

1 Title: **Imaging decision-related neural cascades in the human brain**

2

3 Authors: Jordan Muraskin¹, Truman R. Brown², Jennifer M. Walz³, Bryan Conroy⁴,

4 Robin I. Goldman⁵, Paul Sajda¹

5 ¹ Department of Biomedical Engineering, Columbia University, New York, NY, USA

6 10027

7 ² Center for Biomedical Imaging, Medical University of South Carolina, Charleston, SC,

8 USA, 29425

9 ³ Florey Institute of Neuroscience and Mental Health, Melbourne, Australia

10 ⁴ Philips Research, New York, NY,

11 ⁵ Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA,

12 53705

13

14 Corresponding Author: Paul Sajda, psajda@columbia.edu

15

16

17

18

19

20 ***Abstract***

21 Perceptual decisions depend on coordinated patterns of neural activity cascading across
22 the brain, running in time from stimulus to response and in space from primary sensory
23 regions to the frontal lobe. Measuring this cascade and how it flows through the brain is
24 key to developing an understanding of how our brains function. However observing, let
25 alone understanding, this cascade, particularly in humans, is challenging. Here, we report
26 a significant methodological advance allowing this observation in humans at
27 unprecedented spatiotemporal resolution. We use a novel encoding model to link
28 simultaneously measured electroencephalography (EEG) and functional magnetic
29 resonance imaging (fMRI) signals to infer the high-resolution spatiotemporal brain
30 dynamics taking place during rapid visual perceptual decision-making. After
31 demonstrating the methodology replicates past results, we show that it uncovers a
32 previously unobserved sequential reactivation of a substantial fraction of the pre-response
33 network whose magnitude correlates with decision confidence. Our results illustrate that
34 a temporally coordinated and spatially distributed neural cascade underlies perceptual
35 decision-making, with our methodology illuminating complex brain dynamics that would
36 otherwise be unobservable using conventional fMRI or EEG separately. We expect this
37 methodology to be useful in observing brain dynamics in a wide range of other mental
38 processes.

39

40

41 ***Introduction***

42 The detailed spatiotemporal brain dynamics that underlie human decision-making are
43 difficult to measure. Invasive techniques with sufficient temporal or spatial resolution,
44 such as depth electrodes or cortical arrays used with epilepsy patients, are only feasible in
45 rare cases and, in addition, do not capture activity from the entire brain. In comparison,
46 non-invasive measures such as electroencephalography (EEG) and
47 magnetoencephalography (MEG) suffer from poor spatial resolution, and blood oxygen
48 level dependent functional MRI (BOLD fMRI) from poor temporal resolution and
49 indirect coupling to neural activity (e.g. fMRI)¹. In spite of this, EEG, MEG, and fMRI
50 have been used individually to study perceptual decision-making in the human brain,
51 although, by themselves they provide a limited view of the underlying brain dynamics².

52 Recently, methods enabling simultaneous acquisition of EEG and fMRI
53 (EEG/fMRI) have led to varied analytic approaches aimed at integrating the
54 electrophysiological and hemodynamic information contained in the joint measurements.
55 Such approaches offer the potential to provide a comprehensive picture of global brain
56 dynamics, and will likely offer new insights into how the brain makes rapid decisions^{3,4}.
57 Some of the techniques that have been proposed for combining multi-modal brain signals
58 have separately analyzed the EEG and fMRI data and subsequently juxtaposed the
59 results^{5,6}, while others attempt for a truly integrated approach in order to fully exploit the
60 joint information contained in the data sets⁷. In general, simultaneous EEG/fMRI and the
61 associated analysis techniques have been used to identify neuronal sources of EEG trial-
62 to-trial variability, linking them to cognitive processes such as attention⁸ and inhibition⁹.

63 Many previous studies have used known EEG markers (P1, N2, N170, P300, α -
64 rhythm) or data driven approaches such as Independent Component Analysis (ICA) to
65 combine EEG with fMRI data^{4,8-16}. One promising approach has been to use supervised
66 machine-learning techniques (e.g. classifiers) to find relevant projections of the EEG
67 data, where single-trial variability of the electrophysiological response along these
68 projections can be correlated in the fMRI space. Goldman, et al.¹⁷, Walz, et al.¹⁸ and
69 Fouragnan, et al.¹⁹ have demonstrated this technique on visual and auditory paradigms.
70 This methodology has been shown to localize cortical regions that modulate with the task
71 while preserving the temporal progression of task-relevant neural activity.

72 Here we combine a classification methodology with an encoding model that
73 relates the trial-to-trial variability in the EEG to what is observed in the simultaneously
74 acquired fMRI. Encoding models have become an important machine learning tool for
75 analysis of neuroimaging data, specifically fMRI²⁰. In most cases encoding models have
76 been used to learn brain activity that encodes or represents features of a stimulus, such as
77 visual orientation energy in an image/video²¹⁻²³, acoustic spectral power in sound/speech
78²⁴, or visual imagery during sleep²⁵. In the method presented here, we employ an
79 encoding model to directly relate the simultaneously collected data from the two
80 neuroimaging modalities—instead of features derived from the stimulus, they are derived
81 from EEG component trial-to-trial variability. Specifically, we learn an encoding in the
82 spatially precise fMRI data from the temporally precise trial-to-trial variability of EEG
83 activity predictive of the level of stimulus evidence. This approach leverages the fact that
84 the level of stimulus evidence, as measured via EEG, persists across the trial^{26,27}, and

85 that by discriminating this information in a time-localized way, one can temporally “tag”
86 specific cortical areas by their trial-to-trial variability.

87 Using our framework for learning the BOLD signal encoding of task-relevant and
88 temporally precise EEG component variability, we unravel the cascade of activity from
89 the representation of sensory input to decision formation, decision action, and decision
90 monitoring. A particularly novel finding is that after the activation of decision
91 monitoring regions (i.e. ACC), we see a reactivation of pre-response networks, where the
92 strength of this reactivation correlates with measures of decision confidence. This
93 specific reactivation, as well as the entire spatio-temporal cascade, is completely
94 unobservable using conventional fMRI-only or EEG-only methodologies.

95

96 ***Results***

97 In this study, we used a visual alternative forced choice (AFC) task where
98 subjects were shown brief presentations of pictures corrupted by noise and instructed to
99 rapidly discriminate between object categories. On any given trial, the level of noise, or
100 stimulus evidence, was varied randomly. The task itself, as well as similar visual
101 decision-making tasks²⁸, is believed to engage an extensive set of cortical areas in a
102 coordinated fashion, including regions that are responsible for sensory encoding,
103 evidence accumulation, decision formation, and response and decision monitoring.
104 However, the dynamic interplay of these regions has never been observed in humans.
105 Here we exploit previously reported findings regarding the sensitivity of the EEG and
106 fMRI signals to the level of stimulus evidence during a perceptual decision-making task.
107 Specifically, previous work has shown differential neural responses to high vs. low

108 stimulus evidence in trial averaged EEG event-related potentials (ERPs), where this
109 difference persists across the trial^{26,27}. Similarly, fMRI studies have shown that for
110 perceptual decision making tasks a number of spatially-distributed cortical areas
111 significantly correlate with the level of stimulus evidence^{29,30}. We leverage the fact that
112 the level of stimulus evidence is expressed temporally in the EEG and spatially in the
113 fMRI to “tag” voxels with a time. Specifically, using a classification methodology (i.e.
114 discriminative components) we identify temporally precise expressions of the level of
115 stimulus evidence that then can be spatially localized through an encoding model of the
116 fMRI BOLD data.

117 We collected simultaneous EEG/fMRI data from 21 subjects as they performed a
118 3-AFC task discriminating between faces, cars, and houses (Fig. 1A). Subjects were
119 instructed to discriminate the object class after briefly viewing an image corrupted by
120 varying levels of noise (Fig. 1B) and respond by pressing one of three buttons. Overall,
121 subjects responded with accuracies of $94 \pm 5\%$ and $58 \pm 12\%$ and with response times of
122 $634 \pm 82\text{ms}$ and $770 \pm 99\text{ms}$ for high and low stimulus evidence trials, respectively (Fig.
123 1 C, D). Subject accuracies and response times across stimulus types (faces, cars, houses)
124 for low stimulus evidence trials were similar; however, for high stimulus-evidence trials
125 subject accuracies were higher and response times were shorter for faces than for cars or
126 houses (See Supplemental Information Fig. S1).

127

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

130 A traditional general linear model (GLM) analysis of the fMRI (see Methods)
131 revealed differences in BOLD activation between the two stimulus evidence conditions
132 (Fig. 1F, SI Table 1). Brain regions showing greater BOLD activation to high vs. low
133 stimulus evidence trials included areas associated with early visual perception and the
134 default mode network²⁶, such as fusiform gyrus, parahippocampal gyrus, lateral occipital
135 cortex, superior frontal gyrus, and posterior cingulate cortex. Regions with greater BOLD
136 activation to low vs. high stimulus evidence trials included areas in the executive control
137 and difficulty networks' such as dorsal lateral prefrontal cortex, anterior cingulate cortex,
138 intraparietal sulcus, and insula. Overall, these GLM results for the BOLD data
139 reproduced previous results in the literature where similar stimuli and paradigms were
140 used²⁹(Fig. S2A).

141

142 *Extracting temporally localized EEG signatures of stimulus evidence variability*

143 The traditional fMRI results showed multiple brain regions correlated with the
144 difficulty, or stimulus evidence, of the trial; however, this traditional approach does not
145 enable one to infer the relative timing of these fMRI activations. To infer timing at a
146 scale of tens of milliseconds, we used linear classification^{31,32} of the EEG to extract trial-
147 to-trial variability related to stimulus evidence at specified post-stimulus time points.

