nature human behaviour

SUPPLEMENTARY INFORMATION

VOLUME: 1 | ARTICLE NUMBER: 0139

In the format provided by the authors and unedited.

Perceptual confidence neglects decision-incongruent evidence in the brain

Megan A. K. Peters*[1], Thomas Thesen*[2,3,4], Yoshiaki D. Ko*[5], Brian Maniscalco[6], Chad Carlson[2], Matt Davidson[5], Werner Doyle[2], Ruben Kuzniecky[2], Orrin Devinsky[2], Eric Halgren[3], & Hakwan Lau[1,7]

1. Department of Psychology, University of California, Los Angeles, Los Angeles, California, USA
2. Comprehensive Epilepsy Center, Department of Neurology, New York University Medical Center, New York, New York, USA
3. Multimodal Imaging Laboratory, University of California, San Diego, La Jolla, California, USA
4. Department of Physiology & Neuroscience, St. George's University, Grenada, West Indies
5. Department of Psychology, Columbia University, New York, New York, USA
6. Neuroscience Institute, New York University, New York, New York, USA
7. Brain Research Institute, University of California, Los Angeles, Los Angeles, California, USA

* these authors contributed equally to this work

Correspondence should be addressed to:

Megan A. K. Peters
1285 Franz Hall, Box 951563
University of California, Los Angeles
Los Angeles, CA 90095
(323) 596-1093
meganakpeters@ucla.edu

Supplementary Methods

Behavioral methods

We studied six patients (5 females, 1 male, age range 19-46, all right handed) at the Comprehensive Epilepsy Center of New York University who had surgically implanted intracranial electrodes for monitoring for potential resection due to epilepsy (Supplementary Table 1). The electrodes were implanted on the cortical surface for clinical reasons independent of this research. Sample size was thus determined by availability of patients and data acquisition particulars. This study was approved by the New York University Medical Center ethics board, and all patients gave written consent to participate.

In each trial of the behavioral task, 2s of a fixation point were presented, after which the fixation point disappeared and either a face or a house was presented for 16ms. After that, the screen went blank until the subject made a response or 3.5s had elapsed. Responses were made via 4 keys, representing each combination of face/house and high/low confidence. Responses were made with one hand (Figure 1a, main text). All stimuli were presented on a portable laptop with gamma-corrected screen luminance, and responses were collected via the keyboard via button presses made with one hand. All stimuli were converted to grayscale and matched for size, luminance, contrast, and spectral power. They were then windowed with a blurred oval mask to minimize border effects, and covered with randomly generated noise pixels at run-time.

Subjects first underwent a psychophysics threshold estimation procedure to determine stimuli contrasts that would titrate objective performance at about 75% correct split across two contrast levels. Stimulus contrasts were titrated during a thresholding procedure such that subjects would perform approximately 75% correct in the face/house discrimination task. Two randomly-interleaved staircases presented the stimuli while modulating the contrasts to determine each individual patients' psychometric function. One staircase was a 3-down-1-up and the other was a 2-down-1-up, which were expected to converge on 79% and 71% correct, respectively[1,2]. The staircases started at 60% and 40% contrast, and changed by 6% until 4 reversals occurred, after which they changed by 1%. Thresholding continued until the patients had made 8 reversals in each staircase, at which point the averages of the final 8 inflection points for each staircase were chosen as fixed stimulus contrasts for the main experiment. During the subsequent trials, stimuli were presented at one of the two staircase-determined contrasts in a counterbalanced manner.

Subjects completed an average of 415.17±183.93 face/house discrimination trials each (because of the clinical setting in which these studies were conducted, it is difficult to control the number of trials completed by each subject).

Electrocorticography acquisition and preprocessing

Usable signal was recorded from 874 electrodes across all six subjects ($\mu = 145.7$ electrodes per subject). Three of six subjects had electrodes in one hemisphere only, while the other three had bilaterally implanted electrodes. Subjects demonstrated similarly distributed electrode spread (Supplementary Table 2). MNI coordinates for all electrodes for all subjects are included as a supplementary spreadsheet, and data is available upon request.

ECoG was measured using a custom-built system based on the open-source N-Spike acquisition system, with up to 256 simultaneous channels at a 30 kHz, 16-bit sampling rate. Built-in filters high-pass filtered at .6 Hz, and low-pass filtered at 10 kHz. Signals were subsequently downsampled and converted to 500 Hz, 32-bit floating data prior to preprocessing. After downsampling to 500 Hz, individual trials and channels were manually inspected for artifacts and epileptic activity and removed if necessary as routine procedure at the Schwartz Health Care Center (HCC) of the New York University Medical Center (NYUMC) and the NYUMC Comprehensive Epilepsy Center, and again by the present research team. Thus, epileptic electrodes were not included in the analysis. Line noise and harmonics were removed by notch filters at 60, 120 and 180 Hz. Spectral power was computed via a short-time Fourier transform using the multitaper method[3] in MATLAB R2013a (MathWorks; Natuck, MA). We time-locked the activity for each trial to stimulus onset. The spectrum was normalized using single-trial baseline normalization using the period between 500 and 1500ms before stimulus onset as baseline[4]: for each frequency-time point in each trial, the mean baseline power for that frequency band was subtracted from the raw power, and the difference was then divided by the standard deviation of the baseline power in that frequency across trials. The log was taken to make the power distribution approximately normal. Epochs for decoding were defined as the time period from 250ms before stimulus onset to 1000ms after stimulus onset.

All electrodes' MNI coordinates are available as a supplementary dataset.

Support vector machine decoding

We explored the extent to which two trial-by-trial behavioral factors y could be classified by support vector machine (SVM) estimators (see next section): (1) the perceptual Decision made by the patient (face/house); and (2) the Confidence rating given by the patient (high/low). To define Evidence for the choice probability analyses (see main text Methods), we also trained an additional estimator based on the stimulus presented on each trial.

Support vector machine estimators. Linear multivariate estimators were defined to predict a vector y of trial-by-trial categorical labels (e.g., face vs. house) or ordinal labels (e.g., high vs. low confidence) from a matrix of 'Features': trial-by-trial ECoG power in the high-gamma bands ([80 90 100 110 120 Hz], short-time multitaper method[3], 15 Hz half-bandwidth, 100 ms windows, 10 ms steps) at each sampling timepoint (X, with dimensions $n_{trials} \times (n_{electrodes} \times n_{frequencies} \times 1\ timepoint)$), i.e. 'electrode-frequency-timepoints'. These timepoints were defined on the peri-stimulus window from -250ms (before stimulus onset) to 1000ms after stimulus onset. We focused on high-gamma because this frequency range has demonstrated connection to perceptual processes[5–8], neural firing rates[9–16], and relationship to the BOLD signal in fMRI[17–20]. Importantly, this range was observed to be most salient in the stimulus-locked spectrograms (Supplementary Figure 1). Below we also show that including lower frequency ranges only diluted the decodability of Decision and Confidence behavioral factors, but did not change the qualitative pattern of results (see Supplementary Notes). Because of these considerations, and that SVM decoding can suffer from overfitting if the number of Features vastly outnumbers the number of trials for training and testing (as with all data fitting methods[21]), we focus on the most salient frequencies in order to minimize the possibility of overfitting by reducing the Feature space, while also minimizing computational demands.

Cross-validation and classifier performance. For time-series analyses, each estimator was fitted to each subject individually at each time sample. That is, for each of the behavioral factors described above, we fitted n_{time} estimators on an X matrix (n_{trials} × ($n_{electrodes}$ × $n_{frequencies}$ × 1 *timepoint*)) to predict a vector *y*. We used 5-fold stratified cross-validated L2-regularized L2-loss linear SVM classification (dual) implemented through the `liblinear` package[22] in `libsvm`[23]: each iteration generated predictions on 1/5th of the data (test set, X_{test}) after being fitted to the other 4/5th of trials (training set, X_{train}) while maximizing homogeneity between training and test sets via stratification. All SVM parameters were set to default values as provided in the `liblinear` package (e.g., soft margin parameter C = 1); continuous decision value outputs (coding the strength of categorization results) were taken as trial-by-trial estimates of $\hat{y}$. The performance of the classifier was assessed via comparison of y_{test} to $\hat{y}_{test}$ via an empirical Receiver Operating Characteristic (ROC) analysis[24] taken across all test trials at each fold, which plots the percent of hits (true positives) against false alarms (false positives) as a function of varying criterion values to classify the data into two categories. (The farther from the decision boundary a given decision value $\hat{y}$ is, the more likely it ought to be to reflect a correct classification[21].) When averaged across folds, this analysis resulted in an Area Under the Curve (AUC) score for each estimator, ranging from 0 to 1 with chance classification performance being defined at 0.5. We refer to AUC as 'decoding accuracy' or 'classification accuracy' in the main narrative for ease of interpretation. AUC was defined as being above chance for each subject when empirical p-values for the average AUC in 50ms bins, found via empirical permutation tests (100 permutations at each fold), were below p = 0.05. To reveal global trends as a function of time in visualization (Figure 2, main text), AUC scores were smoothed using a 5-point moving average (window size 50ms). All analyses were completed through custom-written software in MATLAB R2013a (MathWorks; Natuck, MA) and SPSS Statistics 22.0 (IBM).

To quantitatively examine the decoding patterns for the two estimators, we binned the AUC for each estimator as a function of time into post-stimulus time bins of 50ms. We calculated the mean AUC for each estimator in each of these bins, and conducted permutation tests (100 permutations in each fold) to identify an empirical p-value, to evaluate when each estimator achieved above-chance performance across subjects.

Neuroanatomical localization of representations

We used a transformation of the modified F-scores[25,26] for estimators to define electrodes' contribution to decoding. Given training instances $\boldsymbol{x}_i$, $i = 1, \ldots, l$, the F-score of the *j*th Feature for estimator *e* is defined as:

$$F_e(j) \equiv \frac{\left(\bar{x}_j^{(+)} - \bar{x}_j\right)^2 + \left(\bar{x}_j^{(-)} - \bar{x}_j\right)^2}{\frac{1}{n_+ - 1}\sum_{i=1}^{n_+}\left(\bar{x}_{i,j}^{(+)} - \bar{x}_j^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{i=1}^{n_-}\left(\bar{x}_{i,j}^{(-)} - \bar{x}_j^{(-)}\right)^2} \tag{S1}$$

where n_+ and n_- are the number of positive and negative instances, respectively; $\bar{x}_j$, $\bar{x}_j^{(+)}$, $\bar{x}_j^{(-)}$ are the average of the *j*th Feature of the whole, positive-labeled, and negative-labeled data sets; and $\bar{x}_{i,j}^{(+)} / \bar{x}_{i,j}^{(-)}$ is the *j*th Feature of the *i*th positive/negative instance. Following previous work [25], the numerator denotes the inter-class variance, while the denominator is the sum of the variance within each class. A larger F-score indicates that the Feature is more discriminative[26]. Note that the modified F-score is calculated directly from the high-gamma power in each electrode-frequency-timepoint and is therefore independent of any assumptions made during SVM estimator training.

