson, B.S. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **17**, 715–734 (2005).
- Clark, D.L., Brown, E.C., Ramasubbu, R. & Kiss, Z.H.T. Intrinsic local beta oscillations in the subgenual cingulate relate to depressive symptoms in treatment-resistant depression. *Biol. Psychiatry* **80**, e93–e94 (2016).
- Leuchter, A.F., Cook, I.A., Hunter, A.M., Cai, C. & Horvath, S. Resting-state quantitative electroencephalography reveals increased neurophysiologic connectivity in depression. *PLoS One* **7**, e32508 (2012).
- Neumann, W.J. *et al.* Different patterns of local field potentials from limbic DBS targets in patients with major depressive and obsessive compulsive disorder. *Mol. Psychiatry* **19**, 1186–1192 (2014).
- Malone, D.A. Jr. *et al.* Deep brain stimulation of the ventral capsule/ventral striatum for treatment-resistant depression. *Biol. Psychiatry* **65**, 267–275 (2009).
- Mayberg, H.S. *et al.* Deep brain stimulation for treatment-resistant depression. *Neuron* **45**, 651–660 (2005).
- Schlaepfer, T.E. *et al.* Deep brain stimulation to reward circuitry alleviates anhedonia in refractory major depression. *Neuropsychopharmacology* **33**, 368–377 (2008).
- Clark, D.A. & Beck, A.T. Cognitive theory and therapy of anxiety and depression: convergence with neurobiological findings. *Trends Cogn. Sci.* **14**, 418–424 (2010).
- Grimm, S. *et al.* Imbalance between left and right dorsolateral prefrontal cortex in major depression is linked to negative emotional judgment: an fMRI study in severe major depressive disorder. *Biol. Psychiatry* **63**, 369–376 (2008).
- Lee, B.-T., Lee, B.C., Choi, I.G., Lyoo, I.K. & Ham, B.J. The neural substrates of affective processing toward positive and negative affective pictures in patients with major depressive disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **31**, 1487–1492 (2007).
- Bouchard, K.E., Mesgarani, N., Johnson, K. & Chang, E.F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).
- Korzeniewska, A. *et al.* Ictal propagation of high frequency activity is recapitulated in interictal recordings: effective connectivity of epileptogenic networks recorded with intracranial EEG. *Neuroimage* **101**, 96–113 (2014).
- Yaffe, R.B. *et al.* Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proc. Natl. Acad. Sci. USA* **111**, 18727–18732 (2014).
- Nahum, M. *et al.* Immediate Mood Scaler: tracking symptoms of depression and anxiety using a novel mobile mood scale. *JMIR Mhealth Uhealth* **5**, e44 (2017).
- Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning* (Springer, New York, 2001).
- Van Overschee, P. & De Moor, B. *Subspace Identification for Linear Systems: Theory, Implementation and Applications* (Kluwer Academic, Norwell, Massachusetts, USA, 1996).
- Fitzgerald, P.B., Laird, A.R., Maller, J. & Daskalakis, Z.J. A meta-analytic study of changes in brain activation in depression. *Hum. Brain Mapp.* **29**, 683–695 (2008).
- Schieber, M.H. Constraints on somatotopic organization in the primary motor cortex. *J. Neurophysiol.* **86**, 2125–2143 (2001).
- Andersen, R.A., Kellis, S., Klaes, C. & Aflalo, T. Toward more versatile and intuitive cortical brain-machine interfaces. *Curr. Biol.* **24**, R885–R897 (2014).
- Brandman, D.M., Cash, S.S. & Hochberg, L.R. Review: human intracortical recording and neural decoding for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1687–1696 (2017).
- Hatsopoulos, N.G. & Suminski, A.J. Sensing with the motor cortex. *Neuron* **72**, 477–487 (2011).
- Hotson, G. *et al.* Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. *J. Neural Eng.* **13**, 026017–26017 (2016).
- Leuthardt, E.C., Schalk, G., Wolpaw, J.R., Ojemann, J.G. & Moran, D.W. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.* **1**, 63–71 (2004).
- Nicolelis, M.A. & Lebedev, M.A. Principles of neural ensemble physiology underlying the operation of brain-machine interfaces. *Nat. Rev. Neurosci.* **10**, 530–540 (2009).
- Sadtler, P.T. *et al.* Neural constraints on learning. *Nature* **512**, 423–426 (2014).
- Sajda, P., Muller, K.-R. & Shenoy, K. Brain-computer interfaces. *IEEE Signal Process. Mag.* **25**, 16–17 (2008).
- Schwartz, A.B., Cui, X.T., Weber, D.J. & Moran, D.W. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron* **52**, 205–220 (2006).
- Shanechi, M.M. Brain-machine interface control algorithms. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1725–1734 (2017).
- Shanechi, M.M. *et al.* Rapid control and feedback rates enhance neuroprosthetic control. *Nat. Commun.* **8**, 13825 (2017).
- Shenoy, K.V. & Carmena, J.M. Combining decoder design and neural adaptation in brain-machine interfaces. *Neuron* **84**, 665–680 (2014).
- Thakor, N.V. Translating the brain-machine interface. *Sci. Transl. Med.* **5**, 210ps17 (2013).
- Wang, W. *et al.* An electrocorticographic brain interface in an individual with tetraplegia. *PLoS One* **8**, e55344 (2013).
- Yanagisawa, T. *et al.* Electrocorticographic control of a prosthetic arm in paralyzed patients. *Ann. Neurol.* **71**, 353–361 (2012).
- Habel, U., Klein, M., Kellermann, T., Shah, N.J. & Schneider, F. Same or different? Neural correlates of happy and sad mood in healthy males. *Neuroimage* **26**, 206–214 (2005).
- Mak, A.K., Hu, Z.G., Zhang, J.X., Xiao, Z.W. & Lee, T.M. Neural correlates of regulation of positive and negative emotions: an fMRI study. *Neurosci. Lett.* **457**, 101–106 (2009).
- Sartorius, N., Ustün, T.B., Lecrubier, Y. & Wittchen, H.U. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. *Br. J. Psychiatry Suppl.* **1996**, 38–43 (1996).
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (American Psychiatric Publishing, Arlington, Virginia, 2013).

