Cerebral Cortex Advance Access published August 11, 2014

Cerebral Cortex
doi:10.1093/cercor/bhu181

The Relationship between Perceptual Decision Variables and Confidence in the Human Brain

Martin N. Hebart[1,2,3,4], Yoren Schriever[5], Tobias H. Donner[1,6,7,†] and John-Dylan Haynes[1,2,3,8,†]

[1]Bernstein Center for Computational Neuroscience, Charité Universitätsmedizin, 10115 Berlin, Germany, [2]Berlin Center for Advanced Neuroimaging, Charité Universitätsmedizin, 10117 Berlin, Germany, [3]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10099 Berlin, Germany, [4]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany, [5]Department of Psychology, University of Utrecht, 3584 CS, Utrecht, The Netherlands, [6]Department of Psychology, University of Amsterdam, 1018 XA, Amsterdam, The Netherlands, [7]Cognitive Science Center, University of Amsterdam, 1018 WS, Amsterdam, The Netherlands and [8]Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany

Address correspondence to Martin Hebart, Department of Systems Neuroscience, W34, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20251 Hamburg, Germany. Email: m.hebart@uke.de
[†]Tobias H. Donner and John-Dylan Haynes contributed equally.

Perceptual confidence refers to the degree to which we believe in the accuracy of our percepts. Signal detection theory suggests that perceptual confidence is computed from an internal "decision variable," which reflects the amount of available information in favor of one or another perceptual interpretation of the sensory input. The neural processes underlying these computations have, however, remained elusive. Here, we used fMRI and multivariate decoding techniques to identify regions of the human brain that encode this decision variable and confidence during a visual motion discrimination task. We used observers' binary perceptual choices and confidence ratings to reconstruct the internal decision variable that governed the subjects' behavior. A number of areas in prefrontal and posterior parietal association cortex encoded this decision variable, and activity in the ventral striatum reflected the degree of perceptual confidence. Using a multivariate connectivity analysis, we demonstrate that patterns of brain activity in the right ventrolateral prefrontal cortex reflecting the decision variable were linked to brain signals in the ventral striatum reflecting confidence. Our results suggest that the representation of perceptual confidence in the ventral striatum is derived from a transformation of the continuous decision variable encoded in the cerebral cortex.

Keywords: confidence, decision-making, fMRI, multivariate pattern analysis, signal detection theory

Introduction

Human observers are remarkably good at estimating the accuracy of their perceptual judgments, an ability known as perceptual confidence. Our confidence often closely corresponds to the accuracy of categorical perceptual choices (Peirce and Jastrow 1884; Vickers 1979; Baranski and Petrusic 1998), suggesting a close link between the processes underlying confidence and perceptual decision-making (Gold and Shadlen 2007; Heekeren et al. 2008). Despite this connection, surprisingly little is known about the neural mechanisms subserving perceptual confidence judgments (Kepecs et al. 2008; Kiani and Shadlen 2009; Fleming et al. 2012) and in particular how they relate to mechanisms of perceptual decision-making.

Signal detection theory and related accounts (Green and Swets 1966; Macmillan and Creelman 2005) postulate that both perceptual choices and perceptual confidence are based on a continuous decision variable (DV). The DV is typically defined as the amount of available information in favor of one perceptual interpretation of a stimulus and is based on the (immediate or accumulated) sensory evidence available to the observer. (Fig. 1; Green and Swets 1966; Macmillan and Creelman 2005; Gold and Shadlen 2007). Signal detection theory is agnostic to the temporal evolution of this variable. Observers make a categorical choice (e.g., "motion up" vs. "motion down") by comparing the DV against a criterion (Fig. 1*B*), and they generate their choice-independent confidence based on the absolute distance of the DV to this criterion (Fig. 1*A*). In turn, the DV could alternatively be described as a signed version of confidence (Fig. 1*D*). In that way, signal detection theory offers a direct mathematical mapping of the DV and confidence (Fig. 1*D*). This relationship suggests that human observers generate confidence by simply rectifying the DV (Fig. 1*C*).