Because these modified F-scores are approximately exponentially distributed, to define an electrode's contribution to an estimator's predictive capacity, we first took the mean log modified F-score for each electrode-frequency Feature at each timepoint from 0 to 1000ms after stimulus onset across frequency bands. We then z-scored the results across frequency bands and timepoints to create a standardized *contribution index C* such that

$$C_e(E,t) = z\left(\frac{1}{|f*|}\sum_{f \varepsilon [80,120]} log\,(F_e(j,t))\right) \tag{S2}$$

where $C_e(E,t)$ is the contribution to estimator e of electrode E at time t, f indicates the frequency band in the 80-120 Hz range, $|f*|$ indicates the cardinality of the Feature bands to be averaged across (in our case, 80-120 Hz in bands of 10 Hz gives five frequency bands, so in all cases $|f*| = 5$), $F_e(j,t)$ refers to the modified F score (Equation S1) of electrode-frequency Features j at timepoint t (i.e., only electrode-frequencies at timepoint t are taken as contribution at that timepoint), and $z(\cdot)$ refers to the standard z-score transformation across all modified F-scores for all electrode-frequency-timepoint Features j for that estimator. By taking the standard z-score transformation, each Feature's contribution is normalized such that we examine the *relative* contribution of each Feature to an estimator's decodability, to the extent the estimator is actually decodable -- that is, regardless of the magnitude or significance of the decodability itself. We did this separately for Decision and Confidence, and plotted each electrode's average *C* value in each lobe as a function of time after stimulus onset (Figure 2b, main text) and binned in 200ms time bins to facilitate further statistics and interpretation. All electrodes' MNI coordinates are available as a supplementary dataset.

We assigned each electrode as belonging to one of the four neocortical lobes based on its MNI coordinates, and plotted the average contribution index *C* within each lobe as a function of time after stimulus onset (Figure 2c, main text). To quantitatively test for differences in the spatial representation and timecourse for each predictor, we ran a mixed design ANOVA with 'between' factor lobe (4: frontal, parietal, temporal, occipital) and 'within' factors predictor (2: Decision, Confidence) and coarser time bins than used in the significance testing (5: 0-200, 200-400, 400-600, 600-800, 800-1000ms) on electrodes' contribution index *C* values across all subjects. We opted to use coarser time bins to reveal global trends in neuroanatomical localization, as an ANOVA on the 50ms bins used in the significance testing would reveal uninterpretable, complex interactions.

Generative Bayesian models

Details of both the Bayesian ideal observer model[27] and the Decision-Congruent Evidence Bayesian heuristic observer model[28] have previously been described elsewhere. All simulations were completed through custom-written software in MATLAB R2013a (MathWorks; Natuck, MA).

Two-dimensional framework. In the one-dimensional forward model described in the main text methods, the decision value *x* by definition reflects a Balance of Evidence for the two stimulus alternatives: the more *x* indicates Evidence favoring 'face', the less it favors 'house' and vice versa. To independently evaluate the different contributions of Decision-Congruent and Decision-Incongruent evidence with respect to multiple possible computation rules for Confidence, it is necessary to rely on a two-dimensional representation of the decision space to decompose *x* into independent estimates of Face Evidence and House Evidence[27–29] (Supplementary Figure 2).

Bayesian ideal observer. The Bayesian ideal observer makes both Decisions and Confidence judgments according to the same computations. Following previous work[27], we assume each generating category *C* is dependent on the evidence strength *s* favoring one or the other stimulus category (face/house), and can be represented as a bivariate Gaussian distribution such that $C_{s,Face} \sim N([s,\ 0],\ \Sigma)$ for a 'Face' trial and $C_{s,House} \sim N([0,\ s],\ \Sigma)$ for a 'House' trial. In its most basic formulation, we define Σ = I, where I is the 2x2 identity matrix. (Note: because of the scale of Evidence as it is defined, we scale Σ so that the noise does not overwhelm the samples such that $\Sigma^* = 0.01\times\Sigma$).

Because the absolute evidence level is of course unknown to the observer, the face/house Decision is made by marginalizing over possible evidence levels to produce the posterior probability estimate of the trial being a Face or a House trial[30]. Thus the joint probability of each trial type and evidence level is estimated through Bayes' rule,

$$p(C,s|x) = \frac{p(x|C,s)p(C,s)}{p(x)} \tag{S3}$$

and then the secondary variable of evidence strength *s* is integrated out, leaving estimation of the posterior probability of each face/house category via the marginal distribution

$$p(C|x) = \int p(C,s|x)\ ds \tag{S4}$$

In the simplest form, both face/house categories have equal prior probability, and so the observer makes its Decision via

$$C_{chosen} = argmax_i \; p(C_i|x) \tag{S5}$$

Finally, because confidence is judged simply as the probability of being correct, we define

$$confidence \;=\; p(correct) \;=\; p(C_{chosen}|x) \tag{S6}$$

Decision-Congruent Evidence Bayesian heuristic observer. Rather than calculating Confidence in a Decision based on the Balance of Evidence both in favor of the selected choice and related to the unselected choice(s), the Decision-Congruent Evidence Bayesian heuristic model instead discards evidence favoring the unselected choice (Decision-Incongruent Evidence), and to compares the strength of evidence for the selected choice both to a prototype along the selected dimension and to the noise distribution:

$$confidence \;=\; p(\hat{C}_{chosen}|x) \;=\; \frac{p(x|\hat{C}_{chosen})p(\hat{C}_{chosen})}{p(x|\hat{C}_{chosen})p(\hat{C}_{chosen}) + p(x|N)p(N)} \tag{S7}$$

with N representing the noise, or nothingness distribution and $\hat{C}_{chosen}$ representing the prototype. This strategy corresponds to a Decision-Congruent Evidence rule, i.e. that the only relevant information in judging Confidence is the magnitude of evidence favoring the Decision that was made. Note that previous formulations of this model assume that the Decision-Incongruent Evidence is explicitly discarded and the internal evidence remapped, e.g. that if ‘Face’ is selected, the evidence to be evaluated is $x^* = [x(1), 0]$; however, this explicit discarding does not change the behavior of the model from the current formulation.

Model evaluation and comparison. Rather than generate the samples x as would be typical in a Monte Carlo simulation of a generative model, we directly evaluated the Face and House Evidence samples defined from the SVM feature weights as in the choice probability analyses (Methods: Choice probability analysis, Equations 1 & 2, main text). Each trial-by-trial sample, for each subject, is therefore a two-dimensional point, i.e. x = [Evidence$_{\mathrm{Face}}$, Evidence$_{\mathrm{House}}$]. Although the Evidence is defined according to the SVM estimator trained on the presented stimulus, subjects were not at 100% accuracy; this means there is some internal noise in the system that may be attributable to neuronal factors in addition to the decoding noise, $\alpha_{decoding}$, discussed in the main text methods. This noise is not directly estimable, because decoding accuracy did not reach behavioral accuracy, and so the decoding noise overwhelms any internal noise. Therefore, we consider another source of noise, such that $x^* = x + \varepsilon$, with $\varepsilon \sim \mathrm{N}(0,\sigma)$.

Each of the above-described models makes a decision about each noise-corrupted sample according to Equations S3-S5, and then rates confidence according to its own rule (Equation S6 or S7). Because we cannot explicitly estimate internal neural noise, for each model we iteratively examined a range of noise levels, varying from σ of 0 to 0.01 in steps of .001 (note the scale of the Evidence is very small, such that maximal Evidence for most subjects is on the order of 0.1). Because the number of trials completed by each subject is small, we ran each simulation 1000 times with different random seeds on each run. This produced a total of 22,000 sets of trial-by-trial Confidence predictions (2 models x 11 noise levels x 1000 runs) for each subject.

We used standard ROC choice probability (CP) analysis[24] as above to quantitatively evaluate the models' predictions by calculating their choice probabilities: we compared the predicted Confidence ratings for each subject, for each of the 22,000 simulation sets, to the empirical Confidence ratings produced by the subject on a trial-by-trial basis by computing the area under the curves (AUC) as a measure of each model's performance, i.e. its CP.

To evaluate whether the Decision-Congruent-Only rule for computing Confidence is a better predictor of subjects' Confidence behavior than the Balance-Of-Evidence rule, we conducted a 2 (Confidence model: Balance vs. Decision-Congruent) x 11 (noise level) repeated-measures ANOVA on the mean CP for each model at each noise level for each subject. We also calculated the average difference between $CP_{\text{Decision-Congruent}}$ and CP_{Balance} across subjects at each noise level; if the two rules for computing Confidence are truly equivalent, this difference should not deviate from 0 at any noise level. Finally, we calculated the exceedance probability of one model being more likely than the other model across all 33,000 simulation runs and all six subjects. The exceedance probability is particularly intuitive as all exceedance probabilities sum to one over any number of tested models.

Supplementary Results

Behavioral results

Additional behavioral results are presented in Supplementary Figure 3.

Decoding results

Numbers supporting the graphical presentation in Figure 2 (main text) are presented in Supplementary Table 3.

Motor preparation

It is important to note that decodability for all factors reached above-chance levels well before any movement onset (mean RT = 1136ms). Because finger movement preparation can typically be decoded only up to 200ms before movement onset with ECoG[31], and because responses via four unique button-press combinations were executed with one hand, this result rules out the possibility that decoding is based on movement preparation or execution and not internal representations of stimulus properties or decisional calculations (see also next section).

However, to comprehensively rule out the possibility of motor preparation driving our decoding results, we performed a response-locked analysis. We lined up trials by reaction time rather than stimulus onset and calculated the pairwise Contribution Index (see above, Equation S2) for each button, i.e. each finger, compared to the others. We then examined the degree to which electrodes in or adjacent to primary motor cortex (based on MNI coordinates) for each subject would contribute to decoding this actual motor movement as opposed to the Decisions themselves in the 500ms window leading up to the motor movement itself, marked by the reaction time on each trial. We compared this to how these electrodes behaved in the stimulus-locked analysis (main text Results).