50. Russell, J.A. Core affect and the psychological construction of emotion. *Psychol. Rev.* **110**, 145–172 (2003).
51. Kroenke, K., Spitzer, R.L. & Williams, J.B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
52. Spitzer, R.L., Kroenke, K., Williams, J.B. & Löwe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* **166**, 1092–1097 (2006).
53. Hermann, B.P., Seidenberg, M. & Bell, B. Psychiatric comorbidity in chronic epilepsy: identification, consequences, and treatment of major depression. *Epilepsia* **41**(Suppl. 2), S31–S41 (2000).
54. Harden, C.L. *et al.* A pilot study of mood in epilepsy patients treated with vagus nerve stimulation. *Epilepsy Behav.* **1**, 93–99 (2000).
55. George, M.S. *et al.* Vagus nerve stimulation for the treatment of depression and other neuropsychiatric disorders. *Expert Rev. Neurother.* **7**, 63–74 (2007).
56. Etkin, A. & Wager, T.D. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* **164**, 1476–1488 (2007).
57. Goldstein, R.Z. & Volkow, N.D. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat. Rev. Neurosci.* **12**, 652–669 (2011).
58. Tracey, I. & Bushnell, M.C. How neuroimaging studies have challenged us to rethink: is chronic pain a disease? *J. Pain* **10**, 1113–1120 (2009).
59. Yang, Y. & Shanechi, M.M. Generalized binary noise stimulation enables time-efficient identification of input-output brain network dynamics. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2016**, 1766–1769 (2016).
60. Yang, Y., Chang, E.F. & Shanechi, M.M. Dynamic tracking of non-stationarity in human ECoG activity. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2017**, 1660–1663 (2017).

ONLINE METHODS

Subjects. Semi-chronic intracranial ECoG electrodes were surgically implanted in seven subjects with treatment-resistant epilepsy for localization of seizure foci (Supplementary Tables 1 and 2). The surgery was approved by the University of California, San Francisco Institutional Review Board. All subjects provided written consent to participate in this study.

Neural recordings. Raw ECoG signals were continuously recorded during the subject's hospitalization with a sampling rate of either 0.5 kHz or 1 kHz using Nicolet/XLtek electroencephalogram clinical recording systems (Natus Medical, Inc.). ECoG electrodes included four-contact and six-contact strip electrodes with 10 mm center-to-center spacing and 2.3 mm exposed diameter, and four-contact and ten-contact depth electrodes with 6, 5 or 3 mm center-to-center spacing (Ad-Tech Corp.). Overall, electrode coverage was heterogeneous, but some regions (e.g., OFC, ACC, hippocampus; see Supplementary Table 8 for abbreviations) were covered in most subjects. Anatomical locations of contacts on each ECoG electrode—i.e., recording channels—were determined using the FreeSurfer neuroimaging analysis software⁶¹ and for electrodes in key limbic regions were validated by expert examination. The standard Montreal Neurological Institute template brain⁶² was used for visualization in Figure 3.