Recent evidence from animal electrophysiology supports the idea that confidence is closely linked to the DV, by showing that single unit activity in macaque lateral intraparietal area predicted both the choice of the monkey and whether it was going to opt-out on a difficult choice (Kiani and Shadlen 2009). However, these responses were coupled to the specific perceptual decision (e.g., "motion up"), leaving open the question of whether the monkey also computed a more general, choice-independent perceptual confidence signal. Such a signal has been reported in rat orbitofrontal cortex (Kepecs et al. 2008), and it has been suggested that the striatum of macaques carries a similar, choice-independent representation of confidence (Ding and Gold 2012). These two types of decision-related neuronal signals—one reflecting the DV and the other choice-independent confidence—have not yet been measured simultaneously, and it is unknown whether and how they are related to each other in the human brain.

The aim of the present study was to search for representations of the continuous DV and of perceptual confidence throughout the human brain and test whether these two signals are related. To this end, we used fMRI combined with multivariate "searchlight decoding" (Kriegeskorte et al. 2006; Haynes et al. 2007). To link these two signals, we used a multivariate connectivity approach, exploiting the interdependence between patterns of activity in several brain regions. This approach may shed light on the question how brain signals reflecting the DV in one region can be converted into brain signals reflecting confidence in another region.

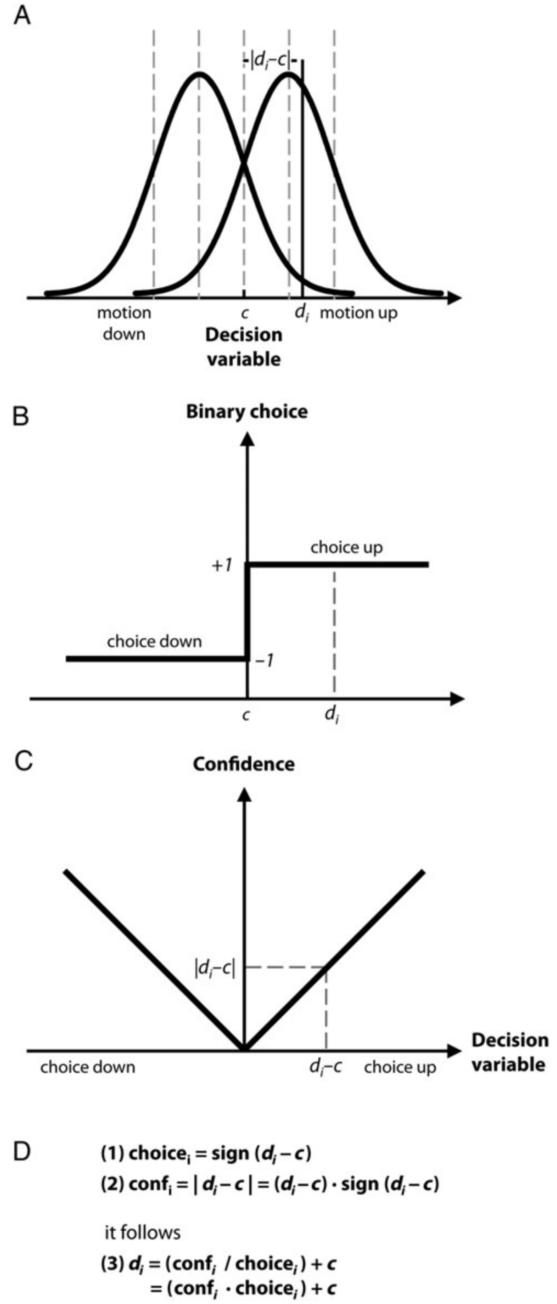
A
$|d_i - c|$
motion down
c
d_i
motion up
Decision variable
B
Binary choice
choice up
+1
choice down
−1
c
d_i
C
Confidence
$|d_i - c|$
choice down
$d_i - c$
choice up
Decision variable
D
(1) $\text{choice}_i = \text{sign}\,(d_i - c)$
(2) $\text{conf}_i = |\,d_i - c\,| = (d_i - c) \cdot \text{sign}\,(d_i - c)$
it follows
(3) $d_i = (\text{conf}_i\ /\ \text{choice}_i) + c$
$= (\text{conf}_i \cdot \text{choice}_i) + c$

Figure 1. (*A*) The representational model underlying signal detection theory. The internal DV defines the *x*-axis and reflects the amount of sensory evidence in favor of one choice. The probability density functions are distributions of DV across repeated presentations of either "motion up" or "motion down" stimuli. The DV for the *i*th choice is denoted by d_i and the criterion of the subject by the constant c. Confidence reflects the distance of d_i to c. (*B*) The DV is transformed into binary choices using a step function, with the step located at the criterion. Please note that the assignment of the sign to the direction of choice is arbitrary and can be used interchangeably. (*C*) The relationship of confidence and the DV can be described by a modulus function: Confidence is an unsigned variable and reflects the absolute value of the difference between the DV and the criterion. (*D*) Equations describing the relationship between DV (*A*), choice (*B*), and confidence (*C*). Please note that c can be treated as a constant.