148 The basic idea is illustrated in Figure 2, where hypothetical neural activity is
149 shown for two different regions that are constituents of the perceptual decision-making
150 network. Averaging over trials would clearly reveal a difference in the mean neural
151 activity between high and low stimulus evidence. However, the two regions contribute
152 differentially to the network, with one region encoding the stimulus evidence (Region 1)

153 and the other integrating it over time (Region 2); both are sensitive to the level of
154 stimulus evidence, though varying so at different times in the trial. By taking
155 advantage of this sensitivity to the stimulus evidence, we can learn EEG discriminant
156 components, i.e. spatial filters, that best classify trials at different time windows given the
157 neural data. We used the trial-to-trial variability along these component directions as
158 features to uniquely tag fMRI voxels with the specific time window of the component.
159 This tagging is done by building an encoding model of the features, given the BOLD
160 signal, details of which are described in the following section.

161 We constructed EEG components by learning linear classifiers at 25ms steps,
162 starting from stimulus onset to 50ms past the average low stimulus evidence response
163 time. We chose a time step of 25ms due to an empirical analysis showing a half width of
164 50ms in the temporal autocorrelation of the EEG data, though in principle this
165 methodology allows for temporal resolution up to the EEG sampling rate. Each classifier
166 was associated with a set of discriminant values, which can be represented as a vector y_{τ} ;
167 each element of the vector is the distance of a given trial to the discrimination boundary
168 for the classifier at time step τ (Fig. 2). This distance can be interpreted as a measure of
169 the EEG classifier's estimate of the level of stimulus evidence for that trial^{17,18,31-34}.

170 Results of the EEG analysis show discriminating information for stimulus
171 evidence spanning the trial (see Fig. 4A), beginning roughly 175ms post-stimulus to past
172 the average response times. A dip occurs around 300ms, indicating stimulus evidence is
173 less discriminative at this time and serves to demarcate early and late cognitive processes.
174 The early process corresponded to the time of the D220 ERP component, which has been
175 shown to modulate with the degree of task difficulty, whether via stimulus noise or task

176 demands³⁵. The later and more prolonged component is likely related to more complex
177 cognitive and motor preparatory processes that differ between high and low stimulus
178 evidence trials. Importantly, although the early and late EEG components were both
179 discriminative, we found their trial-to-trial variability to be uncorrelated (Figs. 4B and
180 S3E), indicating that while the discriminating information (level of stimulus evidence)
181 persists across the trial, it couples differently to processes across time.

182

183 *An encoding model links fMRI activations with temporally distinct EEG trial-to-trial*
184 *variability*

185 After extracting the trial-to-trial variability from the EEG discriminant
186 components, feature vectors y_{τ} are collected across time steps, τ , along with a response
187 time vector to construct a matrix Y . This matrix is the temporally precise representation
188 of the trial-to-trial EEG variability that reflects high vs. low stimulus evidence. An
189 encoding model is then fit, namely a model in which weights are estimated for each time-
190 localized EEG window, to predict the trial-to-trial variability of the BOLD response for
191 each fMRI voxel. Figure 3 shows a schematic of the encoding model framework we used
192 and compares it to a traditional encoding model constructed by using features derived
193 directly from the stimulus. Rather than constructing a map that directly relates each voxel
194 to a type of stimulus feature, such as whether it encodes edges, motion or some semantic
195 concept such as “animal”^{21-23,36-38}, our model is used to construct maps that label voxels
196 by the time window of the variability they encode – i.e. it “tags” each voxel with a
197 “time”, or set of times, when it encodes the variability in the given EEG discriminant
198 component(s).

199 It is important to note that this approach does not attempt to improve source
200 localization typically done for EEG/MEG studies. Our approach instead provides the
201 temporal resolution of EEG (ms) and the spatial resolution of fMRI (mm) without the
202 need to solve the ill-posed inverse solution and make the many associated assumptions
203 required for reliable source-localization results³⁹.

204 An example of the quality of the encoding model is shown in Fig. 4C (see also
205 Fig. S2B) where significant voxels from the encoding model are shown in yellow. Fig.
206 4D shows the trial-to-trial variability of BOLD signal at a specific voxel, comparing it to
207 the variability predicted by the encoding model. Additional validity of the encoding
208 model and single subject results are presented in the Supplemental Information (Fig.
209 S4A/B). The encoding model was also evaluated as a decoding model (see Methods) with
210 the BOLD activity used to predict the trial-to-trial variability in the EEG for unseen
211 data—data on which the encoding model was not trained. Fig. 4E shows these results,
212 expressed as the correlation between the measured and predicted EEG trial-to-trial
213 variability across the 800ms epoch. The shape of the curve is highly consistent with that
214 observed for the EEG data itself (comparing Fig. 4A and Fig. 4E) (additional analysis of
215 the fidelity of the model is provided in the SI, Fig. S3).

216 Given the encoding model, we unwrap the BOLD activity across time by
217 identifying weights that are consistent across subjects in space and time (see Methods).
218 Fig. 5 shows these results for a group level analysis. We observe a progression of activity
219 (see Movie S1), at 25ms resolution, which proceeds simultaneously down the dorsal and
220 ventral streams of visual processing for the first 250ms. After that the cascade becomes
221 more complex with activation in the IPS at 425ms and 750ms (see Fig. 6A), reactivation

222 of the SPL at 675ms and activation of ACC at 600ms along with other regions found in
223 the traditional fMRI results. (see Fig. S5, Tables S2 and an additional analysis using
224 dynamic causal modeling⁴⁰). The reactivation pattern is particularly significant since it
225 would not be observable via a traditional fMRI general linear model (GLM) analysis,
226 which integrates over time and thus superimposes these activities. For example, the
227 changing sign of the middle temporal gyrus (MT) encoding weights in Fig. 6A
228 manifested as no activity in the MT for the traditional fMRI GLM analysis—the change
229 in sign canceled the effective correlation in the GLM (see Fig. 1F and Fig. S1). The areas
230 of activation we find are consistent with previous reports in the literature for human
231 subjects^{29,30}; however, here we are able to link activations across time in a way that was
232 previously only possible with invasive techniques.

233

234 *Cortical reactivation correlates with decision confidence*

235 Further analysis of the spatiotemporal dynamics (see Fig. 6B), shows that the
236 reactivation pattern in the network occurs after decision-monitoring areas become
237 engaged (i.e. after ACC). Spontaneous reactivation, or “replay”, of neural activity in the
238 human brain has been observed and believed to be important for memory consolidation⁴¹
239 and more recently has been hypothesized to play a role in perceptual decision-making by
240 enabling the formation of decision confidence⁴². To test the hypothesis that the
241 reactivation activity we see is in fact related to decision confidence, we used a
242 hierarchical drift diffusion model (DDM)^{43,44} to fit the behavioral data for high and low
243 stimulus evidence conditions (see Methods). Specifically, our model enables us to define
244 a proxy for decision confidence based on the DDM fits to the behavior⁴⁵. Correlating the

245 reactivation level to this confidence proxy shows a strong and significant monotonic
246 relationship between confidence and the level of reactivation (high stimulus evidence-
247 slope= 0.037 ± 0.008 , $t=4.657$, $p=3.2 \times 10^{-6}$; low stimulus evidence-slope= 0.062 ± 0.008 ,
248 $t=7.754$, $p=8.88 \times 10^{-15}$), with low stimulus evidence trials reactivated more strongly than
249 high stimulus evidence trials (difference in slopes= -0.025 ± 0.011 , $t=2.189$, $p=0.029$)(see
250 Fig. 7 and Fig. S7). Additionally, reactivation amplitude correlates with behavioral
251 accuracy (Fig. S8) (high stimulus evidence, slope= 0.0115 ± 0.0047 , $t=2.41$, $p=0.016$; low
252 stimulus evidence, slope= 0.0104 ± 0.0047 , $t=2.19$, $p=0.028$).

253

254

255 *Discussion*

256 We have shown that linking simultaneously acquired EEG and fMRI using a novel
257 encoding model enables imaging of high-resolution spatiotemporal dynamics that
258 underlie rapid perceptual decision-making — decisions made in less than a second. This
259 method, which resolves whole-brain activity with EEG-like temporal resolution, was
260 shown to uncover reactivation processes that would otherwise be masked by the temporal
261 averaging and slow dynamics of traditional fMRI. More broadly, our results
262 demonstrated a general non-invasive data-driven methodology for measuring high
263 spatiotemporal latent neural processes underlying human behavior.

264 This approach temporally “tags” the BOLD fMRI data by encoding the trial-to-
265 trial variability of the temporally precise task relevant components in simultaneously
266 acquired EEG. In effect, the EEG discrimination indexes the activity of interest at high
267 temporal resolution, defining a feature space, and the trial-to-trial variability of these

268 discriminant components becomes the specific feature values used in the encoding model.
269 For the case presented here, this variability was used to tease apart the cascade of activity
270 modulated by stimulus evidence across the trial, and this allowed us to observe, as never
271 seen before, the spatiotemporal brain dynamics underlying a perceptual decision.

272 Previous studies have sought to generalize the timing diagram of a perceptual
273 decision through multi-unit recordings in non-human primates^{46,47} or more broadly in
274 humans^{29,30} using fMRI. Our results confirmed the general temporal ordering of
275 activations found previously (early visual processing, decision formation, decision
276 monitoring). However, there was a possibility the temporal order we observed using our
277 technique was an artifact of our methodology. To assess this possibility, we performed
278 additional analyses using dynamic causal modeling (DCM) to further validate the
279 temporal activation sequence (see Fig. S6) and show, using a different set of assumptions
280 and method, that the temporal sequence we observe is highly likely under a set of
281 alternative sequences. We found that the most likely model is the one consistent with the
282 time course inferred from our encoding model. The DCM results provide additional
283 evidence that the temporal profile uncovered by the encoding model is a likely temporal
284 decomposition of the superimposed fMRI activations.