Mood state measurements. Subjects' mood states were measured with a tablet-based self-report mood assessment questionnaire designed by Posit Science Corp., called the immediate mood scalar (IMS)²⁶. IMS provides a momentary assessment of a set of mood states related to depression and anxiety symptoms and has been validated against standard self-report measures of these disorders²⁶ (Supplementary Note 1). In each of 24 questions, the subject is asked to rate their current mood state ("rate how you feel now"²⁶) by tapping one of seven buttons on a continuum between a pair of negative and positive mood state descriptors (e.g., "depressed" and "happy"; Supplementary Table 4). The buttons on the continuum have scores from -3 to +3. The sum of all 24 scores gives the total IMS. A higher IMS corresponds to a more positive mood state. Only IMS points (i.e., reports) that had concurrent usable ECoG recordings were used for the analyses. Any IMS point with change of more than 50 units from a previous IMS point taken within 2 h was removed due to instability (only 1 point). Each subject had 12 ± 2.4 usable IMS points, and answering the questions took 2.9 ± 1.4 min.

Medication use. Only three subjects (EC79, EC87, and EC150) were using depression or anxiety medication. The same modeling methodology was used in all subjects regardless of the presence or absence of medication and could predict mood in each subject (Fig. 2 and Table 1). Thus, decoding ability did not depend on the presence of medication. In terms of timing, there was no systematic synchronization between the timing of IMS points and medication use. IMS points were measured at irregular times but medications were administered at regular 12- or 24-h intervals. Other medications were for seizure, pain and nausea and were administered on an as-needed basis; they had no systematic relation to IMS times.

Subject selection criteria. We selected subjects with large-scale recordings who had (i) at least ten IMS points so that there were enough IMS data for model fitting, and (ii) a measured IMS range of at least 10% of the total possible IMS range across multiple days so that there was meaningful mood state variations. Each of our subjects had an IMS range of at least 25% of the total possible IMS range (average of 33%; Supplementary Table 3).

ECoG preprocessing. Raw ECoG signals were first preprocessed offline to remove non-neural activity. For each subject, we visually inspected the ECoG signals and marked the ECoG channels and time epochs that contained clearly non-neural motion artifacts. Channels that were found to be noisy for more than 10% of the time during the recording or were noisy during the mood state measurements were excluded from analysis. In other channels, neural features were later linearly interpolated during the noisy epochs.

We downsampled the recordings to a 256-Hz sampling rate (after applying a zero-phase antialiasing Chebyshev type I IIR filter of order 8 with a 100-Hz cut-off), high-pass filtered above 1 Hz (order 2 zero-phase IIR Butterworth

filter), and then applied an order 4 zero-phase notch filter at 60 Hz and harmonics to remove the line noise. We performed common average referencing (CAR) across all channels sharing the same lead. In calculating the reference, we excluded channels and time epochs that had been marked as noisy to prevent noise from spreading to other channels through the CAR.

Neural spectral feature extraction. We extracted log spectral power features from non-overlapping 10-s windows in five frequency bands: 1–8 Hz (delta plus theta), 8–12 Hz (alpha), 12–30 Hz (beta), 30–55 Hz (gamma-1) and 65–100 Hz (gamma-2) by filtering the ECoG signal within these bands (order 8 zero-phase IIR Butterworth filter) and then taking the log of the root mean square of the ECoG signal in each window. We also decoded mood state from coherence features (Supplementary Note 3 and Supplementary Code).

Performance measure. Our decoding measure is the error between the cross-validated prediction of a test IMS point and its true value. We quantify this prediction error using the normalized root mean square error (NRMSE), defined as

$$\text{NRMSE} = \frac{\sqrt{\sum_{j=1}^N (\hat{s}^{(j)} - s^{(j)})^2}}{\sqrt{\sum_{j=1}^N (\bar{s}^{(j)} - s^{(j)})^2}} \quad (1)$$

where N is the total number of IMS points, $s^{(j)}$ is the j th IMS point, $\hat{s}^{(j)}$ is the cross-validated prediction of the j th IMS point (i.e., prediction of the j th IMS point with a decoder that is trained without seeing this IMS point) and $\bar{s}^{(j)}$ is the mean of other IMS points except for the j th IMS point. Thus, NRMSE is the cross-validated root mean square error (RMSE) normalized by the RMSE of predicting the test IMS point as the mean of other IMS points (Supplementary Code).