Materials and Methods

Participants

Nineteen young and neurologically healthy subjects took part in the study and were compensated with 7 € per hour for behavioral pre-tests and with 10 € per hour for the scanning experiment. One participant was subsequently excluded due to excessive head motion, leaving 18 subjects (6 female, 1 left-handed, mean age: 26.4 years, SD: 3.4). All subjects provided informed consent for participation. The study was approved by the local ethics committee of the Humboldt University in Berlin and conducted in accordance with the Declaration of Helsinki.

Procedure

The task of the subject was to observe a random dot motion (RDM) stimulus (Braddick 1974) and then perform two consecutive responses: (1) a binary perceptual judgment of the dominant direction of motion of the stimulus ("motion up" or "motion down") and (2) a rating of confidence of the judgment on a continuous scale (Fig. 2*A*). In the main experiment, the motion coherence of the RDM stimulus was kept fixed to the 75% threshold of each subject that had been estimated in previous practice sessions (see below). A trial began by a fixation period of 2, 4, or 6 s (counterbalanced within each experimental run across motion direction and stimulus-response-mapping screen). Within that period, 0.5 s before the onset of the RDM stimulus, the central fixation cross (0.25 dva) turned yellow for 0.2 s to prepare the subject of the upcoming presentation of the RDM. The RDM stimulus was presented for 1 s, followed by fixation of an otherwise blank screen for 1 s. Then, the stimulus-response-mapping screen was shown for 1.5 s, during which the subject could indicate the perceived direction of motion by a button press corresponding to the chosen stimulus direction indicated by the response-mapping screen (see below). This was followed by another 0.5 s of fixation and the confidence screen for 3 s during which the subject could indicate their confidence by moving a white dot with a trackball to the part of the confidence bar corresponding to their confidence. Each experimental run was divided into 2 longer blocks of 36 trials spaced by 30-s breaks during which the subject was presented a countdown and was instructed to relax the mind and continue fixating. In the MRI scanner, each subject participated in 5 experimental runs of 72 trials each. In total, the experiment lasted about 64 min.

Stimuli

All stimuli were created with Matlab (Mathworks) and presented using the Cogent toolbox (http://www.vislab.ucl.ac.uk/Cogent). Stimuli were all shown on a black background. The RDM stimulus was created in a square region of 10 × 10 dva, but only the region within a circular annulus was visible (outer radius: 5 dva, inner radius: 0.85 dva). The stimulus consisted of 500 dots (5 dots/dva^2) that each moved at a speed of 2.5 dva/s and had a diameter of 5.5 arcmin. Dots were separated in signal and noise dots. Signal dots all moved in the same direction (up or down) whereas noise dots were assigned equally spaced directions of motion including the target motion directions (Fig. 2*A*, top left). We introduced several steps to make sure that subjects had to integrate motion information across space and time, as opposed to tracking only a small set of dots across a few frames. First, dots were placed at random positions on the square region but constrained in a way that minimized the presence of clusters of dots moving in the same direction at stimulus onset (the first 4 frames were discarded to remove any visible pattern emerging from this step). Second, each dot was assigned a limited lifetime of 200 ms, and dots exceeding this time were subsequently placed at a random location in the RDM stimulus. For the first lifetime cycle, dot durations were set to random times between 0 and 200 ms. Finally, dots leaving the square region were redrawn on the opposite side, and dots leaving the annulus were faded out to prevent sharp gradients at the borders.

In the scanner, stimuli were back-projected on a translucent screen (display area: 24.5 × 18.5 cm) in the rear of the scanner that was viewed at an approximate distance of 60 cm through a surface mirror mounted on the head coil. Subjects responded with their right hand using an MR-compatible trackball (Current Designs. Inc.) by pressing the left button with the thumb, the right button with the middle finger, and navigating the trackball with the index finger. In order to experimentally separate motor-related and choice-related brain signals, we used a central response-mapping cue (circle with 0.3 dva diameter) that cued subjects which button to press for which answer (Haynes