285 The approach we present requires that EEG and BOLD data be collected
286 simultaneously and not in separate sessions in order to exploit the correlations in trial-to-
287 trial variability to “tag” the BOLD data. To show the importance of collecting the data
288 simultaneously, we ran a control analysis that randomly permuted the trials within their
289 stimulus evidence class, thus effectively simulating an EEG and BOLD dataset collected
290 separately. By destroying the link between the EEG and BOLD trials, the encoding

291 model failed to find any consistent activation (Fig. S9/10), indicating the necessity of
292 simultaneous acquisition.

293 Alternative techniques for fusing simultaneous EEG-fMRI typically do not
294 exploit EEG across the trial and instead only analyze specific ERP components or time
295 windows of interest^{4,8,10,12-19,48,49}. Results from these techniques identify regions that
296 modulate with the specific components, but yield limited information about the timing of
297 other task-relevant regions seen in traditional fMRI contrasts. The methodology
298 developed here extends the work of¹⁷ and Walz, et al.¹⁸ by combining their EEG data
299 reduction techniques with techniques developed for encoding stimulus features onto
300 BOLD data^{20-23,36,38}, ultimately providing a framework for labeling voxels in task-
301 relevant fMRI contrasts with their timing information (Fig. S2C/E/F).

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

310 Our methodology enabled us to observe reactivation of the pre-response network,
311 spatiotemporal dynamics that would be masked using traditional fMRI analysis.
312 Interestingly, the reactivation terminated in a network that included the MFG, insula, and
313 IPS, similar areas previously reported to be reactivated in metacognitive judgments of

314 confidence in perceptual decisions^{42,50,51}. Gherman and Philiastides⁵² observed this
315 network using a multivariate single-trial EEG approach, coupled with a distributed source
316 reconstruction technique. Fleming, et al.⁴² and Heereman, et al.⁵³ used BOLD fMRI to
317 show that areas in this network negatively correlate with subjective certainty ratings.
318 Unique to our findings, we saw this reactivation on a single-trial basis after engagement
319 of the ACC, which has been shown to be involved in decision monitoring^{52,54}, and also
320 observed the dynamic sequence leading up to this network reactivation. Our results
321 showed that reactivation/replay occurred on a trial-to-trial basis after a decision, was
322 stronger for difficult decisions, and correlated with decision confidence.

323 The encoding model we developed was able to decompose traditional fMRI
324 activation maps into their temporal order with significant voxel overlap between the
325 encoding model results and traditional results. The encoding model was also able to show
326 regions that were activated at multiple time points throughout the decision, indicating
327 temporal dynamics that were hidden previously. The regions of activation we found are
328 consistent with earlier findings; however, the work here provided the precise temporal
329 decomposition of these previously reported, temporally superimposed regions of
330 activation. In general, we have shown that simultaneously acquired EEG/fMRI data
331 enables a novel non-invasive approach to visualize high resolution spatial and temporal
332 processing in the human brain with the potential for providing a more comprehensive
333 understanding of the neural basis of complex behaviors.

334

335 *Methods*

336 *Subjects*

337 21 subjects (12 male, 9 female; age range 20-35 years) participated in the study. The
338 Columbia University Institutional Review Board (IRB) approved all experiments and
339 informed consent was obtained before the start of each experiment. All subjects had
340 normal or corrected-to-normal vision.

341 *Stimuli*

342 We used a set of 30 face (from the Max Planck Institute face database), 30 car, and 30
343 house (obtained from the web) gray scale images (image size 512x512 pixels, 8
344 bits/pixel). They were all equated for spatial frequency, luminance, and contrast. The
345 stimulus evidence (high or low) of the task was modulated by systematically modifying
346 the salience of the image via randomization of image phase (35% (low) and 50% (high)
347 coherence)⁵⁵.

348 *Experimental task*

349 The stimuli were used in an event-related three-alternative forced choice (3-AFC) visual
350 discrimination task. On each trial, an image -- either a face, car, or house -- was presented
351 and subjects were instructed to respond with the category of the image by pressing one of
352 three buttons on an MR compatible button controller. Stimuli were presented to subjects
353 using E-Prime software (Psychology Software Tools) and a VisuaStim Digital System
354 (Resonance Technology) with 600x800 goggle display. Over four runs, a total of 720
355 trials were acquired (240 of each category with 120 high coherence trials) with a random
356 inter-trial interval (ITI) sampled uniformly between 2-2.5s. Each run lasted for 560
357 seconds.

358 *fMRI acquisition*

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

365 *EEG acquisition*

366 We simultaneously and continuously recorded EEG using a custom-built MR-compatible
367 EEG system^{56,57}, with differential amplifiers and bipolar EEG montage. The caps were
368 configured with 36 Ag/AgCl electrodes including left and right mastoids, arranged as 43
369 bipolar pairs. Bipolar pair leads were twisted to minimize inductive pickup from the
370 magnetic gradient pulses and subject head motion in the main magnetic field. This
371 oversampling of electrodes ensured data from a complete set of electrodes even in
372 instances when discarding noisy channels was necessary. To enable removal of gradient
373 artifacts in our offline preprocessing, we synchronized the EEG with the scanner clock by
374 sending a transistor– transistor logic pulse at the start of each image volume. All
375 electrode impedances were kept below 20 k Ω , which included 10 k Ω resistors built into
376 each electrode for subject safety.

377 *Functional image pre-processing.*

378 Image preprocessing was performed with FSL (www.fmrib.ox.ac.uk/fsl/). Functional
379 images were spatially realigned to the middle image in the times series (motion-
380 correction), corrected for slice time acquisition, spatially smoothed with a 6mm FWHM
381 Gaussian kernel, and high pass filtered (100s). The structural images were segmented

382 (into grey matter, white matter and cerebro-spinal fluid), bias corrected and spatially
383 normalized to the MNI template using ‘FAST’⁵⁸. Functional images were registered into
384 MNI space using boundary based registration (BBR)⁵⁹.

385

386 *EEG data preprocessing.*

387 We performed standard EEG preprocessing offline using MATLAB (MathWorks) with
388 the following digital Butterworth filters: 0.5 Hz high pass to remove direct current drift,
389 60 and 120 Hz notches to remove electrical line noise and its first harmonic, and 100 Hz
390 low pass to remove high-frequency artifacts not associated with neurophysiological
391 processes. These filters were applied together in the form of a zero-phase finite impulse
392 response filter to avoid distortions caused by phase delays. We extracted stimulus-locked
393 1500 ms epochs (-500:1000) and subtracted the mean baseline – -200 ms to stimulus
394 onset – from the rest of the epoch. Through visual inspection, we discarded trials
395 containing motion and/or blink artifacts, evidenced by sudden high-amplitude
396 deflections.

397 *Sliding window logistic regression.*

398 We used linear discrimination to associate each trial with the level of stimulus evidence
399 represented in the EEG. We considered high stimulus and low stimulus evidence trials
400 irrespective of behavioral accuracy. Regularized logistic regression was used as a
401 classifier to find an optimal projection for discriminating between high and low stimulus
402 evidence trials over a specific temporal window. A sweep of the regularization
403 parameters was implemented using FaSTGLZ⁶⁰. This approach has been previously

404 applied to identify neural components underlying rapid perceptual decision-making

405 17,18,31,33,34,45,49,61

406 Specifically, we defined 50ms duration training windows centered at time, τ ,
407 ranging from stimulus onset to 800ms following the stimulus in 25ms steps. We used
408 logistic regression to estimate a spatial weighting, on N EEG channels, vector (w_τ which
409 is N x 1) that maximally discriminated between EEG sensor array signals E for each class
410 (e.g., high vs. low stimulus evidence trials):

$$411 \quad y_\tau = w_\tau^T E_\tau \quad (1)$$

412 In eqn. 1, E_τ is an N x p vector (N sensors per time window τ by p trials). For our
413 experiments, the center of the window (τ) was varied across the trial in 25ms time-steps.
414 We quantified the performance of the linear discriminator by the area under the receiver
415 operator characteristic (ROC) curve, referred to here as AUC, using a leave-one-out
416 procedure. We used the ROC AUC metric to characterize the discrimination performance
417 as a function of sliding our training window (i.e., varying τ). For each subject, this
418 produced a matrix Y where the rows corresponded to trials and the columns to training
419 windows, i.e. Y is the combination of the calculated y_τ for each time window.

420 *Traditional fMRI analysis.*

421 We first ran a traditional general linear model (GLM) fMRI analysis in FSL, using
422 event-related (high and low stimulus evidence) and response time (RT) variability
423 regressors. The event-related regressors comprised boxcar functions with unit amplitude
424 and onset and offset matching that of the stimuli. RT variability was modeled using the z-
425 scored RT as the amplitude of the boxcars with onset and offset matching that of the
426 stimulus, and these were orthogonalized to the event-related regressors.