This analysis revealed that electrodes near the motor cortex did indeed carry information about which *finger* was used to press a button when the information was lined up to the motor movement, as shown by the increasing trend of the Contribution Index as the reaction time approaches (Supplementary Figure 4b). However, these electrodes did *not* carry information about the Decision when the trials were locked to motor movements rather than stimulus onset (Supplementary Figure 4b): the Contribution Index for these electrodes was flat across the examined time period leading up to motor movement (button press), suggesting that Decision information is not available in motor planning or execution areas when the information accumulation stage is 'scrambled' due to changing the trial-by-trial alignment from stimulus-locked to response-locked. This dissociation is in stark contrast to the practically null and certainly not increasing contribution of these units to decoding either Decisions or motor movements when the analysis was stimulus-locked (as in the main results) (Supplementary Figure 4a). This result demonstrates that motor preparation and execution does not contribute to decoding of the Decision. We also checked to ensure that high gamma power did not differ between high or low confidence trials when trials were response-locked (Supplementary Figure 5).

Neuroanatomical localization of representations

An axial view and subject-by-subject plots of the Contribution Index (Figure 2b, main text) are presented in Supplementary Figures 6 and 7.

The mixed design ANOVA with 'between' factor lobe (4: frontal, parietal, temporal, occipital) and 'within' factors predictor (2: Decision, Confidence) and coarser time bins than used in the significance testing (5: 0-200, 200-400, 400-600, 600-800, 800-1000ms) revealed main effects for predictor ($F(2,1870) = 1.758$, $p < .001$), time bin ($F(4,3480) = 44.723$, $p < .001$), and lobe ($F(1,870) = 4.668$, $p = .003$), as well as an interaction between the two 'within' factors of predictor and time bin ($F(8,3480) = 6.156$, $p < .001$). Crucially, the 'between' factor of lobe interacted with all 'within' factors: predictor x lobe $F(6,870) = 6.875$, $p < .001$; time bin x lobe $F(12,3480) = 6.972$, $p < .001$; predictor x time bin x lobe $F(12,3480) = 2.294$, $p = .007$.

To explore these interactions, we conducted a series of step-down ANOVAs: two step-down 4 (lobe) x 5 (time bin) mixed design ANOVAs, one for each predictor; and five step-down 4 (lobe) x 2 (predictor) mixed design ANOVAS, one for each time bin (Supplementary Tables 4 & 5).

Representational overlap

To ensure that the observed dissimilarity in representations was not due simply to noise, we directly examined the neuranatomical overlap of the Decision and Confidence representations. We ranked the electrode-frequency-timepoint Features according to their informativeness in decoding each factor independently (Equation S1)[25,26], and selected a subset (the top 25%) of these Features for closer inspection. If Confidence shares the Decision representation, we should expect a large proportion of these most informative Features to be the same between Decision and Confidence even in the presence of significant noise.

32.41% of the top 25% of Features were shared were shared between representations of Decision and Confidence. We assessed the anatomical locus of the overlapping Features by calculating the percentage of overlapping Features in each neocortical lobe at each poststimulus timepoint (note: there are five Features per electrode at each timepoint, corresponding to power in 80, 90, 100, 110, and 120 Hz frequencies).

This analysis revealed a strong overlap in occipital electrodes, especially at 200-500ms post stimulus onset (Supplementary Figure 8). Overlap was second strongest in temporal regions in the same time period, and rose in parietal regions towards the end of the post-stimulus response window. Overlap was smallest in frontal regions across the whole post-stimulus time period.

It is worth noting that although nonzero modified F-scores for a given channel or frequency might indicate an area that is doing something other than the task-relevant processing, we might hope that such task-irrelevant processes would be at least somewhat common across both

Decisions and Confidence judgments -- e.g., pressing keys, remaining focused on the screen, being bored with the task, mind wandering, eye blinking, heartbeat, etc. -- and so if anything would artificially *inflate* the number of shared features between Decisions and Confidence judgments rather than deflate their feature overlap.

As an additional test of representational overlap, we also examined whether electrodes that demonstrate stimulus selectivity (i.e., face- or house-tuned electrodes) might also carry meaningful information about confidence in a tuning-specific manner. For this analysis we defined electrodes as significantly selective to 'face' or 'house' Decisions via one-tailed t-tests of the average high-gamma power in the window between 0ms and 500ms after stimulus onset, based on the observation that the most salient response happens before 500ms post-onset (Supplementary Figure 1). If this average power was significantly greater (at $p < .05$) for trials in which the participant made a 'face' Decision over when the participant made a 'house' Decision, the electrode was designated as 'face-selective'; we used the same approach to define 'house-selective' electrodes.

For electrodes that demonstrated significant selectivity for face or house Decisions, we next examined the average response in trials where the participant responded 'high confidence' versus 'low confidence' as a function of whether the participant's Decision on that trial was congruent with the electrode's tuning. For example, we examined whether for "face-selective" electrodes, average high-gamma power was more different in 'high confidence' versus 'low confidence' trials when the participant responded 'face' than when the participant responded 'house'.

This analysis revealed that despite the largely separable representations for Decisions and Confidence (main text Figure 2), there are also individual electrodes that carry some information about Confidence even though they were also identified as significantly responsive to participants' Decisions. Importantly, as we hypothesized these electrodes' activity reflected Confidence in a manner consistent with the Decision-Congruent Evidence rule: for example, when a 'face-selective' electrode carries information about Confidence, it appears to be primarily on trials when the participant made a 'face' Decision in a manner congruent with the electrode's tuning (Supplementary Figure 9). In contrast, when the subject makes a Decision that is incongruent with the electrode's selectivity (e.g., 'house'), there is a much smaller difference between the high-gamma power in high versus low confidence trials.

To confirm this visual interpretation, we averaged the difference in activity for high versus low confidence trials as a function of choice congruency, and compared their means across the entire time window with a two-sample t-test. This revealed that, confirming visual inspection, the average difference between high versus low confidence activity was significantly higher for trials in which the participant made a Decision congruent with the electrode's tuning preference ($t(114) = 4.133$, $p < .001$), lending further credence to the Decision-Congruent Evidence rule for computing confidence.

Choice probability analysis

Subject-by-subject plots of the choice probability ROC curves are presented in Supplementary Figure 10. Planned *post-hoc* t-tests for Choice Probabilities revealed that all surpassed chance level (Supplementary Table 6).

To ensure that results from this analysis are not overly dependent on our particular definition of Evidence, we performed a confirmatory analysis. In this confirmation, we defined electrodes as being ‘face’ or ‘house’ selective via a set of Bonferroni-corrected t-tests comparing high-gamma power at each post-stimulus timepoint against the average high-gamma power during the baseline (pre-onset) period. This allowed us to identify post-stimulus timepoints at which ‘face’ or ‘house’ vs. baseline high-gamma power was significant, which we interpreted as a set of electrode-frequency-timepoints that were significantly responsive to faces versus houses independently of one another. We then removed all electrode-frequency-timepoints that were common to both faces and houses (which additionally addresses concerns about task-irrelevant processing). Finally, we used the remaining electrode-frequency-timepoints to define Evidence similarly to how it is defined in the main text, but with an important modification that avoids reliance on SVM feature weights.

This new definition of Evidence thus modifies Equation 2 in the main text, which reads

$$f_s(n,t,i) = \left|w_i\right| \cdot g_{n,t,i} \cdot I_i \qquad \text{(main text 2)}$$

to read

$$f_s(n,t,i) = g_{n,t,i} - \frac{1}{N}\sum_{j=1}^{N} g_{n,t,i} \qquad \text{(S8)}$$

By this simpler definition, which does not depend on the sign of the SVM feature weights or the sign of the signal for each feature, we obtained the same results as the main analysis: the Balance-of-Evidence rule predicted Decisions much better than the Decision-Congruent-Only rule (t(5) = 12.31, p < .001) but for Confidence the Choice Probabilities were no different between the two computational rules (t(5) = -1.60, p = .17) (Supplementary Figure 11).

Lag in predicting from Decision to Confidence?

To address the stability of the representations of the Stimulus, Decision, and Confidence, we performed temporal generalization analyses[32] both within each estimator and cross-predicting from Stimulus and Decision to Confidence. This analysis helps explore whether the failure in cross-predicting was due to temporal lag between calculations of Type 1 (Decision) and Type 2 (Confidence) judgments.

To quantify the effect of potential temporal lag, the Decision estimator fitted at each timepoint *t* as above was tested on its ability to decode Confidence at all other timepoints *t’* for each trial after taking the absolute value of the Decision SVM estimator decision values, $\hat{y}$, as above. This type of analysis results in a square generalization-across-time (GAT) matrix, where the y-axis is the time at which the estimator was fitted, and the x-axis is the time at which the estimator was evaluated[32,33]. Temporal generalization analyses were performed on the same time peri-stimulus window as the previous analyses: -250ms to 1000ms. Note that when cross-predicting from either Stimulus or Decision to Confidence, the absolute value of the SVM decision values $\hat{y}$ was taken according to standard signal detection theoretic approaches -- see also the Signal Detection Theoretic Forward Model section in the main text Methods.

All of these temporal generalization analyses suggested that while Decision shares many features with Stimulus -- mostly localized to occipital areas, as presented in the present version of the manuscript -- the overlap from these to the Confidence estimator is minimal even when one applies the decoder to Confidence at a different time than when a Stimulus or Decision estimator was trained (Supplementary Figure 12). This suggests that the lack of feature overlap we observed isn't trivially due to a delay between Type 1 (face/house) and Type 2 (confidence) judgments. In other words, the representation for Decision is separable from that of Confidence not only when the Decision estimator is applied to decode Confidence at its training timepoint; in fact, these representations are separable in space and time, as it is not the case that the representation of Decision can predict Confidence at *any* point in the post-stimulus window[32].