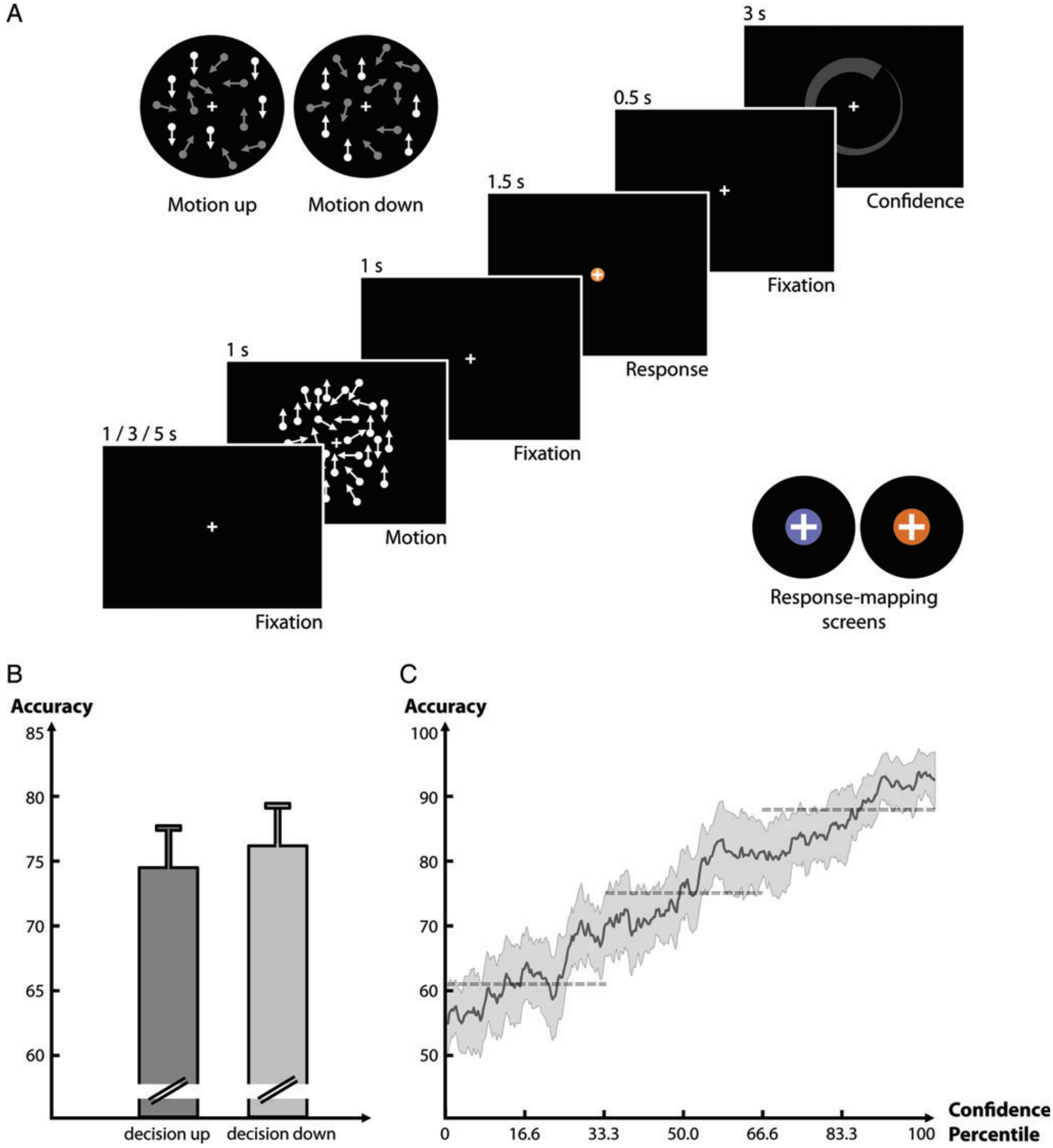

Figure 2. Experimental design and behavioral results. (*A*) Each subject was presented an RDM pattern of which an individually calibrated percentage of dots moved either upward or downward whereas all other dots moved in random directions. The subject had to indicate the dominant motion direction of the dots. For the motor response, a response-mapping screen was introduced to decouple motor-related brain signals from perceptual choices. Finally, subjects used a trackball to indicate the confidence in their choice on a continuous scale. (*B*) Behavioral accuracy, separated for both choices. Error bars denote the standard error of the difference of the mean. (*C*) Accuracy as a function of confidence, using a sliding average with a window width of 5% of all trials. Dashed lines represent the behavioral accuracy for each of the 3 levels of confidence used in later analyses. The shaded area denotes the 95% confidence interval.

et al. 2007; Bennur and Gold 2011; Hebart et al. 2012). The response-mapping rules were as follows: if the choice was "upward," a blue response-mapping cue indicated a left button press, and an orange response-mapping cue indicated a right button press (and vice versa for "downward" choices). Subjects were instructed to respond even when they were very unsure.