427 Orthogonalization was implemented using the Gram-Schmidt procedure⁶² to decorrelate
428 the RT regressor from all other event-related regressors. All regressors were convolved
429 with the canonical hemodynamic response function (HRF), and temporal derivatives
430 were included as confounds of no interest. An event-related high versus low stimulus
431 evidence contrast was also constructed. A fixed-effects model was used to model
432 activations across runs, and a mixed-effects approach was used to compute the contrasts
433 across subjects. Activated regions that passed a family-wise error (FWE)⁶³ corrected
434 cluster threshold of $p < 0.01$ at a z-score threshold of 2.57 were considered significant.
435 *fMRI deconvolution.*

436 Associating fMRI data to each trial is challenging for two main reasons: (a) the temporal
437 dynamics of the hemodynamic response function (HRF) evolve over a longer time-scale
438 than the mean ITI of the event-related design, resulting in overlapping responses between
439 adjacent trials; and (b) the ITI was random for each trial so that the fMRI data was not
440 acquired at a common lag relative to stimulus onset. To overcome these issues, we
441 employed the 'least squares - separate' (LS-S) deconvolution⁶⁴ method to estimate the
442 voxel activations for each trial. For every trial, the time series of each voxel was
443 regressed against a "signal" regressor and a "noise" regressor. The "signal" regressor was
444 the modeled HRF response due to that trial (a delta function centered at stimulus onset
445 convolved with a canonical HRF), while the "noise" regressor was the modeled HRF
446 response due to all other trials (superimposed linearly). The resulting regression
447 coefficients of the "signal" regressor represented the estimated voxel activations due to
448 that trial. These voxel activations were then organized into a single brain volume per trial.
449 We extracted 58697 voxels from a common gray matter group mask at 3 mm³ spatial

450 resolution that excluded white matter and CSF and assembled the resulting voxel
451 activations into rows of the data matrix F .

452 *Single subject encoding model.*

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

$$460 \quad F = WY \quad (2)$$

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

$$462 \quad W = (FY^T)(YY^T)^{-1} \quad (3)$$

463 However, each time point might be highly correlated with its neighbors, which reduces
464 the stability of the least-squares regression. We can use SVD to reduce the feature space
465 and improve our estimation of W (the weights on each window). Then for a leave-one-
466 out cross validation, we hold out a single trial from the EEG Y matrix and the
467 corresponding volume from the fMRI data F and train on the remaining trials. We
468 repeated this for all trials.

$$469 \quad Y^{\text{Train}} = U\Sigma V^T \quad (4)$$

470 Where U is an $r \times r$ orthonormal matrix, Σ is a $r \times p$ diagonal matrix and V is a $p \times p$
471 orthonormal matrix. After SVD on Y^{Train} , we reduced the feature dimensions on Y^{Train} to
472 retain 75% of the variance by only keeping v components. To do this, we selected the

473 first v rows of Σ and zeroed the other rows. We now have $\tilde{\Sigma}$ as our reduced spaced
 474 matrix. If we now recalculate our least squares solution where we have replaced Y by its
 475 reduced form $U\tilde{\Sigma}V^T$ in equation 3:

$$476 \quad \hat{W} = (F^{Train}V\tilde{\Sigma}^T)(\Sigma\Sigma^T)^{-1}U^T \quad (5)$$

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

$$481 \quad \hat{F} = \hat{W}Y^{Test} \quad (6)$$

482 While for the decoding model using the left out test data F^{Test} :

$$483 \quad \hat{Y} = \hat{W}^T F^{Test} (\hat{W}^T \hat{W})^+ \quad (7)$$

484 Here, $\hat{W}^T \hat{W}$ is not invertible, and so we used the pseudo-inverse.

485 At this point, we have \hat{F} , a $m \times p$ matrix with m voxels by p trials. For each voxel
 486 j , we calculated the correlation of \hat{F}_j with F_j , resulting in the matrices R^{fMRI} (Pearson
 487 Correlation Map) and P^{fMRI} (p-value map of the Pearson Correlation) that are $m \times 1$. The
 488 P^{fMRI} was then converted to a z-score map. We constructed the $m \times r$ weight matrix W by
 489 taking the average of all the trained \hat{W} matrices. To test which time windows were
 490 significant, we also calculated, R_{τ}^{EEG} , the correlation between \hat{Y}_{τ} and Y_{τ} .

491 *Group level spatio-temporal analysis.*

492 For group level statistics, we first analyzed the R_{τ}^{EEG} vectors across all subjects. The R_{τ}^{EEG}
 493 vectors were converted into their p-values, and for each time window (τ), used to

494 compute combined Stouffer p-values⁶⁵. These group level results were then false
495 discovery rate corrected (FDR) for multiple comparisons⁶⁶. To identify group level
496 voxels where our model predictions were significant, each subject's p-value maps for the
497 leave-one-out correlation were converted into their respective z-values, and voxel-wise
498 significance was calculated using threshold-free cluster enhancement (TFCE) using a
499 non-parametric randomization procedure implemented in FSL⁶⁷. Voxels were considered
500 significant if they passed a conservative false discovery rate threshold of $p < 0.01$.

501 These significant voxels were then used as a mask to temporally localize
502 activations by computing the voxels that were consistent in their direction (positive (high
503 stimulus evidence) or negative (low stimulus evidence)) and timing (τ window). To this
504 end, we implemented a spatio-temporal TFCE (stTFCE) in both space (neighboring
505 voxels) and time (neighboring time windows - response time window not included) and
506 computed significance through a randomization procedure. 33000 permutations (1000
507 permutations per window) were run by randomly altering the sign of each subject and the
508 temporal ordering of the windows, as we were testing whether the weights were
509 consistent in sign, voxel space, and temporal window. P-values were calculated by
510 comparing the true stTFCE value with the distribution of permuted values. Again, voxels
511 were considered significant if they passed FDR correction at $p < 0.05$ (high stimulus
512 evidence: FDR-Corrected $p < 0.0019$, low stimulus evidence: FDR-Corrected $p < 0.00036$).
513 Note, that now our number of multiple comparisons was the number of voxels in the
514 FDR-mask (20256) times the number of time windows (33). We analyzed the response
515 time separately with a standard TFCE randomization procedure implemented in FSL
516 (Fig. S2D).

517 *Dynamic causal modeling.*

518 To validate the encoding model timing, we implemented single-state linear
519 dynamic causal modeling (DCM) using DCM10 in SPM8⁶⁸, and applied this to the
520 BOLD data to test the hypothesis that the temporal sequence of BOLD activations we
521 found in our EEG-fMRI encoding method was most likely, relative to other possible
522 sequences of these same activations, given only the BOLD data. We used the results of
523 the encoding model to select seven regions of interest that spanned the entire trial. For the
524 first region (labeled 175 in our figures), we computed the union of activations during the
525 175ms and 200ms windows. Activations of the 225ms (225) and 250ms combined with
526 275ms (250) windows become the second and third regions. We computed the union of
527 activations during the 325ms and 350ms windows to create the fourth (325). For the fifth
528 region (400), we computed the union of the activations during the 400ms-450ms
529 windows. For the sixth region (650), we computed the union of the activations during the
530 650ms and 675ms windows. Finally, the union of the activations during the 725-800ms
531 windows was computed to create the seventh region (725). We removed any overlapping
532 voxels between any of the regions and then extracted time series from individual
533 subjects' preprocessed functional data in MNI space by estimation of the first principal
534 component within each region.

535 We constructed 11 models (Figure S6) to investigate the directed connectivity of
536 these regions and validate the temporal ordering found by the encoding model. Each
537 model was feed-forward with first node in each model as the input region. The first
538 model was the temporal ordering of the regions inferred from our EEG-fMRI encoding
539 model analysis. For five of the models, we randomized the temporal ordering of the early

540 regions (175, 225, 250) and the late regions (325, 400, 650, 725) separately. For the other
541 five models, we fully randomized the temporal ordering of all the regions.

542 We used fixed-effects Bayesian model selection (BMS) to compare these 11
543 models both on a single-subject level and at the group level. BMS balances model fit and
544 complexity, thereby selecting the most generalizable model. It estimates the relative
545 model evidence and provides a distribution of posterior probabilities for all of the models
546 considered. We also compared families of similar models⁶⁹; the model space was divided
547 into two families (early/late or fully randomized).

548

549 *Drift Diffusion Model (DDM) and Confidence Proxy.*

550 The DDM models decision-making in two-choice tasks. Here, we treated the decision
551 (correct vs. incorrect) as our two choices. A drift-process accumulates evidence over time
552 until it crosses one of two boundaries (upper or lower) and initiates the corresponding
553 response⁶⁸. The speed with which the accumulation process approaches one of the two
554 boundaries (a) is called drift-rate (v) and represents the relative evidence for or against a
555 particular response. Recently, Philiastides, et al.⁴⁵ showed that for conditions in which
556 the boundary (a) does not change, a proxy for decision confidence for each trial (i) can be
557 computed by $1 / \sqrt{RT_i - T_{non}}$.

558 We used Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python
559 (HDDM) to calculate the drift rate (v), decision boundary (a) and non-decision time T_{non}
560 for each subject⁴³. Specifically, we modeled high and low stimulus evidence response
561 time data separately. This was to ensure our confidence proxies were consistent within
562 trial types. We included the response time and whether the subject got the trial correct.

563 HDDM obtains a sequence of samples (i.e., a Markov chain Monte Carlo; MCMC) from
564 the posterior of each parameter in the DDM. In our model, we generated 5000 samples
565 from the posteriors, the first 1000 (burn-in) samples were discarded, and the remaining
566 samples were thinned by 5%.

567 After modeling the DDM process, each trial's (i) confidence proxy (CP) for each
568 subject (j) was computed by $CP_{i,j} = 1 / \sqrt{RT_i - T_{non,j}}$ and then z-scored across trials where
569 $T_{non,j}$ was varied for high or low stimulus evidence trials, separately. Therefore, CP was a
570 measure of relative trial confidence within difficulty levels.

