Published in final edited form as:

Neuroimage. 2018 October 15; 180(Pt A): 211–222. doi:10.1016/j.neuroimage.2017.06.059.

A multimodal encoding model applied to imaging decision-related neural cascades in the human brain

Jordan Muraskin^a, Truman R. Brown^b, Jennifer M. Walz^c, Tao Tu^a, Bryan Conroy^d, Robin I. Goldman^e, and Paul Sajda^a

^aDepartment of Biomedical Engineering, Columbia University, New York, NY, USA 10027

^bCenter for Biomedical Imaging, Medical University of South Carolina, Charleston, SC, USA, 29425

^cFlorey Institute of Neuroscience and Mental Health, Melbourne, Australia

^dPhilips Research, Cambridge, MA USA, 02141

^eCenter for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA, 53705

Abstract

Perception and cognition in the brain are naturally characterized as spatiotemporal processes. Decision-making, for example depends on coordinated patterns of neural activity cascading across the brain, running in time from stimulus to response and in space from primary sensory regions to the frontal lobe. Measuring this cascade is key to developing an understanding of brain function. Here we report on a novel methodology that employs multi-modal imaging for inferring this cascade in humans at unprecedented spatiotemporal resolution. Specifically we develop an encoding model to link simultaneously measured electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) signals to infer high-resolution spatiotemporal brain dynamics during a perceptual decision. After demonstrating replication of results from the literature, we report previously unobserved sequential reactivation of a substantial fraction of the pre-response network whose magnitude correlates with a proxy for decision confidence. Our encoding model, which temporally tags BOLD activations using time localized EEG variability, identifies a coordinated and spatially distributed neural cascade that is associated with a perceptual decision. In general the methodology illuminates complex brain dynamics that would otherwise be unobservable using fMRI or EEG acquired separately.

Keywords

EEG; fMRI; simultaneous; decision-making; confidence

Correspondence to: Paul Sajda.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

The detailed spatiotemporal brain dynamics that underlie human perception and cognition are difficult to measure. Invasive techniques with sufficient temporal or spatial resolution, such as depth electrodes or cortical arrays used with epilepsy patients, are only feasible in rare cases and, in addition, do not capture activity from the entire brain. In comparison, non-invasive measures such as electroencephalography (EEG) and magnetoencephalography (MEG) suffer from poor spatial resolution, and blood oxygen level dependent functional MRI (BOLD fMRI) from poor temporal resolution and indirect coupling to neural activity [1]. In spite of this, EEG, MEG, and fMRI have been used individually to study functional activity in the human brain, although, by themselves these modalities provide a limited view of the underlying brain dynamics [2].

Recently, methods enabling simultaneous acquisition of EEG and fMRI (EEG/fMRI) have led to varied analytic approaches aimed at integrating the electrophysiological and hemodynamic information acquired through the simultaneous measurements. Such approaches offer the potential to provide a more comprehensive picture of global brain dynamics, and will likely offer new insights into how the brain makes rapid decisions [3, 4]. Some of the techniques that have been proposed for combining multi-modal brain signals have separately analyzed the EEG and fMRI data and subsequently juxtaposed the results [5, 6], while others attempt for a truly integrated approach in order to fully exploit the joint information contained in the data sets [7]. In general, simultaneous EEG/fMRI and the associated analysis techniques have been used to identify neuronal sources of EEG trial-to-trial variability, linking them to cognitive processes such as attention [8] and inhibition [9].

Many previous studies have used known EEG markers (P1, N2, N170, P300, α -rhythm) or data driven approaches such as Independent Component Analysis (ICA) to combine EEG with fMRI data [3, 9, 16–23]. One promising approach has been to use supervised machine-learning techniques (e.g. classifiers) to find relevant projections of the EEG data, where single-trial variability of the electrophysiological response along these projections can be correlated in the fMRI space. Goldman et al.[24], Walz et al.[25], Muraskin et al [26], and Fouragan et al.[27] have demonstrated this technique on visual and auditory paradigms. This methodology has been shown to localize cortical regions that modulate with the task while preserving the temporal progression of task-relevant neural activity. However, these techniques treat neighboring temporal regions independently when fusing the EEG with fMRI. The approach we present in this paper considers the entire temporal progression of task-specific discriminant activity in the EEG and models how this temporal progression is encoded in the fMRI BOLD data. Using the full temporal progression allows us to improve temporal resolution and increase the quality of the fit in the encoding model and the accuracy in the subsequent decoding.

Encoding models have become an important machine learning tool for analysis of fMRI [13], in that they provide a generative model of the BOLD data (Fig. 1A). In most cases encoding models have been used to learn brain activity that encodes or represents features of a stimulus, such as visual orientation energy in an image/video [11, 12, 14], acoustic spectral power in sound/speech [28], visual imagery during sleep [29] and even high level concepts

and semantics [30]. In the method presented here, however, we employ an encoding model methodology to directly relate the simultaneously collected data from the two neuroimaging modalities. Instead of learning features derived from characteristics of the stimulus, they are derived from the trial-to-trial variability of the EEG components learned via a classifier (Fig. 1B). Specifically, we learn an encoding for the spatially precise fMRI data from the temporally precise trial-to-trial variability of the EEG – i.e. we learn a mapping that explains the BOLD data given the EEG variability across the time course of each trial. In contrast to previous simultaneous EEG/fMRI techniques that treat the EEG variability at different time points in the trial as independent, we are able to exploit the temporal structure embedded in the EEG by modeling the EEG variability as a time series to accurately predict time course activations in the spatial domain.

As an example of this method, we apply it to simultaneously acquired EEG/fMRI data collected while subjects performed a perceptual decision-making task. The EEG variability we exploit is derived from a classifier trained to predict the level of stimulus evidence on a trial-to-trial basis from the multichannel EEG data. This approach leverages the fact that in perceptual decision making tasks, the level of stimulus evidence, as measured via EEG, persists across the trial [31, 32]. We can discriminate this information in a time-localized way in the EEG and then temporally “tag” specific cortical areas by their trial-to-trial variability as they become involved and uninvolved in the decision process. We then proceed to discuss the significance of the results of the new method relative to what can be inferred from non-simultaneously acquired EEG and fMRI.

Methods

Subjects

21 subjects (12 male, 9 female; age range 20–35 years) participated in the study. The Columbia University Institutional Review Board (IRB) approved all experiments and informed consent was obtained before the start of each experiment. All subjects had normal or corrected-to-normal vision.

Stimuli

We used a set of 30 face (from the Max Planck Institute face database), 30 car, and 30 house (obtained from the web) gray scale images (image size 512x512 pixels, 8 bits/pixel). They all had identical magnitude spectra (average magnitude spectrum of all images in the database) and their corresponding phase spectra were manipulated using the weighted mean phase (WMP) [33] technique to generate a set of images characterized by their % phase coherence. The stimulus evidence (high or low) for each trial was systematically varied by modifying the salience of the image via randomization of image phase at either 35% (low) or 50% (high) coherence.

Experimental task

The stimuli were used in an event-related three choice reaction time task. On each trial, an image either a face, car, or house was presented for 50 ms and subjects were instructed to respond with the category of the image by pressing one of three buttons on an MR

compatible button controller. Stimuli were presented to subjects using E-Prime software (Psychology Software Tools) and a VisuaStim Digital System (Resonance Technology) with 600x800 goggle display. Images subtended 11° x 8° of visual angle. Over four runs, a total of 720 trials were acquired (240 of each category with 120 high coherence trials) with a random inter-trial interval (ITI) sampled uniformly between 2–4s. Each run lasted for 560 seconds.

fMRI acquisition

Blood-oxygenation-level-dependent (BOLD) T2*-weighted functional images were acquired on a 3T Philips Achieva scanner using a gradient-echo echo-planar imaging (EPI) pulse sequence with the following parameters: Repetition time (TR) 2000ms, echo time (TE) 25ms, flip angle 90, slice thickness 3mm, interslice gap 1mm, in-plane resolution 3x3mm, 27 slices per volume, 280 volumes. For all of the participants, we also acquired a standard T1-weighted structural MRI scan (SPGR, resolution 1x1x1mm).