Generative Bayesian models

The Decision-Congruent Evidence Bayesian heuristic observer significantly outperformed the Bayesian ideal observer (Balance rule) in predicting subjects' Confidence judgments. The Decision-Congruent rule had an exceedance probability of 72.8%, meaning that in 72.8% of the simulations it outperformed the Balance rule. The degree to which the Decision-Congruent rule outperformed the Balance rule interacted with the amount of noise in the simulation, such that with little noise their performance was about equivalent, but with increasing noise the difference became more apparent (2 (Confidence rule) x 11 (noise level) repeated measures ANOVA, results in Supplementary Table 7).

When plotting the difference in CP between the rules across all subjects noise levels, it is clear that the Decision-Congruent rule demonstrates higher choice probability, especially at intermediate noise levels, driving the interaction (Supplementary Figure 13).

Frequencies outside 80-120 Hz

Many studies agree that gamma and high-gamma frequency ranges (~30-190 Hz) together represent a substrate supporting fast, task-relevant processing, including complex cognition, feature selection, attention, memory, and binding of features or distributed responses, among many other higher level functions[15,34–38]. Importantly, however, although precursors of attention and feature binding[39] and object representation[40,41] have been linked to gamma band activity (30-80 Hz), it is power in the high-gamma frequency bands (80-120 Hz) that has been linked to perception itself[5–8] and that appears directly linked to neural firing rates[9–16].

However, broadband activity in the entire gamma-high-gamma frequency range (~30-190 Hz) has been linked to the BOLD signal in functional neuroimaging (fMRI)[17–20]. It has also been suggested that the gamma range has (30-80 Hz) separate neural origins from high-gamma[15], and may provide communication channels between cortical areas[42,43]. Because the relationship between high-gamma power and underlying neural activity is known to be complex and depend on many factors (including anatomical area), to partially alleviate this concern we also re-ran all decoding analyses presented in the main text utilizing a broader frequency range from 30-190 Hz. To provide an easily-digestible summary comparison of high-gamma versus all gamma-high-gamma, we created an index of predictive power *P* of estimator *e* as

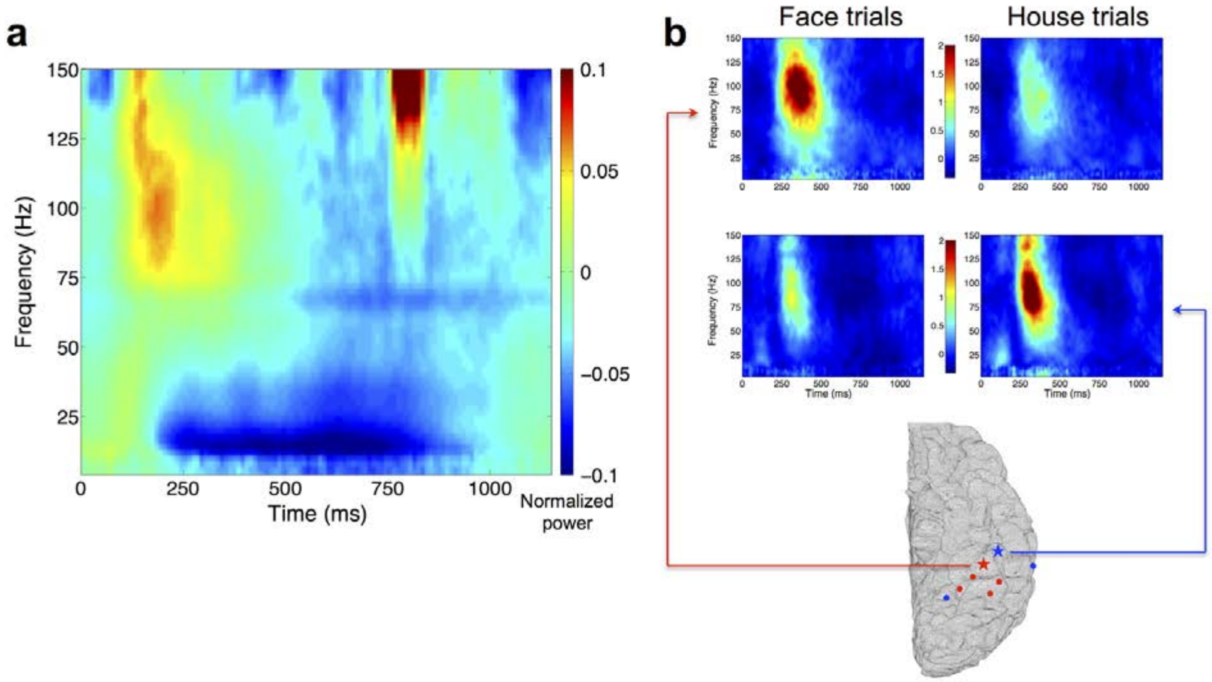
$$P_e = \frac{\sum_{t=0ms}^{t=1000ms} AUC_e(t)}{\sum_{t=0ms}^{t=1000ms} AUC_{chance}} \tag{S9}$$

where $AUC_e(t)$ is the AUC for estimator *e* at timepoint *t* after stimulus onset calculated according to the ROC analysis described above, and $AUC_{chance} = 0.5$. Estimator performance when using all frequency bands from 30-190 Hz ([30 35 40 45 50 55 60 70 80 90 100 110 120 130 140 150 160 170 180 190] Hz) qualitatively followed the same trend as when we focused on high gamma, albeit with slightly poorer decoding performance overall than when focusing on 80-120 Hz (Supplementary Table 8).

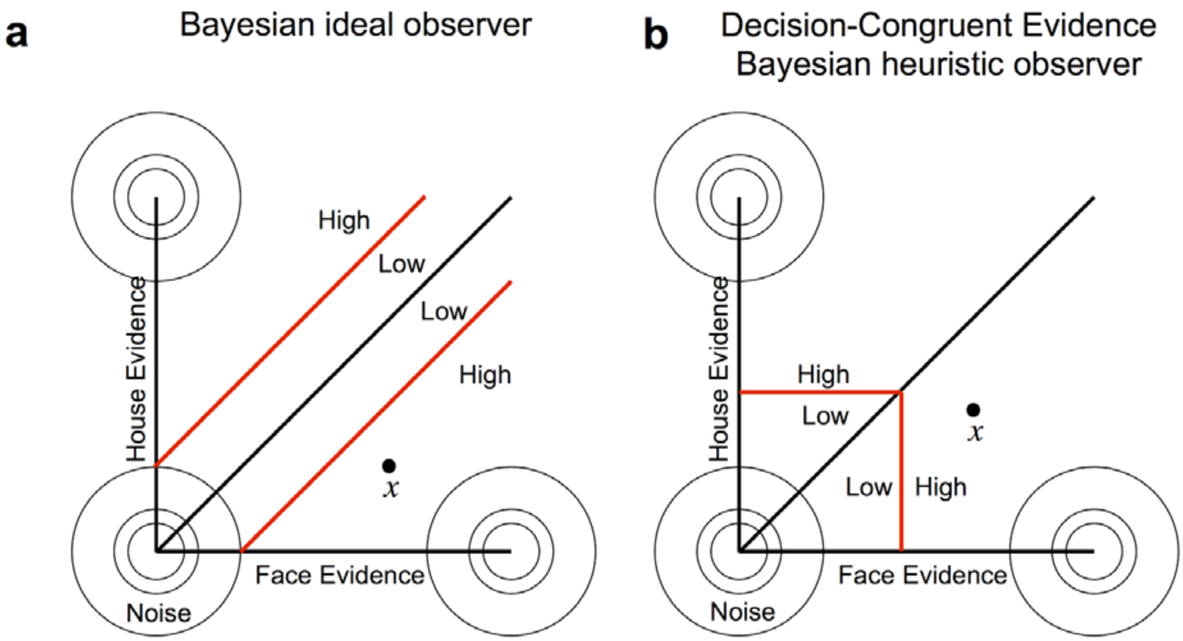
All analyses presented in the main text follow this pattern when conducted on the full 30-190 Hz spectrum: qualitatively similar but quantitatively weaker results than when focusing on 80-120 Hz. This finding also confirms previous reports of the relevance of high-gamma power to the computations and representations underlying perception and perceptual processes.

We also considered including lower frequency phase values as Features in the decoding analysis, as alpha phase has been linked to perception in concert with high-gamma and has been suggested to coordinate high-gamma activity (e.g.,[44,45]). However, unfortunately because low-frequency phase analyses were not planned ahead of time, we did not use an LED to precisely monitor the timing of the actual stimulus onset. Instead, visual stimulus "onset" was monitored from the computer output, meaning a delay between the stimulus presentation trigger and actual stimulus presentation was possible. We expect that this was on the order of milliseconds and relatively constant across trials, and therefore should not affect the interpretation of our findings regarding contributions of high-gamma power in 200ms time bins or smoothed with a 50ms time window, as we present here. However, because phase information cannot be temporally smoothed, the lack of precise stimulus onset monitoring via LED could pose a problem to phase analyses. We therefore elected to conduct this kind of analysis in future studies.