571

572 *Confidence Proxy and Decision Replay.*

573 Trial to trial reactivation amplitude was defined as $Y_{j,i}^R = W_{j,PostACC}^T F_{j,i}$ for each
574 subject (j) and trial (i), where $W_{postACC}$ is the weight matrix from the encoding model
575 thresholded by voxels that were significant in the group results from the 675-800ms
576 windows. The mean of the $Y_{j,i}^R$ across time becomes a measure of “decision replay”
577 strength for that trial (more negative y's indicate more replay activation, more positive y's
578 indicate less replay activation). $Y_{j,i}^R$ was quintiled for high and low stimulus evidence
579 and the average confidence proxy was calculated within each quintile (Fig. 7). A linear
580 mixed effects model⁷⁰ was used to test if the slope of confidences across quintile
581 grouping, $Y_{j,i}^R$, were significantly different from 0 while including stimulus evidence as a
582 condition. Separate similar analyses with non-replay windows (175-250ms) and testing
583 for behavioral accuracy were also performed (Fig. S7-8).

584

585 *Author Contributions*

586 Conceptualization, J.M. and P.S.; Methodology, J.M., T.R.B., J.W. B.C., R.I.G. and P.S.;

587 Investigation, J.M.; Software, J.M., B.C.; Writing – Original Draft, J.M. and P.S.; Writing

588 – Review & Editing, J.M., T.R.B., R.I.G., J.W., and P.S. ; Funding Acquisition, P.S. ;

589 Resources, J.M., T.R.B., J.W. B.C., R.I.G. and P.S.; Supervision, T.R.B and P.S.

590

591 *Acknowledgements*

592 We would like to thank Jianing Shi for assistance in collecting the EEG/fMRI data. This

593 work was funded by National Institutes of Health Grant R01-MH085092, DARPA under

594 Contract NBCHC090029 and the Army Research Laboratory and under Cooperative

595 Agreement Number W911NF-10-2-0022.

596 *References*

- 597 1 Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**,
- 598 869-878, doi:10.1038/nature06976 (2008).
- 599 2 Alexander, D. M., Trengove, C. & van Leeuwen, C. Donders is dead: cortical
- 600 traveling waves and the limits of mental chronometry in cognitive neuroscience.
- 601 *Cognitive Processing*, doi:10.1007/s10339-015-0662-4 (2015).
- 602 3 Jorge, J. o., van der Zwaag, W. & Figueiredo, P. c. EEG-fMRI integration for the
- 603 study of human brain function. *NeuroImage* **102**, 24--34,
- 604 doi:10.1016/j.neuroimage.2013.05.114 (2014).
- 605 4 Huster, R. J., Debener, S., Eichele, T. & Herrmann, C. S. Methods for
- 606 simultaneous EEG-fMRI: an introductory review. *The Journal of neuroscience :*
- 607 *the official journal of the Society for Neuroscience* **32**, 6053-6060,
- 608 doi:10.1523/jneurosci.0447-12.2012 (2012).
- 609 5 Plichta, M. M. *et al.* Simultaneous EEG and fMRI Reveals a Causally Connected
- 610 Subcortical-Cortical Network during Reward Anticipation. *Journal of*
- 611 *Neuroscience* **33**, 14526-14533, doi:10.1523/jneurosci.0631-13.2013 (2013).
- 612 6 Yuan, H. *et al.* Negative covariation between task-related responses in alpha/beta-
- 613 band activity and BOLD in human sensorimotor cortex: an EEG and fMRI study
- 614 of motor imagery and movements. *NeuroImage* **49**, 2596-2606,
- 615 doi:10.1016/j.neuroimage.2009.10.028 (2010).
- 616 7 Dahne, S. *et al.* Multivariate Machine Learning Methods for Fusing Multimodal
- 617 Functional Neuroimaging Data. *Proceedings of the IEEE* **103**, 1507-1530,
- 618 doi:10.1109/JPROC.2015.2425807 (2015).

- 619 8 Warbrick, T., Arrubla, J., Boers, F., Neuner, I. & Shah, N. J. J. Attention to
620 Detail: Why Considering Task Demands Is Essential for Single-Trial Analysis of
621 BOLD Correlates of the Visual P1 and N1. *Journal of cognitive neuroscience* **26**,
622 1--14, doi:10.1162/jocn (2013).
- 623 9 Baumeister, S. *et al.* Sequential inhibitory control processes assessed through
624 simultaneous EEG-fMRI. *NeuroImage*, doi:10.1016/j.neuroimage.2014.01.023
625 (2014).
- 626 10 Novitskiy, N., Ramautar, J. R. & Vanderperren, K. a. The BOLD correlates of the
627 visual P1 and N1 in single-trial analysis of simultaneous EEG-fMRI recordings
628 during a spatial detection task. *NeuroImage* **54**, 824--835,
629 doi:10.1016/j.neuroimage.2010.09.041 (2010).
- 630 11 Nguyen, V. T. & Cunnington, R. The superior temporal sulcus and the N170
631 during face processing: Single trial analysis of concurrent EEG-fMRI.
632 *NeuroImage*, doi:10.1016/j.neuroimage.2013.10.047 (2013).
- 633 12 De Martino, F. *et al.* Multimodal imaging: an evaluation of univariate and
634 multivariate methods for simultaneous EEG/fMRI. *Magnetic resonance imaging*
635 **28**, 1104-1112, doi:10.1016/j.mri.2009.12.026 (2010).
- 636 13 Mayhew, S. D., Ostwald, D., Porcaro, C. & Bagshaw, A. P. Spontaneous EEG
637 alpha oscillation interacts with positive and negative BOLD responses in the
638 visual-auditory cortices and default-mode network. *NeuroImage* **76**, 362-372,
639 doi:10.1016/j.neuroimage.2013.02.070 (2013).
- 640 14 Jann, K. *et al.* BOLD correlates of EEG alpha phase-locking and the fMRI default
641 mode network. *NeuroImage* **45**, 903-916, doi:10.1016/j.neuroimage.2009.01.001
642 (2009).
- 643 15 Jaspers-Fayer, F., Ertl, M., Leicht, G., Leupelt, A. & Mulert, C. Single-trial EEG-
644 fMRI coupling of the emotional auditory early posterior negativity. *NeuroImage*
645 **62**, 1807-1814, doi:10.1016/j.neuroimage.2012.05.018 (2012).
- 646 16 Omata, K., Hanakawa, T., Morimoto, M. & Honda, M. Spontaneous Slow
647 Fluctuation of EEG Alpha Rhythm Reflects Activity in Deep-Brain Structures: A
648 Simultaneous EEG-fMRI Study. *PloS one* **8**, e66869-e66869,
649 doi:10.1371/journal.pone.0066869 (2013).
- 650 17 Goldman, R. I. *et al.* Single-trial discrimination for integrating simultaneous EEG
651 and fMRI: identifying cortical areas contributing to trial-to-trial variability in the
652 auditory oddball task. *NeuroImage* **47**, 136-147,
653 doi:10.1016/j.neuroimage.2009.03.062 (2009).
- 654 18 Walz, J. M. *et al.* Simultaneous EEG-fMRI Reveals Temporal Evolution of
655 Coupling between Supramodal Cortical Attention Networks and the Brainstem.
656 *The Journal of neuroscience : the official journal of the Society for Neuroscience*
657 **33**, 19212-19222, doi:10.1523/jneurosci.2649-13.2013 (2013).
- 658 19 Fouragnan, E., Retzler, C., Mullinger, K. & Philiastides, M. G. Two
659 spatiotemporally distinct value systems shape reward-based learning in the human
660 brain. *Nature communications* **6**, 8107, doi:10.1038/ncomms9107 (2015).
- 661 20 Naselaris, T., Kay, K. N., Nishimoto, S. & Gallant, J. L. Encoding and decoding
662 in fMRI. *NeuroImage* **56**, 400-410, doi:10.1016/j.neuroimage.2010.07.073
663 (2011).

- 664 21 Hansen, K. A., Kay, K. N. & Gallant, J. L. Topographic organization in and near
665 human visual area V4. *Journal of Neuroscience* **27**, 11896--11911,
666 doi:10.1523/JNEUROSCI.2991-07.2007 (2007).
- 667 22 Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. Identifying natural
668 images from human brain activity. *Nature* **452**, 352--355,
669 doi:10.1038/nature06713 (2008).
- 670 23 Nishimoto, S. *et al.* Reconstructing visual experiences from brain activity evoked
671 by natural movies. *Current Biology* **21**, 1641--1646,
672 doi:10.1016/j.cub.2011.08.031 (2011).
- 673 24 Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D. & Hasson, U. Coupled neural
674 systems underlie the production and comprehension of naturalistic narrative
675 speech. *Proceedings of the National Academy of Sciences* **111**, E4687--E4696,
676 doi:10.1073/pnas.1323812111 (2014).
- 677 25 Horikawa, T., Tamaki, M., Miyawaki, Y. & Kamitani, Y. Neural decoding of
678 visual imagery during sleep. *Science (New York, N.Y.)* **340**, 639-642,
679 doi:10.1126/science.1234330 (2013).
- 680 26 Philiastides, M. G., Ratcliff, R. & Sajda, P. Neural representation of task
681 difficulty and decision making during perceptual categorization: a timing
682 diagram. *The Journal of neuroscience : the official journal of the Society for*
683 *Neuroscience* **26**, 8965-8975, doi:10.1523/JNEUROSCI.1655-06.2006 (2006).
- 684 27 Banko, E. M., Gal, V., Kortvelyes, J., Kovacs, G. & Vidnyanszky, Z. Dissociating
685 the effect of noise on sensory processing and overall decision difficulty. *The*
686 *Journal of neuroscience : the official journal of the Society for Neuroscience* **31**,
687 2663-2674, doi:10.1523/JNEUROSCI.2725-10.2011 (2011).
- 688 28 Erickson, D. T. & Kayser, A. S. The neural representation of sensorimotor
689 transformations in a human perceptual decision making network. *NeuroImage*
690 **79C**, 340-350, doi:10.1016/j.neuroimage.2013.04.085 (2013).
- 691 29 Heekeren, H. R., Marrett, S., Bandettini, P. a. & Ungerleider, L. G. A general
692 mechanism for perceptual decision-making in the human brain. *Nature* **431**, 859-
693 862, doi:10.1038/nature02966 (2004).
- 694 30 Philiastides, M. G. & Sajda, P. EEG-informed fMRI reveals spatiotemporal
695 characteristics of perceptual decision making. *The Journal of neuroscience : the*
696 *official journal of the Society for Neuroscience* **27**, 13082-13091,
697 doi:10.1523/JNEUROSCI.3540-07.2007 (2007).
- 698 31 Parra, L. C., Spence, C. D., Gerson, A. D. & Sajda, P. Recipes for the linear
699 analysis of EEG. *NeuroImage* **28**, 326-341,
700 doi:10.1016/j.neuroimage.2005.05.032 (2005).
- 701 32 Sajda, P., Philiastides, M. G. & Parra, L. C. Single-trial analysis of neuroimaging
702 data: inferring neural networks underlying perceptual decision-making in the
703 human brain. *IEEE Rev Biomed Eng* **2**, 97-109,
704 doi:10.1109/RBME.2009.2034535 (2009).
- 705 33 Muraskin, J., Sherwin, J. & Sajda, P. Knowing when not to swing: EEG evidence
706 that enhanced perception-action coupling underlies baseball batter expertise.
707 *NeuroImage* **123**, 1-10, doi:10.1016/j.neuroimage.2015.08.028 (2015).