EEG acquisition

We simultaneously and continuously recorded EEG using a custom-built MR-compatible EEG system [34, 35], with differential amplifiers and bipolar EEG montage. The caps were configured with 36 Ag/AgCl electrodes including left and right mastoids, arranged as 43 bipolar pairs. Further details are given in Supplementary Material.

Functional image pre-processing

Image preprocessing was performed with FSL (www.fmrib.ox.ac.uk/fsl/). Functional images were spatially realigned to the middle image in the times series (motion-correction), corrected for slice time acquisition, spatially smoothed with a 6mm FWHM Gaussian kernel, and high pass filtered (100s). The structural images were segmented (into grey matter, white matter and cerebrospinal fluid), bias corrected and spatially normalized to the MNI template using FAST [36]. Functional images were registered into MNI space using boundary based registration (BBR) [37].

EEG data preprocessing

We performed standard EEG preprocessing offline using MATLAB (MathWorks) with the following digital Butterworth filters: 0.5 Hz high pass to remove direct current drift, 60 and 120 Hz notches to remove electrical line noise and its first harmonic, and 100 Hz low pass to remove high-frequency artifacts not associated with neurophysiological processes. These filters were applied together in the form of a zero-phase finite impulse response filter to avoid distortions caused by phase delays. We extracted stimulus-locked 1500 ms epochs (–500:1000) and subtracted the mean baseline (–200:0) from the rest of the epoch. Through visual inspection, we discarded trials containing motion and/or blink artifacts, evidenced by sudden high-amplitude deflections.

Estimating EEG components and trial-to-trial variability

We used logistic regression to associate each trial with the level of stimulus evidence represented in the EEG. We considered high stimulus and low stimulus evidence trials

irrespective of behavioral accuracy. Regularized logistic regression (r-LR) was used as a classifier to find an optimal projection for discriminating between high and low stimulus evidence trials over a specific temporal window. A sweep of the regularization parameters was implemented using FaSTGLZ [38]. This approach has been previously applied to identify neural components underlying rapid perceptual decision-making [24, 25, 39–44].

Specifically, we defined 50ms duration training windows centered at time, τ , ranging from stimulus onset to 800ms following the stimulus in 25ms steps. We used r-LR to estimate a spatial weighting, on all the N EEG channels, that discriminated high stimulus evidence from low stimulus evidence trials. This identifies an $N \times 1$ vector, w_τ , which defines a hyperplane that maximally discriminates between each class (e.g., high vs. low stimulus evidence trials) in the N dimensional space of electrodes. Using this w_τ we calculate the distance each trial is from the decision boundary

$$y_\tau = w_\tau^T E_\tau \quad (1)$$

In eqn. 1, E_τ is an $N \times p$ vector (N sensors per time window, τ , by p trials). We varied the center of the window (τ) across the trial in 25ms time-steps. For each subject, this produced a matrix Y where the rows corresponded to trials and the columns to training windows, i.e. Y incorporates the time-window localized trial-to-trial variability in the EEG that we subsequently use to tag the BOLD data. We quantified the performance of the linear discriminator by the area under the receiver operator characteristic (ROC) curve, referred to here as AUC, using a leave-one-out procedure. We used the ROC AUC metric to characterize the discrimination performance at each τ .

Traditional fMRI analysis

We first ran a traditional general linear model (GLM) fMRI analysis in FSL, using event-related (high and low stimulus evidence) and response time (RT) variability regressors. The two event related regressors were unit amplitude boxcar functions matching onset and offset of high and low stimulus evidence trials. For each stimulus evidence level, we modeled out the response time variability by constructing regressors with each trials boxcar amplitude matching the z-scored RT within each stimulus evidence level. The onset and offset of the boxcar function of the RT variability regressor matches that of the corresponding stimulus regressor. Activated regions that passed a family-wise error (FWE) [45] corrected cluster threshold of $p < 0.01$ at a z-score threshold of 2.57 were considered significant. Further details can be found in the Supplemental Material.

fMRI deconvolution

Associating fMRI data to each trial is challenging for two main reasons: (a) the temporal dynamics of the hemodynamic response function (HRF) evolve over a longer time-scale than the mean ITI of the event-related design, resulting in overlapping responses between adjacent trials; and (b) the ITI was random for each trial so that the fMRI data was not acquired at a common lag relative to stimulus onset. To overcome these issues, we employed

the ‘least squares-separate’ (LS-S) deconvolution [46] method to estimate the voxel activations for each trial. We extracted 58697 voxels from a common gray matter group mask at 3×3×3mm spatial resolution that excluded white matter and CSF and assembled the resulting voxel activations into rows of the data matrix F .

Single subject encoding model

All encoding model analyses were performed in MATLAB. To relate the EEG data with the fMRI, we devised a subject-wise spatiotemporal decomposition using singular value decomposition (SVD). Let F be an $m \times p$ matrix denoting m -voxels and p -trials that is the deconvolved high and low stimulus evidence fMRI data for each trial. Let Y be the $r \times p$ matrix denoting r -windows (33 EEG_r windows and response time (RT)) and p -trials. For each trial, the first row of Y is the response times while subsequent rows are the y values at each window time. Let W be an $m \times r$ matrix that is the multivariate weights on Y that solve for F with some error ε .

$$F = WY + \varepsilon \quad (2)$$

Normally, if we solve for W using the least squares approach, we obtain:

$$W = (FY^T)(YY^T)^{-1} \quad (3)$$

However, each time point might be highly correlated with its neighbors, which reduces the stability of the least-squares regression. We use SVD to reduce the feature space and improve our estimation of W (the weights on each window).

We include a regressor for RT since it is a strong predictor for evidence type. Higher stimulus evidence leads to faster decisions than low evidence trials (see supplementary material Fig. S1). The faster RTs will manifest itself in EEG variability. We wanted to find neural correlates of the decision process independent of the response time/motor execution. For that reason, we needed to include the RT variability in the SVD to remove any variability in the EEG that correlates with the motor execution. If RT was not added to the SVD, the encoding model weights may be contaminated with differences in motor execution instead of the decision process. It is important to include RT especially because areas that correlate with RT variability overlap with many executive decision areas (see Walz et al., 2013 Fig. 2).

For a leave-one-out cross validation, we hold out a single trial from the EEG Y matrix and the corresponding volume from the fMRI data F and train on the remaining trials. We repeated this for all trials. For each leave one out fold we calculated,

$$Y^{Train} = U\Sigma V^T \quad (4)$$

where U is an $r \times r$ orthonormal matrix, Σ is a $r \times p$ diagonal matrix and V is a $p \times p$ orthonormal matrix. After SVD on Y^{Train} , we reduced the feature dimensions on Y^{Train} to retain 75% of the variance by only keeping v components. To do this, we selected the first v rows of Σ and zeroed the other rows. This results in $\tilde{\Sigma}$ which is our reduced subspace matrix. We recalculate our least squares solution where we have replaced Y by its reduced form $U\tilde{\Sigma}V^T$ in equation 3:

$$\hat{W} = (F^{Train}V\tilde{\Sigma}^T)(\Sigma\Sigma^T)^{-1}U^T \quad (5)$$

For each leave one out fold, we first calculated the SVD of the training set. We then calculated the number of components to keep and then solve for $\hat{W}$, the weight estimate per fold. To test, we applied the weights to the left-out test data Y^{Test} to estimate the encoded fMRI data $\hat{F}$ for the encoding part:

$$\hat{F} = \hat{W}Y^{Test} \quad (6)$$

We constructed the $m \times r$ weight matrix $\bar{W}$ by taking the average of all the trained $\hat{W}$ matrices. For the decoding model we use the left out test data F^{Test} and estimate the EEG variability $\hat{Y}$:

$$\hat{Y} = \hat{W}^T F^{Test} (\hat{W}^T \hat{W})^+ \quad (7)$$

Here, $\hat{W}^T \hat{W}$ is not invertible, and instead use the pseudo-inverse (denoted by $+$).