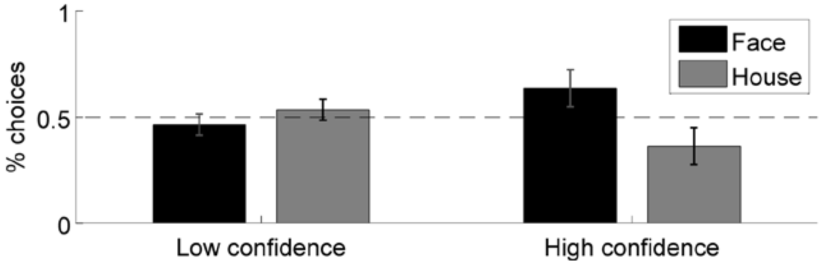
Supplementary Figures



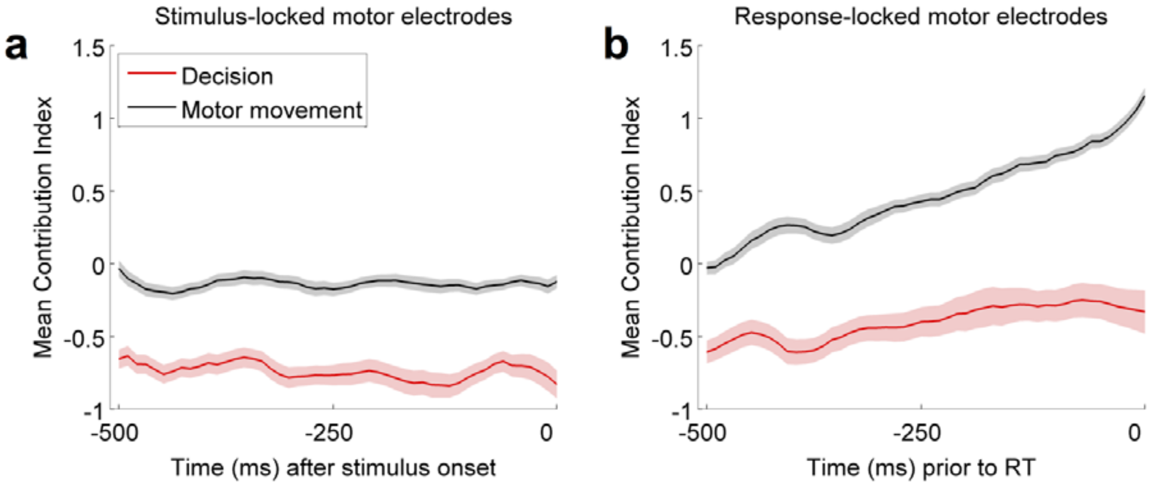
Supplementary Figure 1. Time-frequency spectrograms of electrocorticography data, time-locked to stimulus onset. (a) Normalized average spectrogram across trials and electrodes. (b) Representative electrode spectrograms in response to 'face' versus 'house' stimulus presentations. Power is most salient in high-gamma frequency bands (80-120 Hz) around 250-400ms after stimulus onset. Note that face- and house-selective electrodes may not match known functional or anatomical maps. This depiction is for illustrative purposes only; see main text for more comprehensive discussion of electrode contribution to Decision versus Confidence decoding.



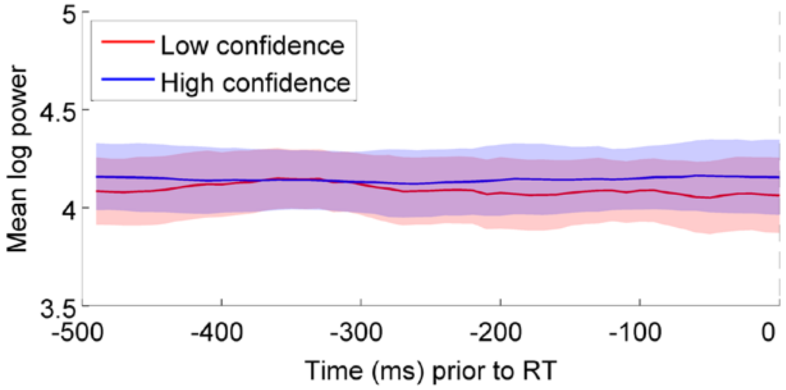
Supplementary Figure 2. Two-dimensional signal detection theoretic and Bayesian ideal observer model space. The abscissa represents Face Evidence and the ordinate represents House Evidence. Concentric circles represent isometric bivariate Gaussian distributions viewed from above. A diagonal decision criterion (black line) divides the space, such that an internal decision variable x for which Face Evidence is higher than House Evidence (below the diagonal) will be categorized as a 'face' by the observer. Confidence criteria are shown in red. Note that criteria are shown for the signal detection theoretic formulations of the models, but represent the shape of the decision and confidence contours for the Bayesian observer models as well. (a) According to normative models, Confidence should be rated on the same internal information as the Decision. The farther x is from the Decision criterion, the more likely the observer is to have made a correct Decision, and so the more confident the observer should be. This leads to diagonal Confidence criteria, or Confidence being rated on the Balance of Evidence. (b) In contrast, the Decision-Congruent Evidence Bayesian heuristic model judges Confidence according to the magnitude of Decision-Congruent Evidence, such that the relevant comparison is now whether the evidence on the current trial better reflects noise or some prototypical example of the chosen stimulus category. See text for details.



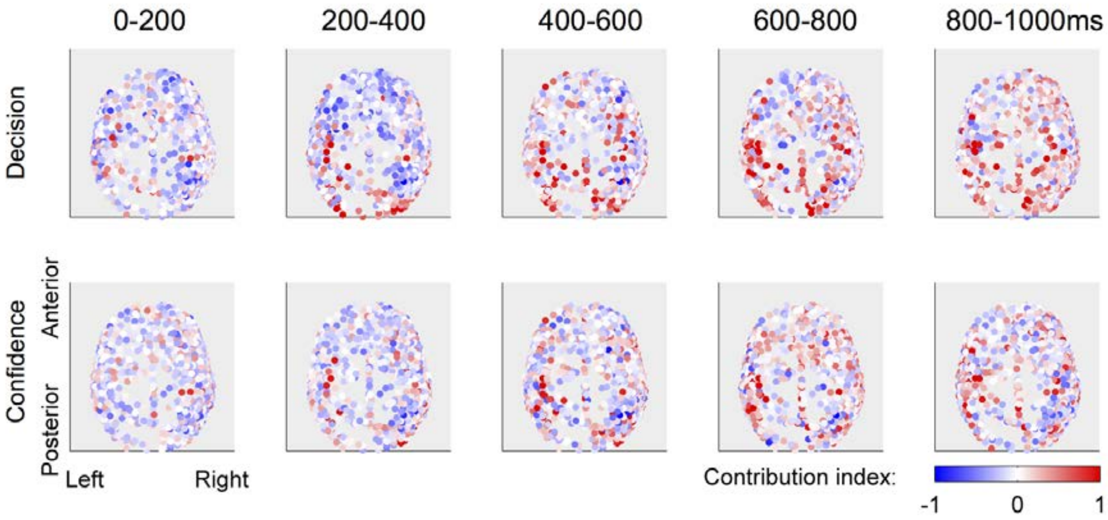
Supplementary Figure 3. Supplementary behavioral results: Face and House response percentages as a function of high versus low confidence. A 2x2 repeated measures ANOVA revealed no main effect of Confidence (F(1,5) < .001, p = 1), no main effect of Face vs. House Decision (F(1,5) = 2.528, p = .173), and no interaction between Confidence and Face vs. House Decision (F(1,5) = 1.885, p = .228).



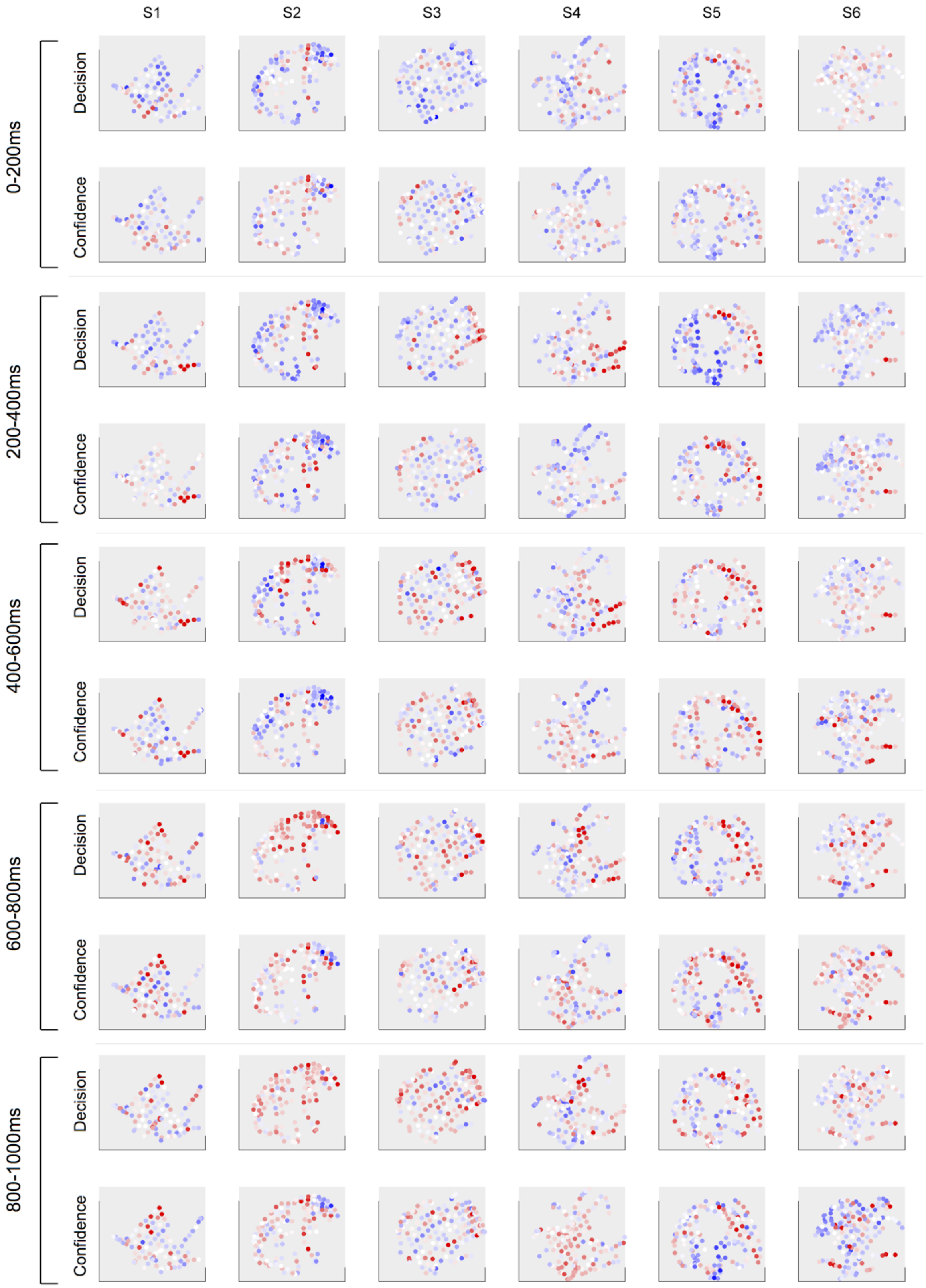
Supplementary Figure 4. Motor preparation cannot explain decoding results. Electrodes near motor cortex carry information about motor preparation only in response-locked analyses (b), not in stimulus-locked analyses (a), and only in time periods directly leading up to motor execution. This result shows that motor preparation or execution should not drive the decoding results presented in the main text.