For the confidence judgment, a confidence scale was created that consisted of a gray bar that was wrapped around a virtual circle of 2 dva radius and that increased linearly in width from 0 to 0.7 dva (Kahnt et al. 2011). This formed a continuous scale for judging the confidence. The orientation and direction of the bar was changed randomly from trial to trial to avoid low-level stimulus confounds. Subjects were instructed that the scale ranged from "very sure about the direction of motion" to "no idea at all," and that the center of the scale indicated the border between rather sure and rather unsure. Furthermore, subjects were instructed during training that they should try not to restrict their judgments to only a small portion of the scale, unless their confidence really did not fluctuate. This instruction was given to improve the discriminatory power of their judgment between trials that may be masked by inaccuracy in pointing responses using the trackball. Please note that this instruction does not change the "relative" confidence ratings and our procedure separated different levels of confidence using percentiles (see below). When subjects noticed that they had made a response-mapping error, they were instructed to withhold a confidence judgment. We wanted to prevent low confidence judgments to be confounded by errors in response mapping (Daniel and Pollmann 2012). To reduce the effect of the thickness of the confidence bar on confidence response difficulty, subjects were told that the mouse did not have to hit the confidence bar but would only have to be close to it.

Practice Sessions Prior to the fMRI Experiment

Each subject practiced the task in two separate psychophysics sessions prior to scanning. Here, subjects viewed stimuli on a 19-inch CRT monitor at a distance of 50 cm, with sizes and speeds of stimuli matched in dva to stimulation in the scanner. In the first session, each subject was familiarized with the stimulus-response-mapping screens and the trackball, followed by an adaptive staircase procedure (QUEST, Watson and Pelli 1983) of 150 trials to find the motion coherence at which the subject performed 75% correct. The staircase procedure was also used to familiarize the subject with the level of motion coherence that would be shown in the scanner. In the second session, the QUEST procedure was repeated with 96 trials to account for between-session learning effects, followed by 3 runs of the actual experiment where the stimulus intensity was fixed and subjects reported both their choice and their confidence. The average coherence used in the experiment was 12.03% (±2.64). Subjects received feedback during the adaptive staircase procedure to speed up learning and increase motivation, but not during the confidence task and not in the scanner. Otherwise, brain signals related to explicit feedback would be difficult to separate from confidence-related brain signals. During the acquisition of the anatomical image, but prior to functional scanning, the coherence threshold was determined again using the QUEST procedure and when necessary adapted slightly to account for differences between experimental sessions. Importantly, motion coherence was then fixed throughout all functional runs from each subject, eliminating any trial-to-trial variations in the strength of the physical stimulus.

Post-Experimental Sessions

After the main experiment, all subjects took part in a functional localizer run to locate motion-sensitive region MT+/V5 in each individual subject (Tootell et al. 1995). Seven of the subjects completed another 144 trials each outside the scanner, now while using a video-based eye tracker (sampling rate: 250 Hz, Cambridge Research Systems). These 7 subjects were able to maintain central fixation, each of them with a maximum of 2 saccades exceeding 1.5 dva during presentation of the RDM stimuli.

Behavioral Data Analysis

Behavioral data of subjects were analyzed for accuracy, response time, and confidence. For subsequent fMRI analyses, the continuous scale of confidence ratings was divided into 3 confidence bins. Confidence bins were created by separating the confidence ratings of all trials of each individual subject at the tertiles, that is, the 33.3 and the 66.6 percentiles. We call these confidence bins "low," "medium," and "high confidence." A comparison of binned confidence with continuous confidence ratings—where possible—yielded very similar results to the use of continuous confidence judgments.