Given $\hat{F}$, an $m \times p$ matrix with m voxels by p trials, for each voxel j , we calculated the correlation over trials of $\hat{F}_j$ with F_j^{Test} , resulting in the matrices R^{fMRI} (Pearson Correlation Map) and P^{fMRI} (p-value map of the Pearson Correlation) that are $m \times 1$. The P^{fMRI} was then converted to a z-score map. Thus, we have a map for each subject indicating that subjects encoding model significance at each voxel. To test which time windows were significant in the decoding model, we also calculated, R_{τ}^{EEG} , the correlation between $\hat{Y}_{\tau}$ and Y_{τ} . For each subject, R_{τ}^{EEG} indicates how well the decoding model predicts EEG variability at each time window from a single fMRI volume.

Group level spatiotemporal analysis

We evaluated both the encoding model and the resulting decoding model at the group level. First, to identify group level voxels where our encoding model predictions were consistent across subjects, each subject's p-value maps for the leave-one-out correlation (R^{fMRI}) were converted into their respective z-values, then a one-sample t-test testing for the mean difference from 0 was calculated for each voxel and voxel-wise significance was calculated using threshold-free cluster enhancement (TFCE) using a non-parametric randomization procedure implemented in FSL[47]. TFCE is a multiple comparison correction technique in

which raw voxel-wise t-statistics are adjusted based on their t-scores and their neighboring spatial clusters, producing corrected whole brain volumes. It generally allows for better sensitivity than other cluster thresholding methods. In our analysis we used the default TFCE parameters provided by FSL (E=2 and H=0.5). Voxels were considered significant if they passed a false discovery rate (FDR) threshold of $p < 0.01$.

These significant voxels were then used as a mask to temporally localize weight values in $\hat{W}$ by computing the voxels that were consistent in their direction (positive (high stimulus evidence) or negative (low stimulus evidence)) and timing (τ window) across subjects. Here, the goal was to develop a statistical parametric map of $\hat{W}$ that identifies significant results in both space and time, thereby creating a spatiotemporal map of a perceptual decision.

First, a one-sample t-test across the subjects $\hat{W}$ was computed for all voxels. This generates a statistical parametric map of T-scores in space (voxels) and time (windows). To correct for multiple comparisons, we implemented a spatiotemporal TFCE (stTFCE) in both space (neighboring voxels) and time (neighboring time windows - response time window not included) and computed significance through a randomization procedure. For each of the 33 temporal windows, we randomly altered the sign of the voxel weights for each subject. We then computed the T-scores (across subjects) at each voxel for all time windows. The spatiotemporal T-score map was then transformed into the stTFCE map. This procedure was repeated 5000 times to generate the empirical null distribution of the stTFCE map. The final statistical parametric map was calculated by comparing the true stTFCE value with the distribution of permuted values for each voxel. Again, voxels were considered significant if they passed FDR correction at $p < 0.05$ (high stimulus evidence: FDR-Corrected $p < 0.0004$, low stimulus evidence: FDR-Corrected $p < 0.0002$). Note that now our number of multiple comparisons was the number of voxels in the FDR-mask (20256) times the number of time windows (33). We analyzed the response time separately with a standard TFCE randomization procedure implemented in FSL.

We also tested the quality of the resulting decoding model at the group level. To construct group level statistics for the decoding model, we first analyzed the R_{τ}^{EEG} vectors across all subjects to identify time windows when the model was consistent across subjects. The correlation vectors were converted into their z-values using the Fisher-Z transform, and for each time window (τ), were used to compute a Z-test across all subjects. These group level results were then corrected for multiple comparisons using $p < 0.05$ FDR.

Drift diffusion model (DDM) and confidence proxy

To analyze and better interpret our novel finding of a reactivation in the neural cascade (see Results), we used a DDM to estimate a proxy of confidence in the decision for each trial. The DDM models decision-making in two-choice tasks. Here, we treated the decision (correct vs. incorrect) as our two choices. A drift-process accumulates evidence over time until it crosses one of two boundaries (upper or lower) and initiates the corresponding response[48]. The speed with which the accumulation process approaches one of the two boundaries (a) is called drift-rate (v) and represents the relative evidence for or against a particular response.

We used Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python (HDDM) to calculate the drift rate (v), decision boundary (a) and non-decision time T_{non} for each subject [49]. Specifically, we modeled high and low stimulus evidence response time data separately. This was to ensure our confidence proxies were consistent within trial types. We included the response time and whether the subject got the trial correct. HDDM obtains a sequence of samples (i.e., a Markov chain Monte Carlo; MCMC) from the posterior of each parameter in the DDM. In our model, we generated 5000 samples from the posteriors, the first 1000 (burn-in) samples were discarded, and the remaining samples were thinned by 5%.

Recently, Philiastides et al. [41] showed that for conditions in which the boundary does not change, a proxy for decision confidence for each trial (i) can be computed by calculating $1/\sqrt{RT_i - T_{non}}$, where T_{non} is a subject dependent non-decision time fit by the model. Thus, after modeling the DDM process, each trial's (i) confidence proxy (CP) for each subject (j) was computed by $CP_{i,j} = 1/\sqrt{RT_i - T_{non,j}}$ and then z-scored across trials where $T_{non,j}$ was varied for high or low stimulus evidence trials, separately, so that CP was a measure of relative trial confidence within difficulty levels.

Using the encoding model weights ($\bar{W}$), we are able to estimate the trial-to-trial activation amplitude at different time windows. To estimate the reactivation amplitudes, we took the mean across time of $Y_{j,i}^R = W_{j,PostACC}^T F_{j,i}$ for each subject (j) and trial (i). Here, $W_{PostACC}$ is the weight matrix from the encoding model restricted to voxels that were significant in the spatiotemporal group results from the 600–800ms windows. By applying the weights restricted to these spatial areas and temporal windows to each trial, we are able to get a trial-to-trial value for reactivation. Note that because the reactivation is stronger in the more difficult trials the significant voxels in W are negative and so more negative y 's indicate stronger reactivation while more positive y 's indicate weaker reactivation. $Y_{j,i}^R$ was quintiled for high and low stimulus evidence and the average confidence proxy was calculated within each quintile. A linear mixed effects model [50] was used to test if the slope of confidences across quintile grouping, $Y_{j,i}^R$, was significantly different from 0 while including stimulus evidence as a condition. Separate similar analyses with non-replay windows (175–250ms) and testing for behavioral accuracy were also performed. To test the contribution of each cluster to the correlation with confidence, we implemented feature elimination procedure [51, 52], where our features were clusters of significant voxels (>48 voxels) during the 600–800ms time window. This procedure removed clusters from the reactivation network before calculating trial-to-trial reactivation similar to the procedure in Sherwin et al. (2015). We then calculated the percent change in slope (reactivation $\times$ confidence proxy) when the cluster was removed compared to the total network ((new-original)/original). This procedure ranks cluster importance by sorting which clusters, when removed, had the strongest negative effect on slope height. For example, a cluster that does not contribute to the reactivation $\times$ confidence proxy slope when removed would not reduce the slope height, however, a cluster that has a large contribution to the reactivation $\times$ confidence proxy slope would reduce the slope height when removed. While such an analysis did not provide insight into the sensitivity of the interactions between clusters, it does provide insight into each cluster on its own.