- 708 34 Sherwin, J., Muraskin, J. & Sajda, P. You Can't Think and Hit at the Same Time:
709 Neural Correlates of Baseball Pitch Classification. *Frontiers in neuroscience* **6**,
710 177, doi:10.3389/fnins.2012.00177 (2012).
- 711 35 Philiastides, M. G. & Sajda, P. Temporal characterization of the neural correlates
712 of perceptual decision making in the human brain. *Cereb Cortex* **16**, 509-518,
713 doi:10.1093/cercor/bhi130 (2006).
- 714 36 Cukur, T., Nishimoto, S., Huth, A. G. & Gallant, J. L. Attention during natural
715 vision warps semantic representation across the human brain. *Nature*
716 *neuroscience* **16**, 763-770, doi:10.1038/nn.3381 (2013).
- 717 37 Naselaris, T., Kay, K. N., Nishimoto, S. & Gallant, J. L. Vol. 56 400-410
718 (2011).
- 719 38 Stansbury, D., Naselaris, T. & Gallant, J. Natural Scene Statistics Account for the
720 Representation of Scene Categories in Human Visual Cortex. *Neuron* **79**, 1025--
721 1034, doi:10.1016/j.neuron.2013.06.034 (2013).
- 722 39 Wendel, K. *et al.* EEG/MEG source imaging: methods, challenges, and open
723 issues. *Computational intelligence and neuroscience*, 656092,
724 doi:10.1155/2009/656092 (2009).
- 725 40 Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *NeuroImage*
726 **19**, 1273--1302, doi:10.1016/S1053-8119(03)00202-7 (2003).
- 727 41 Deuker, L. *et al.* Memory consolidation by replay of stimulus-specific neural
728 activity. *The Journal of neuroscience : the official journal of the Society for*
729 *Neuroscience* **33**, 19373-19383, doi:10.1523/JNEUROSCI.0414-13.2013 (2013).
- 730 42 Fleming, S. M., Huijgen, J. & Dolan, R. J. Prefrontal contributions to
731 metacognition in perceptual decision making. *The Journal of neuroscience : the*
732 *official journal of the Society for Neuroscience* **32**, 6117-6125,
733 doi:10.1523/JNEUROSCI.6489-11.2012 (2012).
- 734 43 Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: Hierarchical Bayesian estimation
735 of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics* **7**, 14,
736 doi:10.3389/fninf.2013.00014 (2013).
- 737 44 Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-
738 choice decision tasks. *Neural Comput* **20**, 873--922, doi:10.1162/neco.2008.12-
739 06-420 (2008).
- 740 45 Philiastides, M. G., Heekeren, H. R. & Sajda, P. Human Scalp Potentials Reflect a
741 Mixture of Decision-Related Signals during Perceptual Choices. *Journal of*
742 *Neuroscience* **34**, 16877--16889, doi:10.1523/JNEUROSCI.3012-14.2014 (2014).
- 743 46 Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during
744 flexible sensorimotor decisions. *Science* **348**, 1352-1355,
745 doi:10.1126/science.aab0551 (2015).
- 746 47 Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annual review*
747 *of neuroscience* **30**, 535-574, doi:10.1146/annurev.neuro.29.051605.113038
748 (2007).
- 749 48 Warbrick, T. *et al.* Single-trial P3 amplitude and latency informed event-related
750 fMRI models yield different BOLD response patterns to a target detection task.
751 *NeuroImage* **47**, 1532-1544, doi:10.1016/j.neuroimage.2009.05.082 (2009).

- 752 49 Walz, J. M. *et al.* Simultaneous EEG-fMRI reveals a temporal cascade of task-
753 related and default-mode activations during a simple target detection task.
754 *NeuroImage* **102 Pt 1**, 229-239, doi:10.1016/j.neuroimage.2013.08.014 (2014).
- 755 50 Yeung, N. & Summerfield, C. Metacognition in human decision-making:
756 confidence and error monitoring. *Philosophical transactions of the Royal Society*
757 *of London. Series B, Biological sciences* **367**, 1310-1321,
758 doi:10.1098/rstb.2011.0416 (2012).
- 759 51 Steinhäuser, M. & Yeung, N. Decision processes in human performance
760 monitoring. *The Journal of neuroscience : the official journal of the Society for*
761 *Neuroscience* **30**, 15643-15653, doi:10.1523/JNEUROSCI.1899-10.2010 (2010).
- 762 52 Gherman, S. & Philiastides, M. G. Neural representations of confidence emerge
763 from the process of decision formation during perceptual choices. *NeuroImage*
764 **106**, 134-143, doi:10.1016/j.neuroimage.2014.11.036 (2015).
- 765 53 Heereman, J., Walter, H. & Heekeren, H. R. A task-independent neural
766 representation of subjective certainty in visual perception. *Front Hum Neurosci* **9**,
767 551, doi:10.3389/fnhum.2015.00551 (2015).
- 768 54 Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D.
769 Conflict monitoring and cognitive control. *Psychological review* **108**, 624-652
770 (2001).
- 771 55 Dakin, S. C., Hess, R. F., Ledgeway, T. & Achtman, R. L. What causes non-
772 monotonic tuning of fMRI response to noisy images? *Current biology : CB* **12**,
773 R476-477; author reply R478 (2002).
- 774 56 Sajda, P., Goldman, R. I., Dyrholm, M. & Brown, T. R. *Signal Processing and*
775 *Machine Learning for Single-trial Analysis of Simultaneously Acquired EEG and*
776 *fMRI*. (Elsevier Inc., 2010).
- 777 57 Sajda, P., Goldman, R. I., Philiastides, M. G., Gerson, A. D. & Brown, T. R. A
778 System for Single-trial Analysis of Simultaneously Acquired EEG and fMRI.
779 *2007 3rd International IEEE/EMBS Conference on Neural Engineering*,
780 doi:10.1109/CNE.2007.369667 (2007).
- 781 58 Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a
782 hidden Markov random field model and the expectation-maximization algorithm.
783 *IEEE transactions on medical imaging* **20**, 45--57, doi:10.1109/42.906424 (2001).
- 784 59 Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using
785 boundary-based registration. *NeuroImage* **48**, 63--72,
786 doi:10.1016/j.neuroimage.2009.06.060 (2009).
- 787 60 Conroy, B. R., Walz, J. M. & Sajda, P. Fast bootstrapping and permutation testing
788 for assessing reproducibility and interpretability of multivariate FMRI decoding
789 models. *PloS one* **8**, e79271, doi:10.1371/journal.pone.0079271 (2013).
- 790 61 Sherwin, J. & Sajda, P. Musical experts recruit action-related neural structures in
791 harmonic anomaly detection: evidence for embodied cognition in expertise. *Brain*
792 *and Cognition* **83**, 190-202, doi:10.1016/j.bandc.2013.07.002 (2013).
- 793 62 Strang, G. Introduction to Linear Algebra. *Mathematics of Computation* **18**, 510,
794 doi:10.2307/2003783 (2003).
- 795 63 Nichols, T. & Hayasaka, S. Controlling the familywise error rate in functional
796 neuroimaging: a comparative review. *Statistical methods in medical research* **12**,
797 419--446, doi:10.1191/0962280203sm341ra (2003).