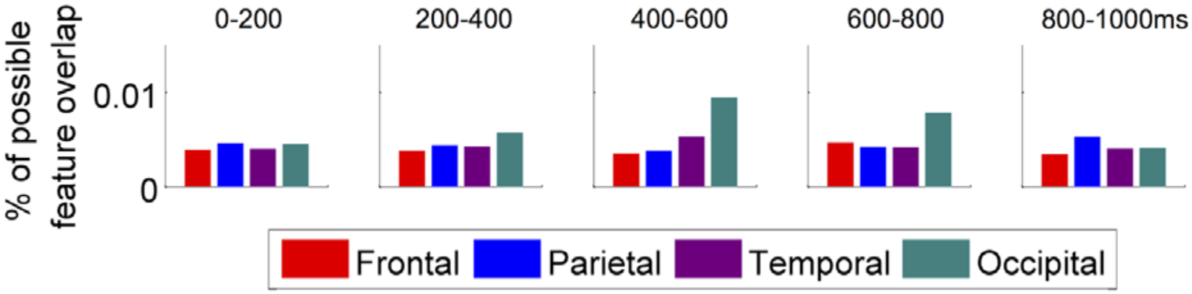
Supplementary Figure 5. Mean high gamma power is not different for high versus low confidence responses when response-locked.



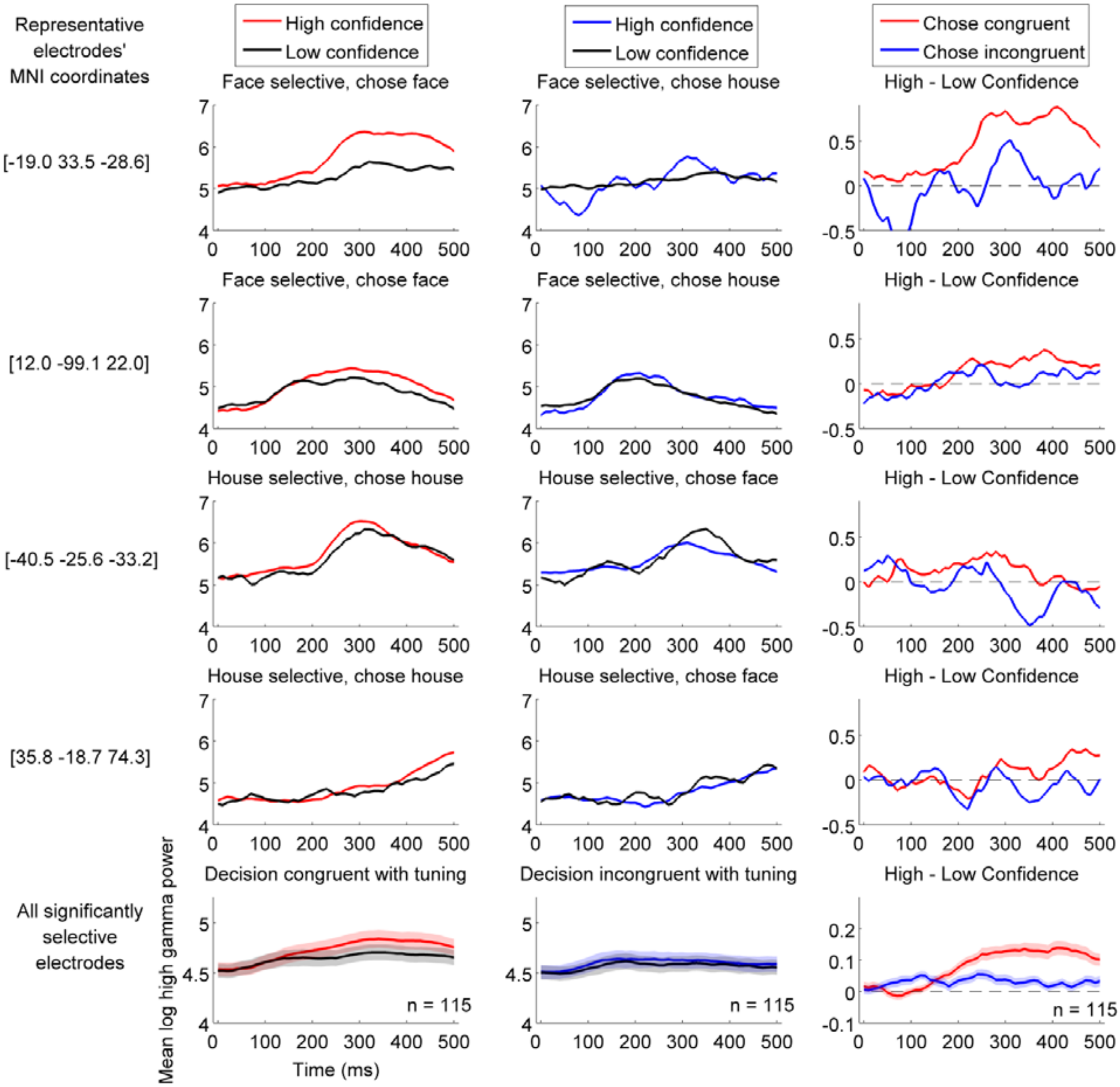
Supplementary Figure 6. Axial view of Figure 2b (main text).



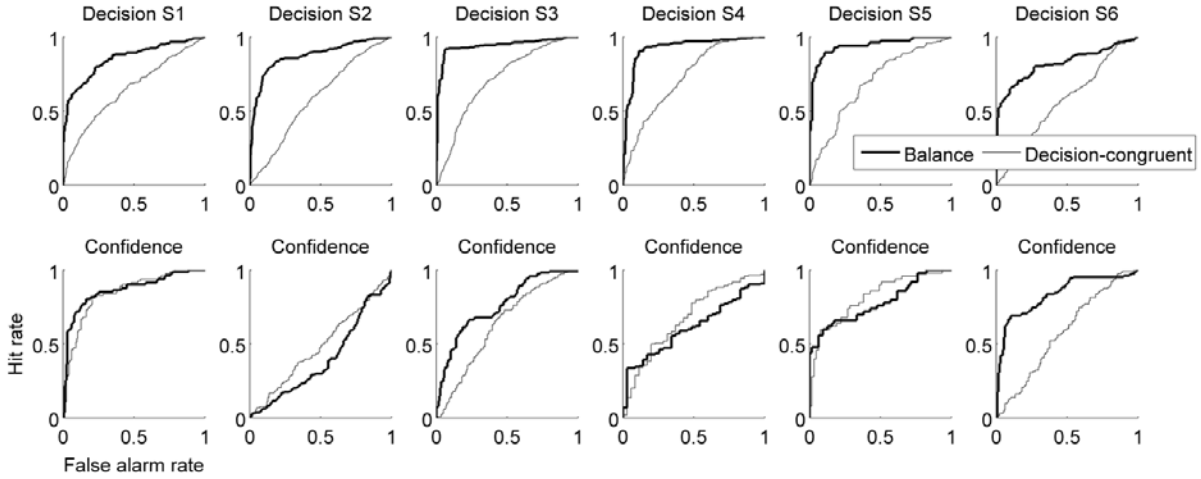
Supplementary Figure 7. Subject-by-subject plots of contribution index *C*, as in Figure 2b (main text).



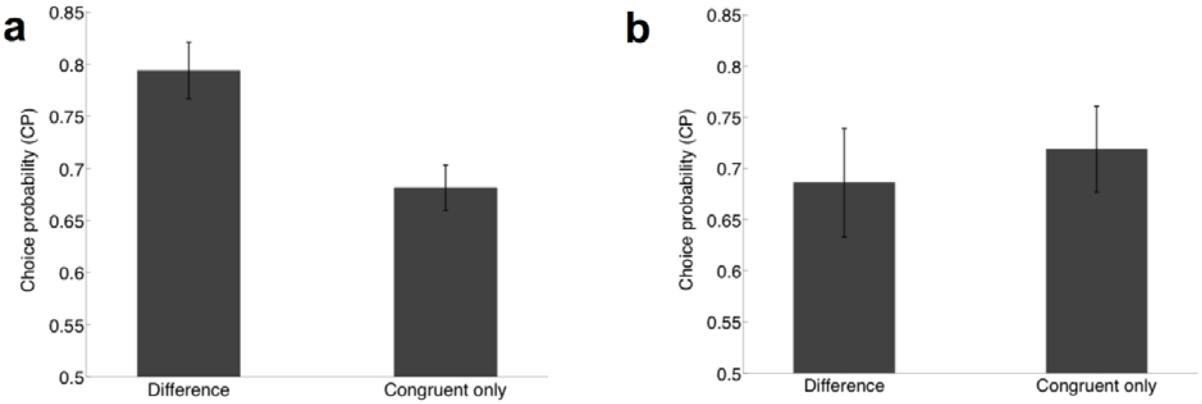
Supplementary Figure 8. Anatomical distribution of top 25% of overlapping Features between Decision and Confidence representations in the same post-stimulus time bins. Most of the overlap is in occipital regions. Because Confidence is more distributed than Decision, it is clear that Confidence calculations rely both on these occipital Features and on other sources of information (e.g., frontal regions; Figure 2c, main text).