Results

As a demonstration of how our method for encoding time-localized EEG variability in simultaneously acquired BOLD data can be used to infer neural cascades underlying perception and cognition, we collected EEG/fMRI data from subjects as they performed a three choice reaction time task discriminating between faces, cars, and houses (Fig. 2A). Subjects were instructed to discriminate the object class after briefly viewing an image corrupted by varying levels of noise (Fig. 2B) and respond by pressing one of three buttons. Overall, subjects responded with accuracies of $94 \pm 5\%$ and $58 \pm 12\%$ and with response times of $634 \pm 82\text{ms}$ and $770 \pm 99\text{ms}$ for high and low stimulus evidence trials, respectively (Fig. 2C, D). Subject accuracies and response times across stimulus types (faces, cars, houses) for low stimulus evidence trials were similar; however, for high stimulus-evidence trials subject accuracies were higher and response times were shorter for faces than for cars or houses (see supplementary material Fig. S1).

GLM based analysis of BOLD fMRI shows superposition of cortical areas correlated with stimulus evidence

We first used a traditional general linear model (GLM) analysis of the fMRI (see Methods) to demonstrate how a difference in stimulus evidence is reflected in the BOLD data (Fig 2E, supplementary material Table S1). These results become a baseline case we will compare to relative to the additional information that we can obtain using our encoding model for fusing the BOLD and EEG data. Brain regions showing greater BOLD activation to high vs. low stimulus evidence trials included areas associated with early visual perception and the default mode network, such as fusiform gyrus, parahippocampal gyrus, lateral occipital cortex, superior frontal gyrus, and posterior cingulate cortex. Regions with greater BOLD activation to low vs. high stimulus evidence trials included areas in the executive control and difficulty networks, such as dorsal lateral prefrontal cortex, anterior cingulate cortex, intraparietal sulcus, and insula. Important to note is that these GLM results, from simultaneous EEG and fMRI acquisitions, reproduced previous results in the literature for fMRI collected alone [53](see supplementary material Fig. S2A). The encoding model results we present below are compared to these results in terms of how the temporal superposition of activations can be separated in time.

Extracting temporally localized EEG signatures of stimulus evidence variability

The traditional fMRI results showed multiple brain regions correlated with the stimulus evidence of the trial; however, this traditional approach does not enable one to infer the relative timing of these fMRI activations. However in the trial averaged event related potentials, we see that the difference in stimulus evidence (high vs. low stimulus evidence) spans the trial (see Fig 2F). To infer timing at a scale of tens of milliseconds, we exploit this difference and use linear classification [40, 54] of the multichannel EEG to estimate EEG components selective to the trial-to-trial variability in the stimulus evidence. Importantly these components were estimated at 25ms steps, locked to either the onset of the stimulus or response, so as to encode the BOLD activity at specified time points in the trial. The basic idea is illustrated in Figure 3, where hypothetical neural activity is shown for two different regions that are constituents of the perceptual decision-making network. Averaging over

trials would clearly reveal a difference in the mean neural activity between high and low stimulus evidence (as we see in Fig 2F). However, the two regions contribute differentially to the network, with one region encoding the stimulus evidence (Region 1) and the other integrating it over time (Region 2); both are sensitive to the level of stimulus evidence, though varying so at different times in the trial. By taking advantage of this sensitivity to the stimulus evidence, we can learn EEG discriminant components, i.e. spatial filters, that best classify trials at different time windows given the neural data. We used the trial-to-trial variability along these component directions as features to uniquely tag fMRI voxels with the specific time window of the component. This tagging is done by building an encoding model of the features, given the BOLD signal, details of which are described in the following section.

We constructed EEG components by learning linear classifiers at 25ms steps, starting from stimulus onset to 50ms past the average low stimulus evidence response time. We chose a time step of 25ms due to an empirical analysis showing a half width of 50ms in the temporal autocorrelation of the EEG data, though in principle this methodology allows for temporal resolution up to the EEG sampling rate. Each classifier was associated with a set of discriminant values, which can be represented as a vector y_{τ} ; each element of the vector is the distance of a given trial to the discrimination boundary for the classifier at time step τ (Fig. 3). This distance can be interpreted as a measure of the EEG classifier's estimate of the level of stimulus evidence for that trial [24, 25, 39, 40, 42, 54].

Results of the EEG analysis show discriminating information for stimulus evidence spanning the trial (see Fig. 4A), beginning roughly 175ms post-stimulus to past the average response times. A dip occurs around 300ms, indicating stimulus evidence is less discriminative at this time and serves to demarcate early and late cognitive processes. The early process corresponded to the time of the D220 ERP component, which has been shown to modulate with the degree of task difficulty, whether via stimulus noise or task demands [32]. The later and more prolonged component is likely related to more complex cognitive and motor preparatory processes that differ between high and low stimulus evidence trials. Importantly, although the early and late EEG components were both discriminative, we found their trial-to-trial variability to be uncorrelated (Fig. 4B and see supplementary material Fig. S3E), indicating that while the discriminating information (level of stimulus evidence) persists across the trial, it couples differently to processes across time.

Using an encoding model to link fMRI activations with temporally distinct EEG trial-to-trial variability

After extracting the trial-to-trial variability from the EEG discriminant components, feature vectors y_{τ} are collected across time steps, τ , along with a response time vector to construct a matrix Y . This matrix is the temporally precise representation of the trial-to-trial EEG variability that reflects high vs. low stimulus evidence. An encoding model (see Methods) is then fit to predict the trial-to-trial variability of the BOLD response for each fMRI voxel. The model can be seen as “tagging” each voxel with a “time”, or set of times, when it encodes the variability in the given EEG discriminant component(s).

The group level result of the encoding model is shown in Fig. 4C (see also supplementary material Fig. S2B) where significant voxels from the encoding model are shown in yellow. Fig. 4D shows the trial-to-trial variability of BOLD signal at a specific voxel, comparing it to the variability predicted by the encoding model. Additional validity of the encoding model and single subject results are presented in the supplementary material Fig. S4A/B. The encoding model was also evaluated as a decoding model (see Methods) with the BOLD activity used to predict the trial-to-trial variability in the EEG for unseen data—data on which the encoding model was not trained. Fig. 4E shows these results, expressed as the correlation between the measured and predicted EEG trial-to-trial variability across the 800ms epoch. The shape of the curve is highly consistent with that observed for the EEG data itself (comparing Fig. 4A and Fig. 4E) (additional analysis of the fidelity of the model is provided in the supplementary material Fig. S3).

Given the encoding model, we construct the progression of the BOLD activity across time by identifying voxel weights that are consistent across subjects in space and time (see Methods). Fig. 5 shows these results for a group level analysis. The weights from the encoding model can be interpreted similarly to the β -weights in a traditional GLM contrast. Here, red voxels indicate significant regions where high stimulus evidence variability positively covary between EEG discrimination y-values and BOLD activation and where blue voxels indicate a negative covariation between more negative EEG discrimination y-values (stronger low evidence discrimination) and BOLD activation. We observe a progression of activity (see also supplementary material Movie S1), at 25ms resolution, which proceeds simultaneously down the dorsal and ventral streams of visual processing for the first 250ms. After that the cascade becomes more complex with activation in the IPS at 425ms and 750ms (see Fig. 6A), reactivation of the SPL at 675ms and activation of ACC at 600ms. The reactivation pattern is particularly significant since it would not be observable via a traditional fMRI general linear model (GLM) analysis (results of which were presented in Fig. 2E), which integrates over time and thus superimposes these activities. For example, the changing sign of the middle temporal gyrus (MT) encoding weights in Fig. 6A manifested as no activity in the MT for the traditional fMRI GLM analysis of this same BOLD data—the change in sign canceled the effective correlation in the GLM. The areas of activation we find are internally consistent with the traditional GLM results (see Fig. S2E/F) and with previous reports in the literature for human subjects [53, 55]; however, here we are able to link activations across time in a way that was previously only possible with invasive techniques.

Cortical reactivation correlates with decision confidence

We identified a reactivation pattern in the spatiotemporal dynamics which was not observable using non-simultaneously acquired EEG or fMRI. Further analysis of the spatiotemporal dynamics (see Fig. 6B) shows that the reactivation pattern in the network occurs after decision-monitoring areas become engaged (i.e. after activation of ACC).