- 798 64 Mumford, J. A., Turner, B. O., Ashby, F. G. & Poldrack, R. A. Deconvolving
799 BOLD activation in event-related designs for multivoxel pattern classification
800 analyses. *NeuroImage* **59**, 2636-2643, doi:10.1016/j.neuroimage.2011.08.076
801 (2012).
- 802 65 Darlington, R. B. & Hayes, A. F. Combining independent p values: extensions of
803 the Stouffer and binomial methods. *Psychological methods* **5**, 496-515,
804 doi:10.1037/1082-989X.5.4.496 (2000).
- 805 66 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical
806 and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical*
807 *Society. Series B (Methodological)* **57**, 289 -- 300, doi:10.2307/2346101 (1995).
- 808 67 Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: addressing
809 problems of smoothing, threshold dependence and localisation in cluster
810 inference. *NeuroImage* **44**, 83-98, doi:10.1016/j.neuroimage.2008.03.061 (2009).
- 811 68 Stephan, K. E. *et al.* Ten simple rules for dynamic causal modeling. *NeuroImage*
812 **49**, 3099--3109, doi:10.1016/j.neuroimage.2009.11.015 (2010).
- 813 69 Penny, W. D. *et al.* Comparing families of dynamic causal models. *PLoS*
814 *Computational Biology* **6**, doi:10.1371/journal.pcbi.1000709 (2010).
- 815 70 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects
816 Models Using lme4. *2015* **67**, 48, doi:10.18637/jss.v067.i01 (2015).
817
818

819

820 **Figure Captions**

821 **Figure 1. Paradigm and traditional EEG and fMRI results**

822 **A**, 3-AFC task where stimulus evidence for each category is modulated by varying the
823 phase coherence in the images. **B**, Example of face images with high stimulus evidence
824 (high coherence: 50%) and low stimulus evidence (low coherence: 35%). **C**, Behavioral
825 performance shows significant differences, as a function of stimulus evidence, in
826 accuracy ($p < 10^{-12}$, paired t-test) and **D**, response time ($p < 10^{-8}$, paired t-test) across the
827 group. **E**, Grand average stimulus-locked event related potentials (ERPs) for electrode Pz
828 show that differences in stimulus evidence span the time from stimulus to response. **F**,
829 fMRI analysis showing cortical areas correlated with high (red) vs. low (blue) stimulus
830 evidence across the entire trial ($Z > 2.57$ with $p < 0.01$ Family-Wise Error cluster
831 corrected).

832 **Figure 2. Temporally precise trial-to-trial EEG variability tags brain regions during**
833 **decision-making**

834 **A**, Illustration of how trial-to-trial variability of neural activity in spatially distinct
835 cortical areas can be used to tag brain regions. In this hypothetical example Region 1 is
836 involved in sensory encoding while Region 2 integrates sensory evidence to form a
837 decision (in NHP literature, Region 1 might represent MT, while Region 2 LIP). Neural
838 activity across the trial is shown for two stimulus types, one with high sensory evidence
839 for the choice (red curves) and one with low sensory evidence (blue curves). Also
840 shown are two temporal windows (τ_1 and τ_2) that represent different times during the
841 trial. **B**, Linear classifiers are trained to separate trials based on the two levels of stimulus
842 evidence at specific temporal windows. Shown are classifiers (parameterized by weight

843 | vectors w_1 and w_2) for two temporal windows (τ_1 and τ_2) with respect to two EEG sensors
844 | (for simplicity only two dimensions of the full $N=43$ sensor space are shown. Though
845 | the component hyperplane is optimal for the full 43 dimensions, when projected to a line
846 | in two dimensions for illustration, it may appear that the separation is sub-optimal). This
847 | yields an EEG discriminant component for each temporal window. Variability along
848 | these components serves as a unique feature vector for temporally tagging the BOLD
849 | data—e.g. variability along an EEG component trained with data from τ_1 tags BOLD
850 | voxels with time τ_1 while variability along an EEG component trained with data from τ_2
851 | tags them with τ_2 .

852

853 | **Figure 3. Encoding models based on stimulus derived features versus EEG**

854 | **variability**

855 | **A**, A traditional encoding model used in fMRI analysis extracts a set of features from the
856 | stimulus that are potentially representative of low level structure and high level semantics
857 | (green box). Weights are learned to model how these stimulus features are encoded in
858 | the fMRI BOLD signal. The resulting encoding model is used to make predictions based
859 | on how well different voxels predict the features from novel stimuli. For example, one
860 | can create maps of the brain that are labeled based on the stimulus features that each
861 | voxel represents. **B**, The same encoding model concept applied to EEG variability (EEG
862 | encoding model). Instead of features being estimated from the stimulus, they are derived
863 | from EEG component trial-to-trial variability (as in Fig 2a) with each temporal window
864 | representing a different feature (green box). Weights are learned so as to model how the
865 | EEG variability at a given time window is encoded in the fMRI BOLD. As in the

866 traditional encoding model, predictions on novel stimuli can be done to test the model
867 and results can be used to construct a map—in this case a map of the brain that shows
868 the timing of the EEG component variability that each voxels represents.

869

870 **Figure 4. EEG discrimination and encoding model results**

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

886

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

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

894

895 **Figure 6. Spatial-temporal event-related activations show coordinated reactivations.**

896 **A**, Union across time windows of significant voxels for high (red) and low (blue)
897 stimulus evidence activations. Voxels with activations for both high and low conditions
898 (at different time windows) are displayed in green. Also shown are the encoding model
899 weights for specific voxels, including fusiform gyrus (FG-R):36/-51/-18, (FG-L):-42/-
900 42/-18, superior lateral occipital cortex (sLOC):24/-63/36, superior parietal lobule
901 (SPL):27/-51/54, anterior cingulate cortex (ACC):-6/24/30, intraparietal sulcus (IPS):-
902 30/-60/39, middle frontal gyrus (MFG):-45/27/30, middle temporal gyrus (MT):-57/-
903 60/0. Asterisks indicate significant windows. **B**, Sequence of significant weights showing
904 a “replay” of the network after the onset of ACC activation (shaded ellipse). “Replay” is
905 faster than the initial stimulus driven sequence and strongest for low evidence trials.
906

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

908 Trial-to-trial reactivation amplitude ($Y_{j,i}^R$ – see Methods) of “replay” correlates with
909 confidence proxy for both high and low stimulus evidence conditions. Error bars
910 represent standard errors across subjects.