Decoding evaluation using leave-one-out cross-validation. We evaluate decoding within rigorous leave-one-out cross-validation. We leave out an IMS point to be predicted (test IMS). We then use the rest of the IMS points as training data (training IMS) to select a network and fit its neural encoding model, and thus to build the decoder (Fig. 1a). This decoder is then used to predict the test IMS (Fig. 1c). Critically, the model fitting and selection have no knowledge of the test IMS. We repeat this leave-one-out procedure for all IMS points and compute the cross-validated IMS prediction NRMSE in equation (1). We perform statistical tests on this cross-validated NRMSE (using random-test and permuted-test). We only evaluate our decoders for prediction of mood state at the discrete points in time when an IMS point (i.e., report) is available.

Statistical tests. We conduct two statistical tests to assess decoding. (i) In the random-test, for each subject, we generate 1,000 sets of random integer numbers drawn from the same range as the true IMS points using a uniform distribution and place these at the same times as the true IMS points. Each random set has as many points as there are true IMS points. We then repeat the same cross-validated modeling and decoding using the same neural data for each set of random IMS points. This results in a distribution of the 1,000 random-IMS cross-validated prediction errors for each subject. We define the random-test P value as the probability that random IMS points will have lower cross-validated prediction error than the true IMS points (Supplementary Note 6). (ii) In the permuted-test, we randomly permute the time indices of the IMS points 1,000 times for each subject and repeat the procedure in the random-test to get the permuted-test P values. We take the random-test as the main criteria for significance and provide the permuted-test to show robustness to statistical tests.

Overview of model construction, fitting and evaluation. Within cross-validation, our modeling framework first reduces the dimensionality of the neural features to obtain a small number of hidden variables to serve as neural predictors, and then only relates these neural predictors to the training IMS via a linear regression. It then uses the resulting trained decoder to predict the test IMS, which is not seen by any step of model training. We provide an overview (Supplementary Fig. 2) first before giving details.

We reduce dimensionality by considering candidate models that (i) preferably consist of a small number of regions (see “Progressive region selection method”) and (ii) describe the neural activity in these regions in terms of a low-dimensional hidden variable (see “Neural encoding model and the corresponding mood state decoder”). We fit and compare a limited number of candidate models with the above qualities and then select one of them (see “Model fitting” and “Model selection”), all using only the training IMS within cross-validation.

The candidate models are generated through progressive region selection and dynamic modeling, which constitute our dimensionality reduction steps. In progressive region selection (**Supplementary Fig. 2a**), we start from considering the smallest possible networks—namely, all possible networks with a single region. To model activity in each region (**Supplementary Fig. 2b**), we use dynamic modeling consisting of principal component analysis (PCA)²⁷ and a linear state-space model (LSSM)²⁸ that—completely unsupervised with respect to IMS points and purely based on neural features—extract a low-dimensional hidden neural state. This neural state constitutes the neural predictors. Thus, in all candidate models, we restrict this neural state dimension to be similar to or smaller than the number of available IMS points. We then fit a regularized linear regression from this low-dimensional hidden neural state to training IMS. Given that the only parameters in a candidate model (**Supplementary Fig. 2b**) that need to be fitted using the training IMS are the regression parameters—which are equal in number to the small number of neural predictors—the dimensionality reduction process avoids overfitting to training IMS.

Once we have all the candidate models (e.g., corresponding to different single regions), we need to select one of them purely using the training IMS (**Supplementary Fig. 2c**). To promote generalizability, we select a model with not only small prediction error for training IMS but also low sensitivity to training IMS (see “Model selection”). We then construct the decoder for this model and use it to predict the test IMS, which has not been seen by any step of modeling (**Supplementary Fig. 2d**). If the prediction of test IMS is significant (random-test), we stop (**Supplementary Fig. 2a**). In five of seven subjects, we indeed stopped at one region. If not, we declare that a one-region network is not sufficient for decoding and start the cross-validated modeling process from scratch, but this time with candidate models consisting of two regions (see “Progressive region selection method”). This progressive process continues until either prediction is possible or we have considered all regions and prediction is not possible. Below we describe the details.

Progressive region selection method. The method consists of progressive stages to find the smallest network size sufficient for decoding (**Supplementary Fig. 2a**). We define each ECoG strip or depth electrode (each of which has several channels) as a region. In the first stage, we fit neural encoding models for every single region on its own—i.e., networks of size 1 consisting of a single electrode’s recording channels (**Supplementary Fig. 2b**; see “Model fitting”). Among the candidate single-region neural encoding models, we select a generalizable model by considering both the prediction error and the sensitivity measures of candidate models, which are computed purely from the training data (**Supplementary Fig. 2c**; see “Model selection”). Finally, we build a mood state decoder based on the selected neural encoding model. We then use this decoder to predict the test IMS—which has not been seen by any step of model fitting or selection—and thus compute the cross-validated prediction error (**Supplementary Fig. 2d**). We repeat this process for all IMS points (i.e., conduct leave-one-out cross-validation) to compute the average cross-validated prediction error (i.e., NRMSE in equation (1)). If the cross-validated prediction of test IMS is significant ($P \leq 0.05$, random-test), we declare that mood state can be decoded from a network of size 1 and terminate the region selection process. This was the case for five of seven subjects.