Spontaneous reactivation, or “replay,” of neural activity in the human brain has been observed and is believed to be important for memory consolidation [56] and more recently has been hypothesized to play a role in perceptual decision-making by enabling the

formation of decision confidence [57]. To test the hypothesis that the reactivation activity we see may be related to decision confidence, we used a hierarchical drift diffusion model (DDM) [48, 49] to fit the behavioral data for high and low stimulus evidence conditions (see Methods). Specifically, our model enables us to define a proxy for decision confidence based on the DDM fits to the behavior [41, 58].

Correlating the reactivation level to this confidence proxy shows a strong and significant monotonic relationship between confidence proxy and the level of reactivation (high stimulus evidence-slope= 0.037 ± 0.008 , $t=4.657$, $\text{dof}=206$, $p = 3.2 \times 10^{-6}$; low stimulus evidence-slope= 0.062 ± 0.008 , $t=7.754$, $\text{dof}=206$, $p = 8.88 \times 10^{-15}$), with low stimulus evidence trials reactivated more strongly than high stimulus evidence trials (difference in slopes= -0.025 ± 0.011 , $t=2.189$, $\text{dof}=206$, $p=0.029$) (see Fig. 7 and supplementary material Fig. S5a). Additionally, reactivation amplitude correlates with behavioral accuracy (supplementary material Fig. S5b) (high stimulus evidence, slope= 0.0115 ± 0.0047 , $t=2.41$, $\text{dof}=206$, $p=0.016$; low stimulus evidence, slope= 0.0104 ± 0.0047 , $t=2.19$, $\text{dof}=206$, $p=0.028$). Additional control analyses from early time windows (175–250ms) shows no significant correlations between activation and confidence proxy (supplementary material Fig. S5).

We next analyzed which cortical regions most contributed to this reactivation. Specifically, we used a recursive feature elimination to identify those regions contributing the most to the correlation between the level of reactivation and the confidence proxy. We found that the IPS/SPL and dorsal lateral prefrontal cortex (DLPFC) contributed the most (Fig. 7C). Interestingly, these two areas have been previously implicated in metacognitive judgments of confidence [57, 59, 60].

Discussion

In this paper we have shown that linking simultaneously acquired EEG and fMRI using a novel encoding model enables us to probe the high-resolution spatiotemporal dynamics in the human brain. We demonstrate the methodology using neural data acquired during a rapid perceptual decision-making task. Our method, which resolves whole-brain activity with EEG-like temporal resolution, uncovers what appears to be a reactivation of neural activity in perceptual decision making, dynamics that would otherwise be masked by the temporal averaging and slow dynamics of traditional fMRI. More broadly, our results demonstrate a general non-invasive data-driven methodology for measuring, at high spatiotemporal resolution, latent neural processes underlying human behavior. Below we discuss the methodology and our specific findings for perceptual decision-making with respect to the current literature.

The trial-to-trial variability we leverage in our methodology can be viewed as a combination of exogenous and endogenous neural variability expressed in the EEG and fMRI BOLD. The exogenous variability is related to the stimulus presentation while the endogenous variability is subject and trial specific. Endogenous trial-to-trial variability during perceptual decisions has been shown to be associated with attention [25], reward [27], confidence [61] and other internal mental states that vary at different times during the perceptual decision.

The classifier applied to the EEG data is thus sensitive to both the exogenous variability (stimulus differences) and the endogenous variability at the multiple time steps throughout the decision process.

The approach we present requires that EEG and BOLD data be collected simultaneously and not in separate sessions in order to exploit the correlations in exogenous and endogenous trial-to-trial variability between EEG and BOLD to temporally “tag” voxels in the space of the fMRI. To show the importance of collecting the data simultaneously, we ran a control analysis that randomly permuted the trials within their stimulus evidence class, thus effectively simulating an EEG and BOLD dataset collected separately. By destroying the link between the EEG and BOLD trials, the encoding model failed to find any consistent activation (supplementary material Fig. S8), indicating the necessity of simultaneous acquisition.

Alternative techniques for fusing simultaneous EEG-fMRI typically do not exploit EEG across the trial and instead only analyze specific ERP components or time windows of interest [3, 8, 16–19, 21–25, 27, 44]. Results from these techniques identify regions that modulate with the specific components, but yield limited information about the timing of other task-relevant regions seen in traditional fMRI contrasts. The methodology developed here extends the previously reported work of Goldman et al. [24] and Walz et al. [25] by combining their EEG data reduction techniques with techniques developed for encoding stimulus features onto BOLD data [10–15], ultimately providing a framework for labeling voxels in task-relevant fMRI contrasts with their timing information (supplementary material Fig. S2C/E/F).

Clearly, other EEG components that are task-related can be isolated and could potentially be used to “tag” BOLD data. The sliding window linear classification used here acts to reduce the EEG data along a dimension that categorizes stimulus evidence; however, this could be replaced by any other data reduction technique, such as temporally windowed ICA or PCA. Variability along these component directions could then be used in the encoding model to link with the simultaneously collected BOLD data. The choice of data reduction technique (i.e. feature space) would be highly dependent on the nature of the inferences.

In terms what our methodology is able to potentially reveal about the spatiotemporal brain dynamics of perceptual decision making, we observe what appears to be reactivation of the pre-response network, spatiotemporal dynamics that would be masked using traditional fMRI analysis. Interestingly, the reactivation terminated in a network that included the MFG, SPL, and IPS, similar areas previously reported to be reactivated in metacognitive judgments of confidence in perceptual decisions [57, 59, 60]. In addition, these areas contributed the most to the correlation to our confidence proxy. Gherman and Philiastides [61] observed this network using a multivariate single-trial EEG approach, coupled with a distributed source reconstruction technique. Fleming et al. [57] and Heereman et al. [62] used BOLD fMRI to show that areas in this network negatively correlate with subjective certainty ratings. Unique to our findings, we saw this reactivation on a single-trial basis after engagement of the ACC, which has been shown to be involved in decision monitoring [61, 63], and also observed the dynamic sequence leading up to this network reactivation. Our results showed that

reactivation/replay occurred on a trial-to-trial basis after a decision, was stronger for difficult decisions, and correlated with a proxy for decision confidence (Fig. 7).

To ensure that our findings were not simply an artifact of a stimulus-locked analysis, we performed the same analysis response-locked. We found that the reactivation, both its correlation with the confidence proxy and the cortical areas implicated, to be consistent with the stimulus locked analysis (see supplementary material Fig. S6). In addition we found that the reactivation clearly begins pre-response and extends through and past the timing of the behavioral response (supplementary material Fig. S7). Thus our complete analysis suggests that while in the process of making and executing a perceptual decision, the brain may replay activity to develop a metacognitive representation of its decision.

We note several caveats in terms of the interpretation of our novel results as they relate to perceptual decision-making. One potential criticism is that the variability we discriminate and encode in the BOLD is related to the stimulus-evidence and not a decision variable such as the choice (e.g. trial-to-trial variability of a subject's decision of a face, car or house). Though our previous work has shown that there are EEG components that are sensitive to stimulus category or choice, these components tend to be isolated at specific time windows [32, 64]. The level of stimulus evidence, on the other hand, spans most of the trial and thus can be used to provide a more complete picture of brain dynamics during the task. Ultimately the reactivation we observe in these dynamics is intriguing and thus we conducted additional analyses that points to a novel “replay” hypothesis, namely that the reactivation is a replay of activity associated with the human brain generating, on a trial-to-trial basis, a meta-cognitive judgment of confidence immediately after the decision. We note that this hypothesis is possible only by linking the EEG and BOLD data.

Another potential source of bias in our results comes from the autocorrelation of the beta estimates in our deconvolution procedure. Since there is temporal autocorrelation in the beta's our leave-one-out cross validation is not the optimal analysis design [65]. However we decided to run the leave-one-out cross validation to remain consistent with previous analyses.