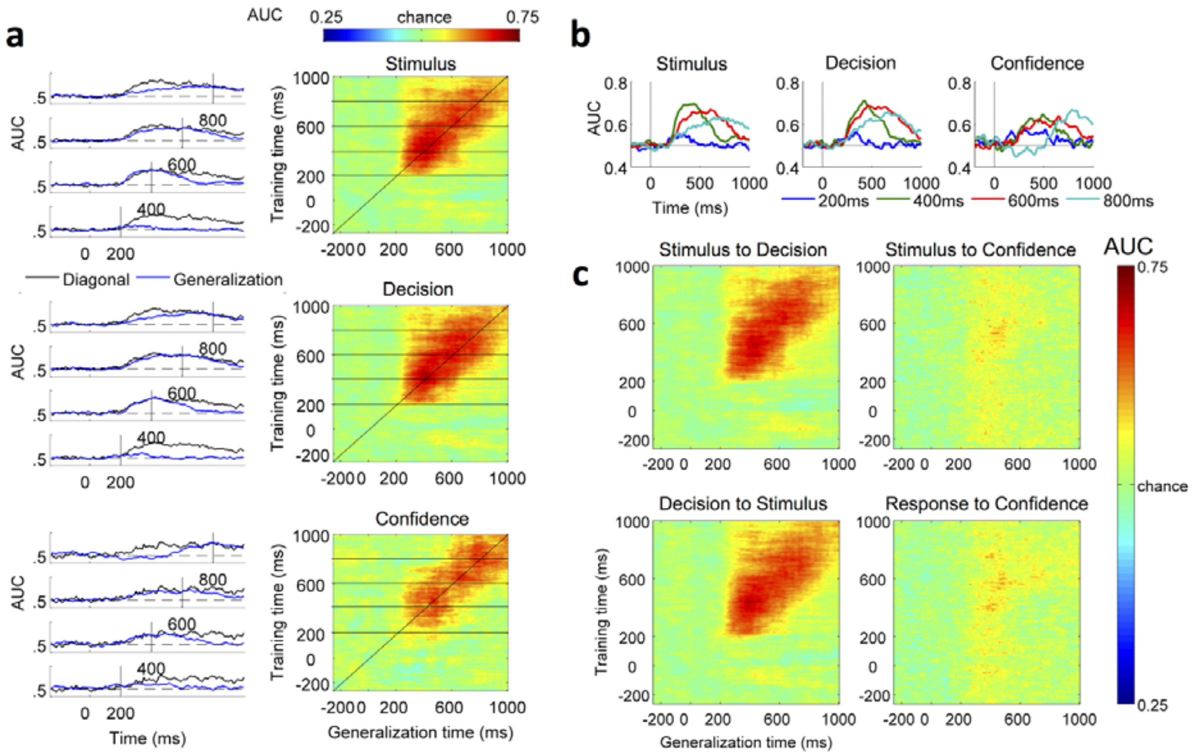
Supplementary Figure 9. Electrodes significantly responsive to participants' Decisions may also carry some information about Decision-Congruent Evidence for confidence. Rows 1-4 show individual representative 'face-selective' and 'house-selective' electrodes, and the bottom row shows the average response for all electrodes demonstrating significant selectivity according to participants' Decisions. In electrodes significantly tuned to Decisions, mean log high-gamma power between high versus low confidence trials differs only when the participant's Decision is congruent with the electrode's tuning preference. For example, 'face-selective' electrodes show differences in high-gamma power primarily when the participant makes a 'face' Decision, and not nearly as much when the participant makes a 'house' decision. The opposite is true for 'house-selective' electrodes. This response pattern is consistent with the Decision-Congruent Evidence rule for computing perceptual confidence. Shaded regions (bottom row only) indicate the standard error of the mean.



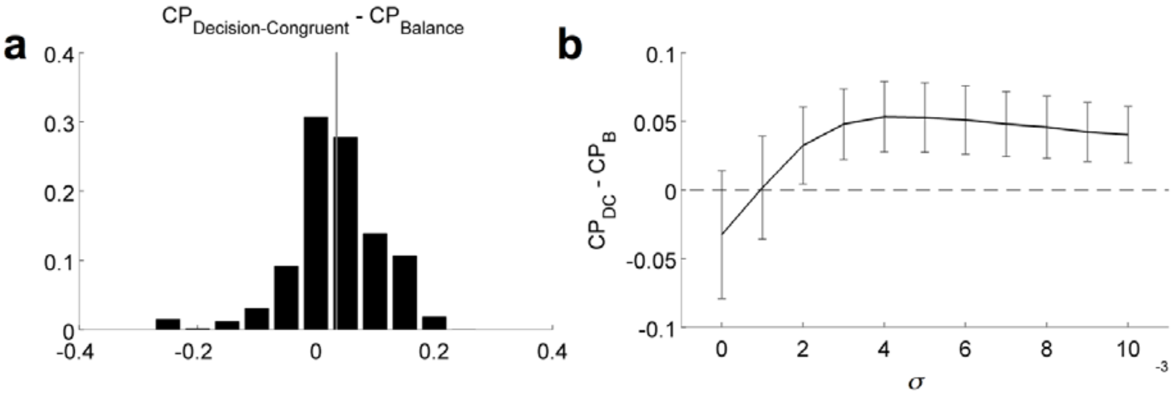
Supplementary Figure 10. Subject-by-subject plots showing choice probability ROC analyses. choice probability (CP) for Balance is always much higher than for the Decision-Congruent Evidence rule for Decision, but the two are statistically indistinguishable for Confidence.



Supplementary Figure 11. Results of the choice probability confirmatory analysis demonstrate that choice probability results are robust to the particularities of the definition of Evidence presented in the main text.



Supplementary Figure 12. Results of temporal generalization analysis to examine potential lag in predicting from Decision to Confidence. The Stimulus and Decision estimators trained at any timepoint poorly predicted Confidence at all timepoints, indicating that the poor performance of the forward model is not due to temporal dissociations in processing, but must be due to actual differences in representations and computations.



Supplementary Figure 13. The Decision-Congruent Evidence Bayesian heuristic observer predicted subjects' Confidence responses better than the Bayesian ideal observer across noise levels. (a) Histogram of difference in CP for each rule across all noise levels, Decision-Congruent-Only minus Balance-Of-Evidence. The Decision-congruent model on average outperforms the Balance model. (b) The difference in CPs changes as a function of the internal (neuronal) noise assumed in the observer. Across noise levels except for no noise, CP for the Decision-Congruent model matches or exceeds CP for the Balance model. Error bars indicate the standard error of the mean across subjects.

Supplementary Tables

Patient	Hand	VCI	POI	WMI	PSI	Language
1	R	82	97	91	88	English
2	R	-	-	-	-	Spanish
3	R	134	84	102	86	English
4	R	-	-	-	-	Farsi
5	R	95	109	108	124	English
6	R	-	-	-	-	Mandarin

Supplementary Table 1. Patients' handedness, spoken language, and WAIS results. Prior to completing the experimental task, subjects with English as a first language were evaluated on the Weschler Adult Intelligence Scale (WAIS): Verbal Comprehension Index (VCI), Perceptual Organization Index (POI) Working Memory Index (WMI), and Processing Speed Index (PSI).

		Frontal		Parietal		Temporal		Occipital	
Patient	**Hemisphere**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**
1	Left	26	28.6%	17	18.7%	40	44.0%	8	8.8%
2	Right	75	54.0%	34	24.5%	25	18.0%	5	3.6%
3	Right	33	22.3%	49	33.1%	49	33.1%	17	11.5%
4	Bilateral	48	32.2%	36	24.2%	63	42.3%	2	1.3%
5	Bilateral	67	39.0%	37	21.5%	59	34.3%	9	5.2%
6	Bilateral	70	40.0%	41	23.4%	54	30.9%	10	5.7%
Average		**53.2**	**36.0%**	**35.7**	**24.2%**	**48.3**	**33.6%**	**8.5**	**6.0%**

Supplementary Table 2. Electrode distribution by subject and cortical lobe.

	Decision		Confidence	
Time window beginning at (ms)	AUC	% significant subjects	AUC	% significant subjects
0	0.4945	0%	0.4862	0%
50	0.5013	0%	0.5168	0%
100	0.5101	0%	0.5087	0%
150	0.5238	0%	0.5257	0%
200	0.5536	16.67%	0.5327	0%
250	0.6176	50%	0.5595	16.67%
300	0.6490	83.33%	0.6042	16.67%
350	0.6757	66.67%	0.6216	16.67%
400	0.6999	100%	0.6144	16.67%
450	0.6789	66.67%	0.6612	50%
500	0.6775	83.33%	0.6239	33.33%
550	0.6700	83.33%	0.6268	33.33%
600	0.6797	100%	0.6361	16.67%
650	0.6721	100%	0.6654	33.33%
700	0.6726	83.33%	0.6244	33.33%
750	0.6401	50%	0.6574	33.33%
800	0.6157	50%	0.6452	33.33%
850	0.6053	16.67%	0.6138	33.33%
900	0.5885	33.33%	0.5999	0%
950	0.5963	0%	0.6182	0%

Supplementary Table 3. Results of SVM decoding in 50ms timebins as average AUC in each bin and percentage of subjects reaching significance in each bin. This analysis confirms that Decision reached significant decodability earlier, and for more subjects, than Confidence.

	Factor	**F(df)**	**p**
Decision	**Time bin**	F(4,3480) = 41.272	< .001*
	Time bin x Lobe	F(4,3480) = 7.843	< .001*
	Lobe	F(3,870) = 7.748	< .001*
Confidence	**Time bin**	F(4,3480) = 18.728	< .001*
	Time bin x Lobe	F(4,3480) = 2.475	.003*
	Lobe	F(3,870) = 1.896	.129

Supplementary Table 4. Results of step-down mixed design ANOVAs within each predictor to test for neuroanatomical distribution of Decision versus Confidence representations. All effects are significant, with the exception of the main effect of lobe for Confidence; this indicates that the neuroanatomical localization of the representation for Decision is specific to certain neocortical lobes, but the distribution of Confidence is highly distributed.

	Factor	**F(df)**	**p**
0-200ms	**Predictor**	F(1,870) = .556	.456
	Predictor x Lobe	F(3,870) = 1.528	.206
	Lobe	F(3,870) = .419	.740
200-400ms	**Predictor**	F(1,870) = 3.358	.067
	Predictor x Lobe	F(3,870) = 8.864	< .001*
	Lobe	F(3,870) = 17.010	< .001*
400-600ms	**Predictor**	F(1,870) = 15.826	< .001*
	Predictor x Lobe	F(3,870) = 1.682	.169
	Lobe	F(3,870) = 4.640	.003*
600-800ms	**Predictor**	F(1,870) = 2.871	.091
	Predictor x Lobe	F(3,870) = 5.165	.002*
	Lobe	F(3,870) = 2.000	.113
800-1000ms	**Predictor**	F(1,870) = 21.880	< .001*
	Predictor x Lobe	F(3,870) = 2.389	.067
	Lobe	F(3,870) = 3.150	.024

Supplementary Table 5. Results of step-down mixed design ANOVAs within each post-stimulus time bin, again to check for differences between predictors. No differences are apparent in the 0-200ms time bin, consistent with decoding results. However, beginning in the 200-400ms bin, a significant main effect of lobe and an interaction between lobe and predictor appear, with a trending main effect of predictor. This pattern is relatively consistent through the remaining three time bins.

		Mean (± SD)	t(5)	p
Decision	**Balance**	0.8940 ± 0.050	43.630	< .001
	Decision-Congruent	0.6560 ± 0.066	24.058	< .001
Confidence	**Balance**	0.7085 ± 0.175	9.922	< .001
	Decision-Congruent	0.6721 ± 0.137	11.987	< .001

Supplementary Table 6. Planned *post-hoc* t-tests for Choice Probabilities (CPs) to test whether they are significantly above chance (CP = 0.5). All results surpass the criterion for significance even after Bonferroni-Holm correction for multiple comparisons[46].