Only if decoding is not possible with a one-region network do we proceed to the second stage of the method to construct two-region networks. We repeat the cross-validated model fitting, selection and evaluation exactly as in the first stage, this time for candidate two-region networks. To make the search space practical, the first region of all candidate two-region networks is fixed as the region that was selected purely on the basis of training IMS in the first stage (see “Model selection”). Also, in the second stage, we make a correction for making two comparisons (one for the one-region network and one for the

two-region network) using false discovery rate (FDR)⁶³ control. If the FDR-corrected $P \leq 0.05$, we declare that mood state can be decoded from a network of size 2 and terminate the region selection. If not, we repeat the process by progressively increasing the network size and performing FDR correction, until we have a stage in which the FDR corrected $P \leq 0.05$; at such a stage, we declare that decoding is possible with a network size corresponding to that stage. If the procedure does not lead to significant decoding (i.e., $P > 0.05$) even when including all regions, we declare that decoding is not possible. We also devised an alternative statistical test that empirically accounts for multiple comparisons and therefore does not require FDR correction. Results were consistent using both tests across our subjects (**Supplementary Note 7**).

In the following sections, we provide details of each element in progressive region selection (**Supplementary Fig. 2**). We emphasize that all model fitting and selection described below is done purely on the basis of the training data within cross-validation.

Neural encoding model and the corresponding mood state decoder. For a given candidate network in progressive region selection, we build a neural encoding model that describes the temporal dynamics of neural features in terms of a low-dimensional hidden neural state and then relates this neural state to training IMS points.

Assuming n_z neural features in the network, we denote the features at time t by $z_t \in \mathbb{R}^{n_z}$. We identify the low-dimensional hidden neural state completely unsupervised with respect to (i.e., without knowing) IMS points and merely to describe the temporal dynamics of neural features. We first perform PCA and extract the top n_y principal components (PCs), $y_t \in \mathbb{R}^{n_y}$, via a linear transform

$$y_t = Pz_t \quad (2)$$

where rows of P consist of eigenvectors corresponding to the n_y largest eigenvalues of the neural feature sample covariance matrix. We then fit an LSSM to these PCs y_t to define the low-dimensional hidden neural state $x_t \in \mathbb{R}^{n_x}$, written as

$$\begin{cases} x_{t+1} = Ax_t + Ke_t \\ y_t = Cx_t + e_t \end{cases} \quad (3)$$

Here we model the neural noise and model mismatch as a white Gaussian noise $e_t \in \mathbb{R}^{n_y}$ with covariance matrix $E\{e_t e_t^T\} = R\delta_{ij}$ ($\delta_{ij} = 1$ if $i = j$ and 0 otherwise). To enable generalizability, we restrict the upper bounds on the number of retained PCs n_y and the neural state dimension n_x to be similar to or smaller than the typical number of available IMS points. We describe how we select n_y and n_x under “Model selection.” For a given network, n_y and n_x , the model parameters are the PCA matrix $p \in \mathbb{R}^{n_y \times n_z}$ and the LSSM matrices $A \in \mathbb{R}^{n_x \times n_x}$, $C \in \mathbb{R}^{n_y \times n_x}$, $K \in \mathbb{R}^{n_x \times n_y}$ and $R \in \mathbb{R}^{n_y \times n_y}$. All model parameters are fitted unsupervised with respect to IMS points and only on the basis of neural features using standard methods (see “Model fitting”). Note that LSSMs have also been used in BMI decoders^{32,33,36,38–41,43,64–71}, where the state is largely taken as the behavioral (e.g., kinematic) state to be decoded or sometimes as the dynamical state of a population of spiking neurons⁶⁶ (**Supplementary Note 8**). We used LSSM as one component of dimensionality reduction to help obtain a low-dimensional hidden neural state.

Once the PCA and LSSM are fitted and the neural state is identified, we build a linear regression model that relates the neural state x_t to training IMS points, denoted by s_t , as

$$s_t = Tx_t + s_0 + \varepsilon_t \quad (4)$$

Here $s_0 \in \mathbb{R}$ is the mean of the training IMS and $\varepsilon_t \in \mathbb{R}$ is zero-mean white Gaussian noise.