Of course additional experiments must be done to more thoroughly test our “replay” hypothesis. For example future work will investigate other measures of confidence that are more direct than the proxy we estimated from the DDM. We will also consider different types of rapid perceptual decision-making tasks to investigate whether the results we see generalize across stimulus types and/or the modality of stimulus presentation. In general, we have shown that simultaneously acquired EEG/fMRI data enables a novel non-invasive approach to visualize high resolution spatial and temporal processing in the human brain with the potential for providing a more comprehensive understanding of the neural basis of complex behaviors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Jianing Shi for assistance in collecting the EEG/fMRI data. This work was funded by National Institutes of Health Grant R01-MH085092, DARPA under Contract NBCHC090029 and the Army Research Laboratory and under Cooperative Agreement Number W911NF-10-2-0022.

References

1. Logothetis NK. What we can do and what we cannot do with fMRI. *Nature*. 2008; 453(7197):869–78. DOI: 10.1038/nature06976 [PubMed: 18548064]
2. Alexander DM, Trengove C, van Leeuwen C. Donders is dead: cortical traveling waves and the limits of mental chronometry in cognitive neuroscience. *Cognitive Processing*.
3. Huster RJ, Debener S, Eichele T, Herrmann CS. Methods for simultaneous EEG-fMRI: an introductory review. *Journal of Neuroscience*. 2012; 32(18):6053–60. DOI: 10.1523/JNEUROSCI.0447-12.2012 [PubMed: 22553012]
4. Jorge J, van der Zwaag W, Figueiredo P. EEG-fMRI integration for the study of human brain function. *NeuroImage*. 2014; 102:24–34. DOI: 10.1016/j.neuroimage.2013.05.114 [PubMed: 23732883]
5. Plichta MM, Wolf I, Hohmann S, Baumeister S, Boecker R, Schwarz aJ, Zangl M, Mier D, Diener C, Meyer P, Holz N, Ruf M, Gerchen MF, Bernal-Casas D, Kolev V, Yordanova J, Flor H, Laucht M, Banaschewski T, Kirsch P, Meyer-Lindenberg a, Brandeis D. Simultaneous EEG and fMRI Reveals a Causally Connected Subcortical-Cortical Network during Reward Anticipation. *Journal of Neuroscience*. 2013; 33(36):14526–14533. DOI: 10.1523/JNEUROSCI.0631-13.2013 [PubMed: 24005303]
6. Yuan H, Liu T, Szarkowski R, Rios C, Ashe J, He B. Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: an EEG and fMRI study of motor imagery and movements. *NeuroImage*. 2010; 49(3):2596–606. DOI: 10.1016/j.neuroimage.2009.10.028 [PubMed: 19850134]
7. Dahne S, Biessmann F, Samek W, Haufe S, Goltz D, Gundlach C, Villringer A, Fazli S, Muller K-R. Multivariate Machine Learning Methods for Fusing Multimodal Functional Neuroimaging Data. 2015; doi: 10.1109/JPROC.2015.2425807
8. Warbrick T, Arrubla J, Boers F, Neuner I, Shah NJ. Attention to Detail: Why Considering Task Demands Is Essential for Single-Trial Analysis of BOLD Correlates of the Visual P1 and N1. *Journal of cognitive neuroscience*. 2013; :1–14. DOI: 10.1162/jocn [PubMed: 24047384]
9. Baumeister S, Hohmann S, Wolf I, Plichta M, Rechtsteiner S, Zangl M, Ruf M, Holz N, Boecker R, Meyer-Lindenberg A, Holtmann M, Laucht M, Banaschewski T, Brandeis D. Sequential inhibitory control processes assessed through simultaneous EEG-fMRI. *NeuroImage*.
10. Kuker T, Nishimoto S, Huth AG, Gallant JL. Attention during natural vision warps semantic representation across the human brain. *Nature neuroscience*. 2013; 16(6):763–770. DOI: 10.1038/nn.3381 [PubMed: 23603707]
11. Hansen KA, Kay KN, Gallant JL. Topographic organization in and near human visual area V4. *Journal of Neuroscience*. 2007; 27:11896–11911. DOI: 10.1523/JNEUROSCI.2991-07.2007 [PubMed: 17978030]
12. Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. *Nature*. 2008; 452(7185):352–355. DOI: 10.1038/nature06713 [PubMed: 18322462]
13. Naselaris T, Kay KN, Nishimoto S, Gallant JL. Encoding and decoding in fMRI. 2011
14. Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*. 2011; 21:1641–1646. DOI: 10.1016/j.cub.2011.08.031 [PubMed: 21945275]
15. Stansbury D, Naselaris T, Gallant J. Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex. *Neuron*. 2013; 79:1025–1034. DOI: 10.1016/j.neuron.2013.06.034 [PubMed: 23932491]
16. De Martino F, Valente G, de Borst AW, Esposito F, Roebroeck A, Goebel R, Formisano E. Multimodal imaging: an evaluation of univariate and multivariate methods for simultaneous EEG/

- fMRI. Magnetic resonance imaging. 2010; 28(8):1104–1112. DOI: 10.1016/j.mri.2009.12.026 [PubMed: 20097029]
17. Jann K, Dierks T, Boesch C, Kottlow M, Strik W, Koenig T. BOLD correlates of EEG alpha phase-locking and the fMRI default mode network. *NeuroImage*. 2009; 45(3):903–916. DOI: 10.1016/j.neuroimage.2009.01.001 [PubMed: 19280706]
  18. Jaspers-Fayer F, Ertl M, Leicht G, Leupelt A, Mulert C. Single-trial EEG-fMRI coupling of the emotional auditory early posterior negativity. *NeuroImage*. 2012; 62(3):1807–1814. DOI: 10.1016/j.neuroimage.2012.05.018 [PubMed: 22584235]
  19. Mayhew SD, Ostwald D, Porcaro C, Bagshaw AP. Spontaneous EEG alpha oscillation interacts with positive and negative BOLD responses in the visual-auditory cortices and default-mode network. *NeuroImage*. 2013; 76:362–372. DOI: 10.1016/j.neuroimage.2013.02.070 [PubMed: 23507378]
  20. Nguyen VT, Cunnington R. The superior temporal sulcus and the N170 during face processing: single trial analysis of concurrent EEG-fMRI. *NeuroImage*. 2014; 86:492–502. DOI: 10.1016/j.neuroimage.2013.10.047 [PubMed: 24185024]
  21. Novitskiy N, Ramautar JR, Vanderperren K, De Vos M, Mennes M, Mijovic B, Vanrumste B, Stiers P, Van den Bergh B, Lagae L, Sunaert S, Van Huffel S, Wagemans J. The BOLD correlates of the visual P1 and N1 in single-trial analysis of simultaneous EEG-fMRI recordings during a spatial detection task. *NeuroImage*. 2011; 54(2):824–835. DOI: 10.1016/j.neuroimage.2010.09.041 [PubMed: 20869449]
  22. Omata K, Hanakawa T, Morimoto M, Honda M. Spontaneous Slow Fluctuation of EEG Alpha Rhythm Reflects Activity in Deep-Brain Structures: A Simultaneous EEG-fMRI Study. *PloS one*. 2013; 8(6):e66869.doi: 10.1371/journal.pone.0066869 [PubMed: 23824708]
  23. Warbrick T, Mobascher a, Brinkmeyer J, Musso F, Richter N, Stoecker T, Fink GR, Shah NJ, Winterer G. Single-trial P3 amplitude and latency informed event-related fMRI models yield different BOLD response patterns to a target detection task. *NeuroImage*. 2009; 47(4):1532–44. DOI: 10.1016/j.neuroimage.2009.05.082 [PubMed: 19505583]
  24. Goldman RI, Wei C-Y, Philiastides MG, Gerson AD, Friedman D, Brown TR, Sajda P. Single-trial discrimination for integrating simultaneous EEG and fMRI: identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task. *NeuroImage*. 2009; 47(1):136–47. DOI: 10.1016/j.neuroimage.2009.03.062 [PubMed: 19345734]
  25. Walz JM, Goldman RI, Carapezza M, Muraskin J, Brown TR, Sajda P. Simultaneous EEGfMRI reveals a temporal cascade of task-related and default-mode activations during a simple target detection task. *NeuroImage*. 2014; 102:229–239. DOI: 10.1016/j.neuroimage.2013.08.014 [PubMed: 23962956]
  26. Muraskin J, Sherwin J, Lieberman G, Garcia JO, Verstynen T, Vettel JM, Sajda P. Fusing Multiple Neuroimaging Modalities to Assess Group Differences in Perception Action Coupling. *Proceedings of the IEEE*. 2016; (99):1–18. DOI: 10.1109/jproc.2016.2574702
  27. Fouragnan E, Retzler C, Mullinger K, Philiastides MG. Two spatiotemporally distinct value systems shape reward-based learning in the human brain. *Nature Communications*. 2015; 6:8107.doi: 10.1038/ncomms9107
  28. Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*. 2014; 111:E4687–E4696. DOI: 10.1073/pnas.1323812111
  29. Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y. Neural decoding of visual imagery during sleep. *Science (New York, NY)*. 2013; 340(6132):639–42. DOI: 10.1126/science.1234330
  30. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*. 2016; 532(7600):453–8. DOI: 10.1038/nature17637 [PubMed: 27121839]
  31. Banko EM, Gal V, Kortvelyes J, Kovacs G, Vidnyanszky Z. Dissociating the effect of noise on sensory processing and overall decision difficulty. *J Neurosci*. 2011; 31(7):2663–2674. DOI: 10.1523/JNEUROSCI.2725-10.2011 [PubMed: 21325535]