	F	p
Rule: Balance or Decision-Congruent	F(1,5) = 1.952	.221
Noise	F(10,50) = 11.613	< .001
Interaction: Rule x noise	F(10,50) = 3.399	.002

Supplementary Table 7. Results of 2 x 11 repeated measures ANOVA testing the predictive power of two generative models of Confidence: Bayesian ideal observer (Balance-Of-Evidence) and Decision-Congruent Evidence Bayesian heuristic observer.

	P **(80-120 Hz)**	***P*** **(30-190 Hz)**	***P*** **(80-120 Hz) /** ***P*** **(30-190 Hz)**
Decision	1.2333 ± 0.0634	1.2168 ± 0.0610	1.0135
Confidence	1.1942 ± 0.1296	1.1850 ± 0.1255	1.0078

Supplementary Table 8. Predictive power comparisons between high-gamma and all gamma-high-gamma frequencies.

Supplementary Notes

One potential concern in the analyses presented here relates to the limited spatial resolution of ECoG in clinical patients: electrodes are implanted in heterogeneous locations of clinical relevance, so spatial coverage is both relatively sparse and varied across subjects in comparison to whole-brain imaging methods. Indeed, a common approach is to specifically investigate regions of interest known to be involved in a particular computation, for example posterior parietal cortex for dot-motion discrimination perceptual decisions. Likewise, areas presumed to be involved in confidence computations (e.g., pulvinar, orbitofrontal cortex, intraparietal sulcus, etc.) are often targeted in whole-brain analyses. Unfortunately, the use of ECoG in human patients undergoing surgery for reasons unrelated to the present research precludes specific targeting of these regions of interest in the present investigation, as few or no electrodes reached these areas of interest.

However, concluding that Decisions and Confidence judgments rely on spatiotemporally dissociable computations requires only that we demonstrate any difference, not that we are committed to claiming this difference is ubiquitous across all brain areas. Although there are shared Features between Decision to Confidence representations (clustered primarily in occipital regions), the degree of Feature overlap is small. In this way, ECoG not only provides evidence that Decision and Confidence are dissociable, but also provides key information about the neuroanatomical loci of their similarities and differences. That we have quantified and localized these dissociations suggests that the difference between Type 1 (Decision) and Type 2 (Confidence) judgments is robust despite the limitations of ECoG spatial coverage.

Indeed, the results of all analyses demonstrate that (a) there is a difference in how Decisions and Confidence are computed from available evidence (e.g. the significant ANOVA interaction in the choice probability analyses), (b) with a forward model we can estimate the trial-by-trial accuracy of a subject's Decisions and Accuracy but not his/her Confidence judgments (suggesting that Confidence relies on something different from simply a readout of trial-by-trial Accuracy), and (c) the Decision-Congruent model fits subjects' data *better* than the Balance model in the Bayesian observer analysis. Although the spatial coverage may not allow us to sample every neuron, the evidence available indicates that the computations leading to Decisions and Confidence judgments are indeed dissimilar in a way that is consistent across three different analysis approaches. More studies should be done using neuroimaging methods with more comprehensive spatial coverage and better resolution to confirm the results presented here.

Supplementary References

1. Levitt, H. Transformed Up-Down Methods in Psychoacoustics. *J. Acoust. Soc. Am.* **49,** 467–467 (1971).
2. Watson, A. B. & Pelli, D. G. QUEST: A Bayesian adaptive psychometric method. *Percept. Psychophys.* **33,** 113–120 (1983).
3. Percival, D. B. & Walden, A. T. *Spectral analysis for physical applications*. (Cambridge University Press, 1993).
4. Grandchamp, R. & Delorme, A. Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Front. Psychol.* **2,** 236–236 (2011).
5. Crone, N. E., Sinai, A. & Korzeniewska, A. in (ed. Research, C. N. A. W. K. B. T.-P. in B.) **159,** 275–295 (Elsevier, 2006).
6. Crone, N. E., Boatman, D., Gordon, B. & Hao, L. Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. *Clin. Neurophysiol.* **112,** 565–582 (2001).
7. Hermes, D., Miller, K. J., Wandell, B. A. & Winawer, J. Stimulus Dependence of Gamma Oscillations in Human Visual Cortex. *Cereb. Cortex* **25,** 2951–2959 (2015).
8. Hipp, J. F., Engel, A. K. & Siegel, M. Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* **69,** 387–396 (2011).
9. Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J. & Hsiao, S. S. Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. *Journal of Neuroscience* **28,** 11526–11536 (2008).
10. Liu, J. & Newsome, W. T. Local Field Potential in Cortical Area MT: Stimulus Tuning and Behavioral Correlations. *J. Neurosci.* **26,** 7779–7790 (2006).
11. Belitski, A. *et al.* Low-frequency local field potentials and spikes in primary visual cortex convey

independent visual information. *J. Neurosci.* **28,** 5696–5709 (2008).

12. Rasch, M. J., Gretton, A., Murayama, Y., Maass, W. & Logothetis, N. K. Inferring Spike Trains From Local Field Potentials. *J. Neurophysiol.* **99,** 1461–1476 (2008).
13. Ray, S., Hsiao, S. S., Crone, N. E., Franaszczuk, P. J. & Niebur, E. Effect of Stimulus Intensity on the Spike–Local Field Potential Relationship in the Secondary Somatosensory Cortex. *J. Neurosci.* **28,** 7334–7343 (2008).
14. Whittingstall, K. & Logothetis, N. K. Frequency-Band Coupling in Surface EEG Reflects Spiking Activity in Monkey Visual Cortex. *Neuron* **64,** 281–289 (2009).
15. Ray, S. & Maunsell, J. H. R. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9,** (2011).
16. Ritaccio, A. *et al.* Proceedings of the Second International Workshop on Advances in Electrocorticography. *Epilepsy Behav.* **22,** 641–650 (2011).
17. Winawer, J. *et al.* Asynchronous broadband signals are the principal source of the bold response in human visual cortex. *Curr. Biol.* **23,** 1145–1153 (2013).
18. Kunii, N., Kamada, K., Ota, T., Kawai, K. & Saito, N. Characteristic profiles of high gamma activity and blood oxygenation level-dependent responses in various language areas. *Neuroimage* **65,** 242–249 (2013).
19. Esposito, F. *et al.* Cortex-based inter-subject analysis of iEEG and fMRI data sets: application to sustained task-related BOLD and gamma responses. *Neuroimage* **66,** 457–468 (2013).
20. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412,** 150–157 (2001).
21. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning, with Applications in R*. (Springer, 2015).
22. Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. & Lin, C.-J. LIBLINEAR: A Library for Large

Linear Classification. *J. Mach. Learn. Res.* **9,** 1871–1874 (2008).

23. Chang, C. C. & Lin, C. J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2,** 1–27 (2011).

24. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics*. (John Wiley & Sons, Inc., 1966).

25. Chen, Y.-W. & Lin, C.-J. in (eds. Guyon, I., Nikravesh, M., Gunn, S. & Zadeh, L. A.) 315–324 (Springer Berlin Heidelberg, 2006).

26. Chang, Y. W. & Lin, C. J. Feature ranking using linear SVM. *J. Mach. Learn. Res.* **3,** 53–64 (2008).

27. Peters, M. A. K. & Lau, H. Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *Elife* 10.7554/eLife.09651 (2015).

28. Maniscalco, B., Peters, M. A. K. & Lau, H. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Atten. Percept. Psychophys.* (2016). doi:10.3758/s13414-016-1059-x

29. King, J.-R. & Dehaene, S. A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369,** (2014).

30. Yuille, A. L. & Bülthoff, H. H. in (eds. Knill, D. C. & Richards, W.) 123–161 (Cambridge University Press, 1996).

31. Kubánek, J., Miller, K. J., Ojemann, J. G., Wolpaw, J. R. & Schalk, G. Decoding flexion of individual fingers using electrocorticographic signals in humans. *J. Neural Eng.* **6,** 066001–066001 (2009).

32. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn. Sci.* **18,** 203–210 (2014).

33. King, J.-R., Pescetelli, N. & Dehaene, S. Selective maintenance mechanisms of seen and unseen sensory features in the human brain. *Neuron* 1–30 (2016).

34. Fries, P., Nikolic, D. & Singer, W. The gamma cycle. *Trends Neurosci.* **30,** 309–316 (2007).

35. Giraud, A.-L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15,** 511–517 (2012).

36. Haig, A. R., Gordon, E., Wright, J. J., Meares, R. A. & Bahramali, H. Synchronous cortical gamma-band activity in task-relevant cognition. *Neuroreport* **11,** 669–675 (2000).

37. Jensen, O., Kaiser, J. & Lachaux, J. P. Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci.* **30,** 317–324 (2007).

38. van de Nieuwenhuijzen, M. E. *et al.* Decoding of task-relevant and task-irrelevant intracranial EEG representations. *Neuroimage* **137,** 132–139 (2016).

39. Fell, J., Fernández, G., Klaver, P., Elger, C. E. & Fries, P. Is synchronized neuronal gamma activity relevant for selective attention? *Brain Res. Rev.* **42,** 265–272 (2003).

40. Keil, A., Müller, M. M., Ray, W. J., Gruber, T. & Elbert, T. Human gamma band activity and perception of a gestalt. *J. Neurosci.* **19,** 7152–7161 (1999).

41. Tallon-Baudry, C. & Bertrand, O. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci.* **3,** 151–162 (1999).

42. Womelsdorf, T. *et al.* Modulation of neuronal interactions through neuronal synchronization. *Science* **316,** 1609–1612 (2007).

43. Schoffelen, J.-M., Oostenveld, R. & Fries, P. Neuronal Coherence as a Mechanism of Effective Corticospinal Interaction. *Science* **308,** 111–113 (2005).

44. Bahramisharif, A. *et al.* Propagating neocortical gamma bursts are coordinated by traveling alpha waves. *J. Neurosci.* **33,** 18849–18854 (2013).

45. Potes, C., Brunner, P., Gunduz, A., Knight, R. T. & Schalk, G. Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing. *Neuroimage* **97,** 188–195 (2014).

46. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. Stat. Theory Appl.* 65–70 (1979).