Together, equations (2)–(4) provide a parametric candidate neural encoding model (**Supplementary Fig. 2b**). The only parameters of a candidate model that are fitted using the training IMS are the regression parameters $T \in \mathbb{R}^{1 \times n_x}$ and s_0 (see “Model fitting”). The neural encoding model is dynamic, which introduces several benefits—for example, denoising neural features and accumulating information from them over time, as well as timescale computations (**Supplementary Note 8**).

The optimal decoder for the neural encoding model in equations (2)–(4) consists of three main steps (**Supplementary Fig. 2d** and **Supplementary Code**). First, it applies the linear transformation in equation (2) to the neural features z_t to obtain the PCs, y_t . Second, it uses the optimal recursive Kalman filter⁷² for the LSSM in equation (3) to estimate the neural state \hat{x}_t from the PCs y_t as:

$$\hat{x}_t = A\hat{x}_{t-1} + K(y_t - C\hat{x}_{t-1}) \quad (5)$$

Here, as in equation (3), K is the gain matrix^{28,72}. The decoder then applies a moving average to estimate the neural state, \bar{x}_t , over the past 4 min, which is roughly the time it takes to fill out one full IMS questionnaire²⁶. Third, using equation (4), the decoder estimates IMS, \hat{s}_t as:

$$\hat{s}_t = T\bar{x}_t + s_0 \quad (6)$$

Model fitting. For a given network and a given number of retained PCs n_y and neural state dimension n_x , the fitting of model parameters consists of an unsupervised learning step and a supervised learning step (**Supplementary Fig. 2b**).

In the unsupervised learning step, we fit the PCA and LSSM parameters P , A , C , K and R purely from neural data, without any information from the training IMS. We concatenate the neural spectral features within 10 h of IMS points (union of 20-h windows centered at each IMS point) to form z_t . We then conduct PCA on z_t to fit P and retain the first n_y PCs as y_t . The LSSM parameters A , C , K and R are then fitted using subspace identification²⁸, which is a stable numerical algorithm to fit LSSMs only from observed data y_t when the neural state x_t is hidden. Since the fitting of PCA and LSSM is based on days of continuous neural recordings, there is no shortage of data in this step.

In the supervised learning step, we use the fitted LSSM to estimate the averaged neural state \bar{x}_t from y_t using equation (5) and the 4-min averaging. We then use ridge regularization to fit T and s_0 as the optimal solution²⁷ for

$$\min_{T, s_0} \sum_{k=1}^M (s^{(k)} - T\bar{x}^{(k)} - s_0)^2 + \lambda \|T\|_2^2 \quad (7)$$

where $s^{(k)}$ is the k th training IMS point, $\bar{x}^{(k)}$ is the average state around the k th training IMS, $M = N - 1$ is the number of training IMS points, and $\lambda > 0$ is the regularization parameter selected using standard inner-level cross-validation²⁷ (**Supplementary Note 9**). The total number of regression parameters is $n_x + 1$ (n_x parameters for T and 1 for s_0). Ridge regularization further ensures that the effective number of regression parameters²⁷ is strictly less than $n_x + 1$ and M (**Supplementary Note 9**). Together the progressive region selection, PCA, LSSM and regularization could make the effective number of regression parameters to be about 30–50% of the number of available IMS points in every subject (**Supplementary Table 6**).

Model selection. In progressive region selection, at each stage, we can fit a pool of candidate models (for different choices of regions, number of retained PCs n_y and neural state dimension n_x), and we need to select one of them (**Supplementary Fig. 2c**). We thus need a model selection technique that uses only the training IMS to pick a model that is likely to generalize to the test IMS. Thus, we introduce a measure that evaluates the sensitivity of a given model to the training IMS. We then pick among models that have low sensitivity to the training IMS even if they do not have the lowest prediction error for the training IMS. For simplicity, we first present the model selection procedure for a network size of 1; i.e., with a single region. We then extend to larger network sizes.

For G total regions, a single-region network can be built using any one of these G regions. For each region, we consider four choices of PC dimension n_y from {1,5,10,15} and four choices of neural state dimension n_x from {3,6,9,DR}. Here DR represents a special case of LSSM corresponding to a direct regression model from y_t to s_t (equivalent to a degenerate LSSM with $A = C = 0_{n_y \times n_y}$ and $K = I_{n_y \times n_y}$, where $0_{n_y \times n_y}$ and $I_{n_y \times n_y}$ are $n_y \times n_y$ zero and identity matrices, respectively). We choose the upper limits of n_y and n_x to be comparable to the typical number of available IMS points. This ensures that the number of neural predictors are at most comparable to the number of available IMS points, thus helping prevent overfitting to the training IMS. For each combination of region, n_y and n_x , we fit a neural encoding model consisting of equations (2)–(4) using the training data (**Supplementary Fig. 2b**

and “Model fitting”). This forms a pool of 16G candidate models from which we need to select one generalizable model.