32. Philiastides MG, Sajda P. Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb Cortex*. 2006; 16(4):509–518. DOI: 10.1093/cercor/bhi130 [PubMed: 16014865]
33. Dakin SC, Hess RF, Ledges T, Achtmann RL. What causes non-monotonic tuning of fMRI response to noisy images? *Current biology : CB*. 2002; 12(14):476–477.
34. Sajda P, Goldman R, Philiastides M, Gerson A, Brown T. A System for Single-trial Analysis of Simultaneously Acquired EEG and fMRI. 2007; 3rd International IEEE/EMBS Conference on Neural Engineering;
35. Sajda P, Goldman RI, Dyrholm M, Brown TR. Signal Processing and Machine Learning for Single-trial Analysis of Simultaneously Acquired EEG and fMRI. Elsevier Inc; 2010.
36. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*. 2001; 20:45–57. DOI: 10.1109/42.906424 [PubMed: 11293691]
37. Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*. 2009; 48:63–72. DOI: 10.1016/j.neuroimage.2009.06.060 [PubMed: 19573611]
38. Conroy BR, Walz JM, Sajda P. Fast bootstrapping and permutation testing for assessing reproducibility and interpretability of multivariate FMRI decoding models. *PloS one*. 2013; 8(11):e79271.doi: 10.1371/journal.pone.0079271 [PubMed: 24244465]
39. Muraskin J, Sherwin J, Sajda P. Knowing when not to swing: EEG evidence that enhanced perception-action coupling underlies baseball batter expertise. *NeuroImage*. 2015; 123:1–10. DOI: 10.1016/j.neuroimage.2015.08.028 [PubMed: 26299795]
40. Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. *NeuroImage*. 2005; 28:326–341. DOI: 10.1016/j.neuroimage.2005.05.032 [PubMed: 16084117]
41. Philiastides MG, Heekeren HR, Sajda P. Human Scalp Potentials Reflect a Mixture of Decision-Related Signals during Perceptual Choices. *Journal of Neuroscience*. 2014; 34(50):16877–16889. DOI: 10.1523/JNEUROSCI.3012-14.2014 [PubMed: 25505339]
42. Sherwin J, Muraskin J, Sajda P. You Can't Think and Hit at the Same Time: Neural Correlates of Baseball Pitch Classification. *Frontiers in neuroscience*. 2012; 6(December):177.doi: 10.3389/fnins.2012.00177 [PubMed: 23267311]
43. Sherwin J, Sajda P. Musical experts recruit action-related neural structures in harmonic anomaly detection: evidence for embodied cognition in expertise. *Brain and cognition*. 2013; 83(2):190–202. DOI: 10.1016/j.bandc.2013.07.002 [PubMed: 24056235]
44. Walz JM, Goldman RI, Carapezza M, Muraskin J, Brown TR, Sajda P. Simultaneous EEG-fMRI Reveals Temporal Evolution of Coupling between Supramodal Cortical Attention Networks and the Brainstem. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2013; 33(49):19212–22. DOI: 10.1523/JNEUROSCI.2649-13.2013 [PubMed: 24305817]
45. Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research*. 2003; 12:419–446. DOI: 10.1191/0962280203sm341ra [PubMed: 14599004]
46. Mumford JA, Turner BO, Ashby FG, Poldrack RA. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*. 2012; 59:2636–2643. DOI: 10.1016/j.neuroimage.2011.08.076 [PubMed: 21924359]
47. Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*. 2009; 44(1):83–98. DOI: 10.1016/j.neuroimage.2008.03.061 [PubMed: 18501637]
48. Ratcliff R, McKoon G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*. 2008; 20(4):873–922. DOI: 10.1162/neco.2008.12-06-420 [PubMed: 18085991]
49. Wiecki TV, Sofer I, Frank MJ. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics*. 2013; 7:14.doi: 10.3389/fninf.2013.00014 [PubMed: 23935581]
50. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015; 67(1):48.doi: 10.18637/jss.v067.i01

51. Martino FD, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. Combining multivariate voxel selection and support vector machines for mapping and classification of fmri spatial patterns. *NeuroImage*. 2008; 43(1):44–58. DOI: 10.1016/j.neuroimage.2008.06.037 [PubMed: 18672070]
52. Sherwin JS, Muraskin J, Sajda P. Pre-stimulus functional networks modulate task performance in time-pressured evidence gathering and decision-making. *NeuroImage*. 2015; 111:513–525. DOI: 10.1016/j.neuroimage.2015.01.023 [PubMed: 25614974]
53. Heekeren HR, Marrett S, Bandettini Pa, Ungerleider LG. A general mechanism for perceptual decision-making in the human brain. *Nature*. 2004; 431(7010):859–862. DOI: 10.1038/nature02966 [PubMed: 15483614]
54. Sajda P, Philiastides MG, Parra LC. Single-trial analysis of neuroimaging data: inferring neural networks underlying perceptual decision-making in the human brain. 2009; doi: 10.1109/RBME.2009.2034535
55. Philiastides MG, Sajda P. EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2007; 27(48):13082–91. DOI: 10.1523/JNEUROSCI.3540-07.2007 [PubMed: 18045902]
56. Deuker L, Olligs J, Fell J, Kranz TA, Mormann F, Montag C, Reuter M, Elger CE, Axmacher N. Memory consolidation by replay of stimulus-specific neural activity. *J Neurosci*. 2013; 33(49):19373–19383. DOI: 10.1523/JNEUROSCI.0414-13.2013 [PubMed: 24305832]
57. Fleming S, Huijgen J, Dolan R. Prefrontal contributions to metacognition in perceptual decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2012; 32(18):6117–6125. DOI: 10.1523/JNEUROSCI.6489-11.2012 [PubMed: 22553018]
58. Kiani R, Corthell L, Shadlen MN. Choice certainty is informed by both evidence and decision time. *Neuron*. 2014; 84(6):1329–1342. DOI: 10.1016/j.neuron.2014.12.015 [PubMed: 25521381]
59. Steinhauser M, Yeung N. Decision processes in human performance monitoring. *J Neurosci*. 2010; 30(46):15643–15653. DOI: 10.1523/JNEUROSCI.1899-10.2010 [PubMed: 21084620]
60. Yeung N, Summerfield C. Metacognition in human decision-making: confidence and error monitoring. *Philos Trans R Soc Lond B Biol Sci*. 2012; 367(1594):1310–1321. DOI: 10.1098/rstb.2011.0416 [PubMed: 22492749]
61. Gherman S, Philiastides MG. Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage*. 2015; 106:134–143. DOI: 10.1016/j.neuroimage.2014.11.036 [PubMed: 25463461]
62. Heereman J, Walter H, Heekeren HR. A task-independent neural representation of subjective certainty in visual perception. *Frontiers in Human Neuroscience*. 9(OCTOBER)
63. Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. *Psychol Rev*. 2001; 108(3):624–652. [PubMed: 11488380]
64. Philiastides MG, Ratcliff R, Sajda P. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *Journal of Neuroscience*. 2006; 26(35):8965–8975. DOI: 10.1523/JNEUROSCI.1655-06.2006 [PubMed: 16943552]
65. Mumford JA, Davis T, Poldrack RA. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*. 2014; 103:130–138. DOI: 10.1016/j.neuroimage.2014.09.026 [PubMed: 25241907]