911 **Figures:**

912

913

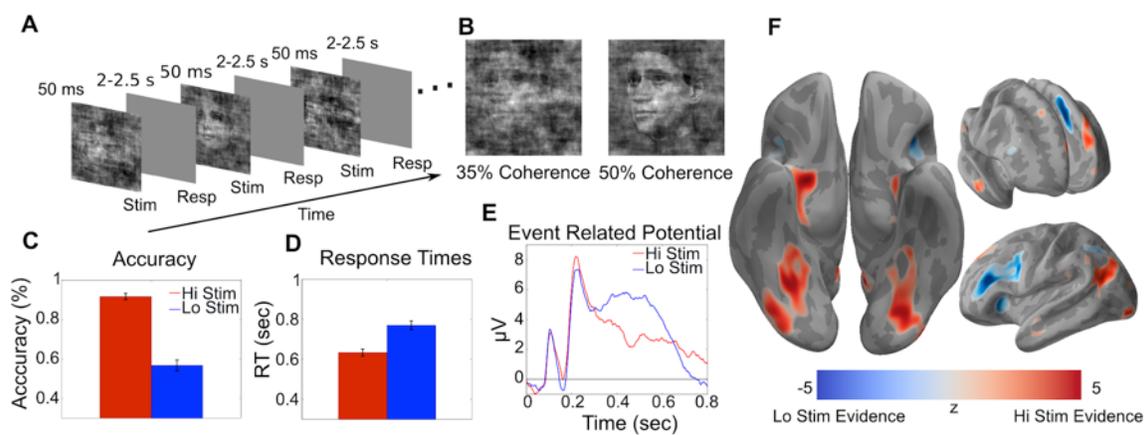
914

915

916

917

918



919

920 *Figure 1.*

921

922

923

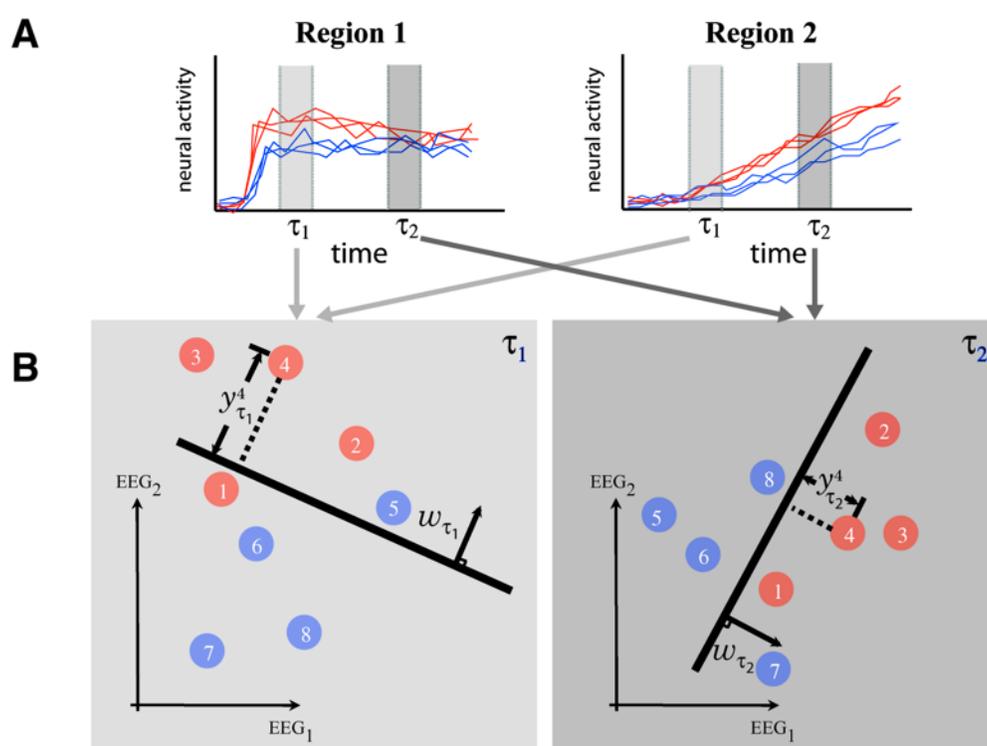
924

925

926

927

928



929

930 *Figure 2.*

931

932

933

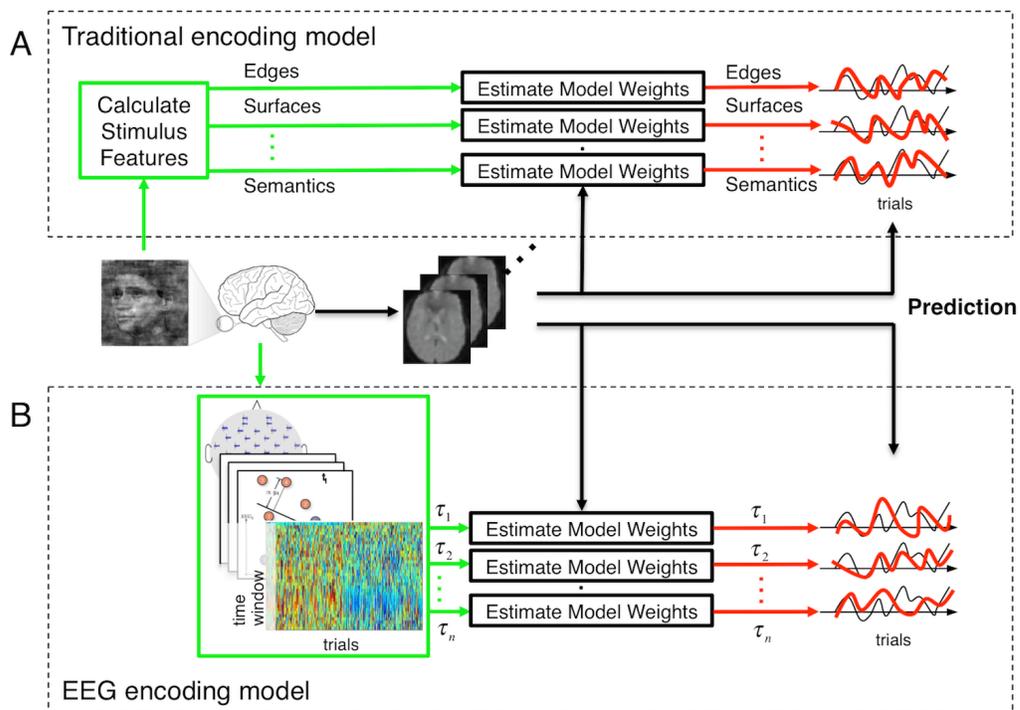
934

935

936

937

938



939

940 *Figure 3.*

941

942

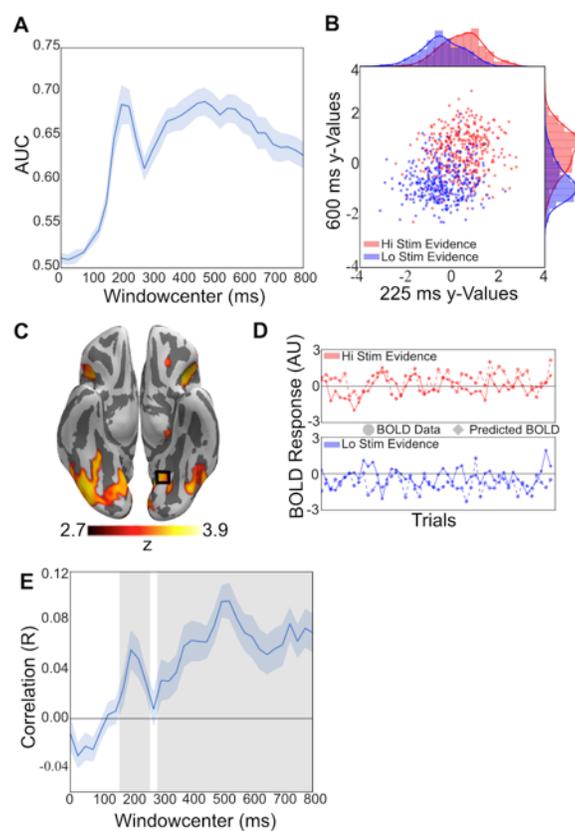
943

944

945

946

947



948

949 *Figure 4.*

950

951

952

953

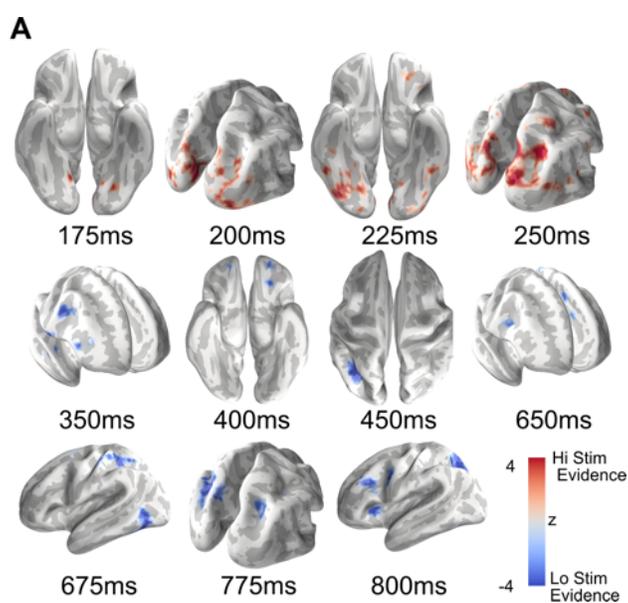
954

955

956

957

958



959

960 *Figure 5.*

961

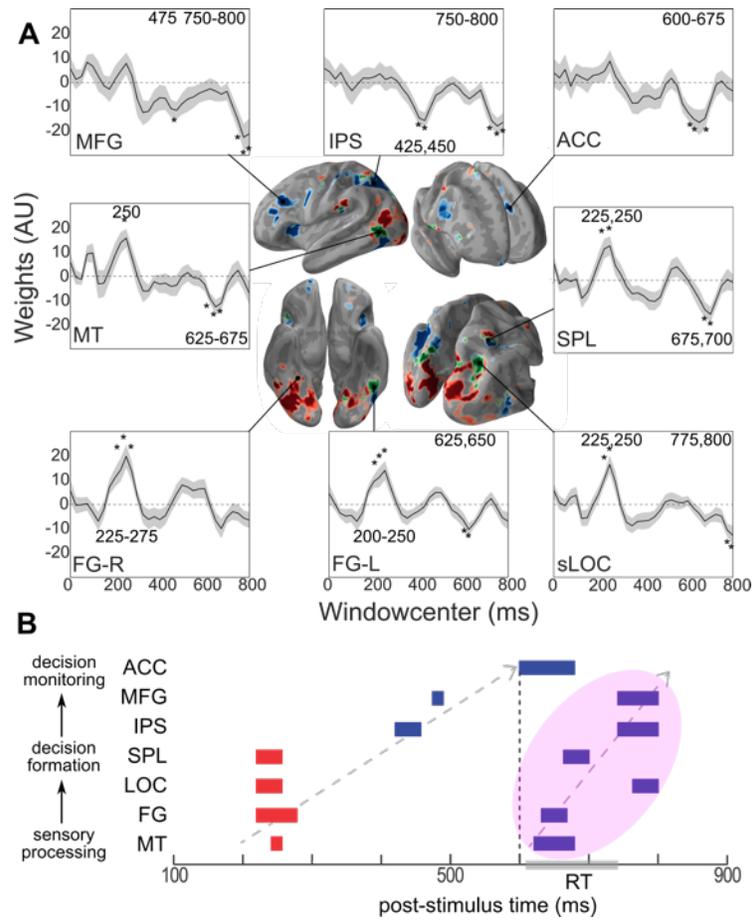
962

963

964

965

966



967

968 *Figure 6.*

969

970

971

972

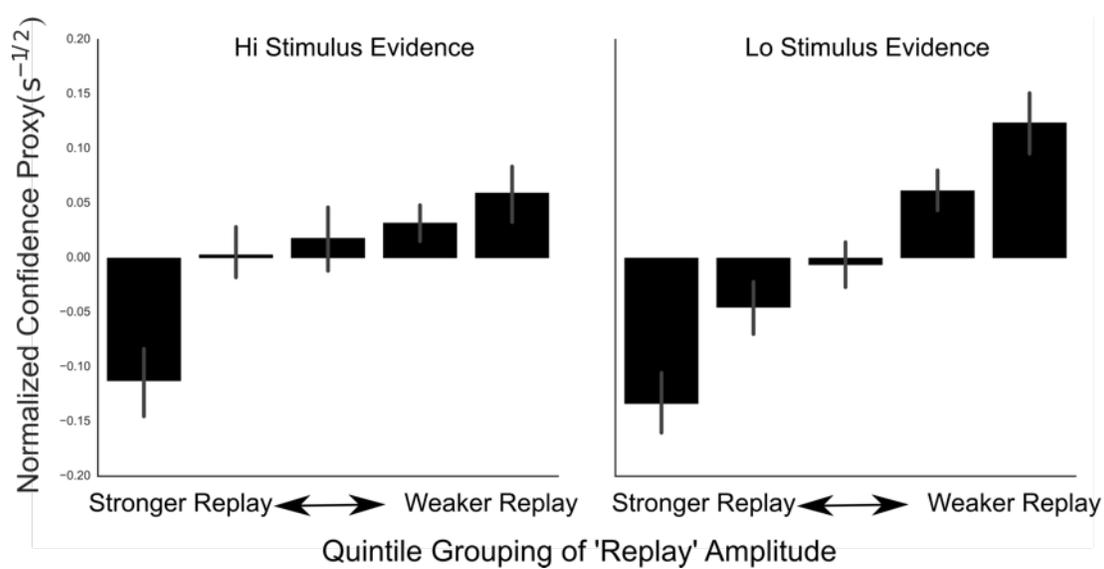
973

974

975

976

977



978

979 *Figure 7.*