To select a model (**Supplementary Fig. 2c**), we compute two performance measures for each candidate model. First, we calculate the prediction error (i.e., NRMSE) of that model for the training IMS using an internal leave-one-out cross-validation. Second, on the basis of Cook’s distance⁷³, we quantify the model sensitivity to training IMS points, denoted by D . For the j th training IMS point, Cook’s distance D_j is defined as

$$D_j = \frac{\sum_{k=1}^M (\hat{s}^{(k)} - \hat{s}^{(k,-j)})^2}{f\hat{\sigma}_e^2} \quad (8)$$

where $\hat{s}^{(k)}$ is the prediction of the k th training IMS by a model fitted to all training IMS, $\hat{s}^{(k,-j)}$ is the prediction of the k th training IMS by a model fitted to all training IMS except the j th training IMS, f is the effective number of parameters in the regression model (**Supplementary Note 9**) and

$$\hat{\sigma}_e^2 = \frac{1}{(M-f)} \sum_{k=1}^M (s^{(k)} - \hat{s}^{(k)})^2$$

D_j shows how much influence the j th training IMS point has on the fitting of the neural encoding model. For a given fitted model, we compute the maximum sensitivity to training IMS points, $D = \max_{j \in \{1, \dots, M\}} D_j$, as our sensitivity

measure. Since $D > 1$ implies strong model sensitivity to data⁷³, we choose the model with the smallest prediction error for the training IMS among those that have $D \leq 0.5$. If no model satisfies this limit, we gradually relax the maximum allowed D to 1 in 0.1 steps until we find a model that satisfies the limit. If no model has $D \leq 1$, we discard D and select the model with the smallest prediction error for the training IMS. Note that the entire model selection described here is performed completely unaware of the test IMS to be decoded later using the selected model (**Supplementary Fig. 2d**).

The model selection procedure is the same for larger networks, except in how we grow the network size. At stage h , to build candidate networks of size h , we fix the first $h - 1$ regions as those selected in stage $h - 1$ on the basis of the training data, and thus have $G - h + 1$ possible h -region networks to choose from. This greedy algorithm limits the search scope to be much smaller than all possible combinations of regions. Finally, note that all candidate models include all frequency bands (**Supplementary Note 10**).

Brain region search space to evaluate decoding. In our main decoding analysis, we first focus on limbic regions—including OFC, ACC, insular cortex, amygdala and hippocampus—in progressive region selection, given their central role^{1–11,74–80}. From each ECoG strip or depth electrode, we only include channels that are confirmed to be within these limbic regions by FreeSurfer⁶¹ and further expert verification. Only if the limbic regions alone are not sufficient for significant prediction (which happened only for EC137) do we extend our search to all channels from all ECoG electrodes (excluding channels in white matter). We then repeat the progressive region selection from scratch. In this case, in addition to the region selection stages, we also correct for the initial search within the limbic regions with FDR control.

We also perform two control analyses. First, we search within all electrodes for all subjects. Second, we search only within electrodes outside limbic regions—i.e., we exclude all limbic electrodes from the search.

Population-level prediction evaluation. To evaluate decoding at the population level, we first standardize the IMS points in each subject by z -scoring the true and cross-validated IMS predictions on the basis of the true IMS points. We then pool these standardized values from all subjects and compute the cross-validated NRMSE. To assess statistical significance, we randomly select one set of the random IMS points from each subject and pool them with a similar procedure. We repeat this process 10⁹ times and calculate the P value on the basis of the resulting NRMSE distribution (**Supplementary Note 6**). We repeat this procedure to get the permuted-test P value.

Life Sciences Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The custom computer code in this study is provided as **Supplementary Code** and is also available from the corresponding authors (M.M.S. and E.F.C.).

Data availability. The data that support the findings of this study are available upon reasonable request from the corresponding authors (M.M.S. and E.F.C.).

61. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
62. Mazziotta, J. *et al.* A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Phil. Trans. R. Soc. Lond. B* **356**, 1293–1322 (2001).
63. Yekutieli, D. & Benjamini, Y. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plan. Inference* **82**, 171–196 (1999).
64. Donoghue, J.P. Bridging the brain to the world: a perspective on neural interface systems. *Neuron* **60**, 511–521 (2008).
65. Hsieh, H.-L. & Shanechi, M.M. Optimizing the learning rate for adaptive estimation of neural encoding models. *PLoS Comput. Biol.* **14**, e1006168 (2018).
66. Kao, J.C. *et al.* Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat. Commun.* **6**, 7759 (2015).
67. Pistohl, T., Ball, T., Schulze-Bonhage, A., Aertsen, A. & Mehring, C. Prediction of arm movement trajectories from ECoG-recordings in humans. *J. Neurosci. Methods* **167**, 105–114 (2008).
68. Shanechi, M.M., Orsborn, A.L. & Carmena, J.M. Robust brain-machine interface design using optimal feedback control modeling and adaptive point process filtering. *PLoS Comput. Biol.* **12**, e1004730 (2016).
69. So, K., Dangi, S., Orsborn, A.L., Gastpar, M.C. & Carmena, J.M. Subject-specific modulation of local field potential spectral power during brain-machine interface control in primates. *J. Neural Eng.* **11**, 026002 (2014).
70. Stavisky, S.D., Kao, J.C., Nuyujukian, P., Ryu, S.I. & Shenoy, K.V. A high performing brain-machine interface driven by low-frequency local field potentials alone and together with spikes. *J. Neural Eng.* **12**, 036009 (2015).
71. Yang, Y. & Shanechi, M.M. An adaptive and generalizable closed-loop system for control of medically induced coma and other states of anesthesia. *J. Neural Eng.* **13**, 066019 (2016).
72. Anderson, B.D. & Moore, J.B. *Optimal Filtering* (Dover, New York, 1979).
73. Cook, R.D. & Weisberg, S. *Residuals and Influence in Regression* (Chapman and Hall, New York, 1982).
74. Burgos-Robles, A. *et al.* Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment. *Nat. Neurosci.* **20**, 824–835 (2017).
75. Ebmeier, K.P., Donaghey, C. & Steele, J.D. Recent developments and current controversies in depression. *Lancet* **367**, 153–167 (2006).
76. Kim, S.-Y. *et al.* Diverging neural pathways assemble a behavioural state from separable features in anxiety. *Nature* **496**, 219–223 (2013).
77. Michopoulos, V. *et al.* Association of CRP genetic variation and CRP level with elevated PTSD symptoms and physiological responses in a civilian population with high levels of trauma. *Am. J. Psychiatry* **172**, 353–362 (2015).
78. Mulders, P.C., van Eijndhoven, P.F., Schene, A.H., Beckmann, C.F. & Tendolkar, I. Resting-state functional connectivity in major depressive disorder: A review. *Neurosci. Biobehav. Rev.* **56**, 330–344 (2015).
79. Namburi, P. *et al.* A circuit mechanism for differentiating positive and negative associations. *Nature* **520**, 675–678 (2015).
80. Tye, K.M. *et al.* Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature* **493**, 537–541 (2013).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

The immediate mood scaler (IMS) questionnaire was provided to subjects via an iPad application developed by Posit Science Corp. ECoG data was collected using Natus Nicolet (EC79) or XLTek (other subjects) EEG clinical recording systems with a Natus Quantum Amplifier, and Natus NeuroWorks software version 8 (Natus Medical, Inc.).

Data analysis

All algorithms were implemented as custom MATLAB code and ran on MATLAB version R2017a. Anatomical location of contacts on each ECoG electrode, i.e., recording channels, were determined using the FreeSurfer neuroimaging analysis software, version 5.3.0 (EC79 to EC137) or 6.0.0-2beb96c (EC150 and EC166).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All possible subjects were used (N=7 subjects, see Methods). The results are based on personalized per-subject decoding. So conclusions about decoding can be made for each subject even individually (see Fig. 2).
Data exclusions	We visually inspected the ECoG signals and marked the ECoG channels and time epochs that contained clearly non-neural motion artifacts. Channels that were found to be noisy for more than 10% of the time during the recording or were noisy during the mood state measurements were excluded from analysis. Only IMS points (i.e., reports) that had concurrent usable ECoG recordings were used for the analyses. Any IMS point with change of more than 50 units from a previous IMS point taken within 2 hours prior was removed due to instability (only 1 point). Each subject had 12 +/- 2.4 usable IMS points. All exclusion criteria were pre-established and were done to remove non-neural/electronic noise and artifacts in data.
Replication	Decoding was successfully achieved in each of the 7 subjects. Each subject provides an independent reproduction attempt, and all attempts led to successful personalized decoding of mood.
Randomization	Not relevant to the study. Each subject was used for decoding, and conclusions were made for each subject. There was no grouping of subjects.
Blinding	Not relevant to study. There was no group allocation. Decoding was attempted for each subject individually.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Seven human subjects (2 male, 5 female, aged 40 ± 17 years). See Methods and Supplementary Table 1 for more details.
Recruitment	All epilepsy subjects with large-scale neural recordings, and sufficient mood state reports and mood state range were recruited (see Methods for details).