Figure 1. Encoding models based on stimulus derived features versus EEG variability

A, A traditional encoding model used in fMRI analysis extracts a set of features from the stimulus that are potentially representative of low level structure and high level semantics (green box). Weights are learned to model how these stimulus features are encoded in the fMRI BOLD signal. The resulting encoding model is used to make predictions based on how well different voxels predict the features from novel stimuli. For example, one can create maps of the brain that are labeled based on the stimulus features that each voxel represents.

B, The same encoding model concept applied to EEG variability (EEG encoding model). Instead of features being estimated from the stimulus, they are derived from EEG component trial-to-trial variability (as in Fig 2a) with each temporal window representing a different feature (green box). In traditional encoding models, the stimulus features or concepts are derived through non-linear filters [10–15], however in our model, the non-linearity step is produced by the subjects brain response to each stimuli which we are able to record with EEG. Weights are learned so as to model how the EEG variability at a given time window is encoded in the fMRI BOLD. As in the traditional encoding model, predictions on novel stimuli can be done to test the model and results can be used to construct a map in this case a map of the brain that shows the timing of the EEG component variability that each voxels represents.

Figure 2. Paradigm and traditional EEG and fMRI results

A, three choice reaction time task where stimulus evidence for each category is modulated by varying the phase coherence in the images. **B**, Example of face images with high stimulus evidence (high coherence: 50%) and low stimulus evidence (low coherence: 35%). **C**, Behavioral performance shows significant differences, as a function of stimulus evidence, in accuracy ($p < 10^{-12}$, degrees of freedom (dof)=20, paired t-test) and **D**, response time ($p < 10^{-8}$, dof=20 paired t-test) across the group. **E**, fMRI analysis showing cortical areas correlated with high (red) vs. low (blue) stimulus evidence across the entire trial ($Z > 2.57$ with $p < 0.01$ Family-Wise Error cluster corrected). **F**, Grand average stimulus-locked event related potentials (ERPs) for electrode Pz show that differences in stimulus evidence span the time from stimulus to response.

Figure 3. Temporally precise trial-to-trial EEG variability tags brain regions during decision-making

A, Illustration of how trial-to-trial variability of neural activity in spatially distinct cortical areas can be used to tag brain regions. In this hypothetical example Region 1 is involved in sensory encoding while Region 2 integrates sensory evidence to form a decision (in NHP literature, Region 1 might represent MT, while Region 2 LIP). Neural activity across the trial is shown for two stimulus types, one with high sensory evidence for the choice (red curves) and one with low sensory evidence (blue curves). Also shown are two temporal windows (τ_1 and τ_2) that represent different times during the trial. **B**, Linear classifiers are trained to separate trials based on the two levels of stimulus evidence at specific temporal windows. Shown are classifiers (parameterized by weight vectors w_1 and w_2) for two temporal windows (τ_1 and τ_2) with respect to two EEG sensors (for simplicity only two dimensions of the full $N=43$ sensor space are shown). Though the component hyperplane is optimal for the full 43 dimensions, when projected to a line in two dimensions for illustration, it may appear that the separation is sub-optimal). This yields an EEG discriminant component for each temporal window. Variability along these components serves as a unique feature vector for temporally tagging the BOLD datae.g. variability along an EEG component trained with data from τ_1 tags BOLD voxels with time τ_1 while variability along an EEG component trained with data from τ_2 tags them with τ_2 .

Figure 4. EEG discrimination and encoding model results

A, Group average area under the receiver operating curve (AUC) for the sliding window logistic regression EEG discrimination analysis, comparing high versus low stimulus evidence trials; standard error across subjects is shown with shading. **B**, A single subject's discriminating y-value distributions for high (red) and low stimulus evidence (blue) trials for two EEG time points (225ms and 600ms). **C**, Significant fMRI voxels resulting from the group level analysis for the encoding model ($p < 0.01$ TFCE-False Discovery Rate (FDR) corrected). Activity is seen encompassing early visual processing regions, attention networks, and the task positive network. **D**, A random subset of 100 (50 for each stimulus evidence condition) from 700 total trials of the actual (circle) and predicted (diamond) BOLD responses from the encoding model, for an example subject at a single voxel (MNI X/Y/Z mm: $-27/-54/-15$, $r=0.206$, $dof=709$, $p < 10^{-6}$). High and low stimulus evidence trials are shown separately for clarity. **E**, The averaged correlation of the predicted y-values with the true y-values across the trial duration. Blue shading represents the standard error across subjects. Grey shading indicates significant time windows ($p < 0.05$ FDR-corrected).

Figure 5. Group-level encoding model weights results show neural activation cascade

Subset of thresholded ($p < 0.05$ FDR-Corrected, $k=10$) group level statistical parametric maps created by stTFCE randomization procedure on the encoding model weight matrices show the progression of spatial activity across the trial. Activation can be seen early in the trial in the occipital regions while progressing more anteriorly later in the trial to executive control areas. Activations in red indicate areas where high stimulus evidence trials had larger activations than low stimulus evidence trials, and blue the inverse.

Figure 6. Spatiotemporal event-related activations show coordinated reactivations

A, Union across time windows of significant voxels for high (red) and low (blue) stimulus evidence activations. Voxels with activations for both high and low conditions (at different time windows) are displayed in green (also see supplementary material Fig. S2C). Also shown are the encoding model weights for specific voxels, including fusiform gyrus (FG-R): 36/−51/−18, (FG-L): −42/−42/−18, superior lateral occipital cortex (sLOC): 28/−64/38, superior parietal lobule (SPL): 32/−52/54, anterior cingulate cortex (ACC): −8/24/28, intraparietal sulcus (IPS): −30/−60/39, middle frontal gyrus (MFG): −45/27/30, middle temporal gyrus (MT): −57/−60/0. Asterisks indicate significant windows. **B**, Sequence of significant weights showing a “replay” of the network after the onset of ACC activation (shaded ellipse). “Replay” is faster than the initial stimulus driven sequence and strongest for low evidence trials.

Figure 7. Trial-to-trial reactivation correlates with decision confidence

Trial-to-trial reactivation amplitude ($\gamma_{j,i}^R$, see Methods) of “replay” correlates with confidence proxy for both high (**A**) and low (**B**) stimulus evidence conditions. Error bars represent standard errors across subjects. Response locked analysis results are shown in the insets for (**A**) and (**B**) as well as supplementary material Figs. S6–7. **C**, Stimulus-locked replay activation clusters and feature importance. (inset) Regions of interest used in computing the reactivation values for computing confidence proxy correlations. These regions were taken from significant group activations from 600–800ms post stimulus. Regions were then clustered (> 48 voxels) and a secondary analysis for feature importance was performed. Here, we removed each cluster before computing trial-to-trial reactivations and compared the slope of reactivation $\times$ confidence proxy when all clusters were present. Panel **C** shows the ranking of feature importance for each cluster (more negative % change = more importance). Negative changes in slopes indicate that by removing that cluster the slope of the correlation between reactivation and confidence decreases indicating the importance of that cluster.