MRI Data Acquisition

All scanning was performed using a 3.0 Tesla TIM Trio MRI (Siemens) and a 12-channel head coil. An anatomical scan was acquired for intersubject normalization and mapping of functional to reference structural images (T1-weighted MPRAGE volume, TE: 2.52 ms, TR: 1900 ms, flip angle: 9°, FOV: 256 mm, matrix size: 256 × 256, slice thickness: 1 mm, 192 slices). In addition, 2247 functional volumes were acquired per subject (T2*-weighted gradient-echo echo-planar images, TE: 30 ms, TR: 2000 ms, flip angle: 78°, FOV: 192 mm, matrix size: 64 × 64, in-plane voxel size: 3 × 3 mm, slice thickness: 3 mm, gap: 0.3 mm, 33 slices, descending sequence), corresponding to the experimental runs and the functional localizer. The first 2 functional volumes were automatically discarded in each run to allow for T1 equilibrium effect.

MRI Data Preprocessing and Analysis Streams

MRI data preprocessing and general linear model (GLM)-based statistical analyses were performed using SPM8 (SPM8, Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/). Using the structural images of all subjects, a template structural reference image was created using DARTEL (Ashburner 2007). For comparability with other studies, this template was brought to MNI space using affine transformation, and the transformation matrices of each subject were saved. Three analysis streams were carried out on the functional images: the standard mass-univariate approach in fMRI data analysis, a multivariate searchlight-based classification method, and a multivariate searchlight-based regression method. We will refer to these steps as the "univariate analysis," the "multivariate classification analysis," and the "multivariate regression analysis," respectively. In the univariate analysis, we were interested to find brain regions in which the BOLD signal amplitude co-varied with confidence, independent of the direction of the moving dots. In the multivariate classification analysis, we sought to identify brain regions that represented the decision of the subject independent of the motion direction of the stimulus. Finally, in the multivariate regression analysis, we were looking for brain regions where patterns of brain activity predicted the subject's decision and their associated confidence. In all three analyses, functional images of each subject were first spatially realigned and slice-timing corrected. For the univariate analysis, the images were then spatially normalized using the abovementioned transformation matrix and smoothed with a Gaussian kernel (6 mm FWHM). For both multivariate analyses, functional data were not spatially warped or smoothed.

Confidence: Mass-Univariate fMRI Data Analysis

A GLM with 10 regressors per run was used to estimate BOLD response amplitudes in each voxel separately. Two regressors modeled the BOLD response related to perceptual decision-making. The first was a stimulation regressor that consisted of a boxcar function with onsets at RDM stimulus onset and duration of 2 s, convolved with a canonical hemodynamic response function (HRF). This duration was the time from RDM stimulus onset to the onset of the response-mapping screen. This regressor reflected common neural responses to all stimulation conditions. The second regressor was a parametric modulator that reflected the linear increase or decrease in BOLD signal across different confidence levels. To this end, we split subjects' confidence ratings into 3 different bins. Although the latter was not a methodological requirement for the mass-univariate analysis, it was necessary for the multivariate regression analysis (see below). We chose to bin the trials by confidence in both analysis approaches for consistency.

In addition to these 2 regressors, missed trials were modeled in a separate third regressor convolved with an HRF, as well as 6 additional head motion regressors and a constant term. To correct for low-frequency drifts, a 1/128-Hz high-pass filter was applied to the data. Across all runs, the model contained 50 regressors. Using the parametric regressor, we separately tested for positive or negative linear relationships between the BOLD signal amplitude and the confidence of the subject across all experimental runs. The resulting contrast images were submitted to two group-level t-tests, one for a positive and one for a negative relationship. The statistical cutoff was set to $P < 0.0001$, corrected family-wise for cluster size at $P < 0.05$.

Since response time correlated with confidence, we were interested to see which brain regions exhibited confidence-related responses after accounting for response time. For this, we repeated the same analysis as mentioned above but used response time as the first and confidence as the second parametric regressor in the analysis. We orthogonalized confidence with respect to response time, that is, the confidence regressor could only explain additional variance that was not already explained by response time.

Searchlight-Based Multivariate Classification Analysis: Binary Choices

In a first step, we delineated brain regions in which activity reflected the subjects' choices. To this end, we classified choices based on brain signals, separately for both directions of motion, similar to the computation of choice probabilities (Britten et al. 1996). Since motion coherence was kept constant, this eliminates regions where signals are selective to physical motion direction alone and are not linked to trial-by-trial fluctuations in choices, which would be required for a DV. Please note that this post hoc separation neither affects the representation of the stimulus-related probability density functions (Fig. 1A) nor the DV. Please also note that this contrast could alternatively be explained by correctness which we addressed in additional control analyses (see Results and Supplementary Analyses). We could then in a second step limit our search for the DV (see below) to those brain regions that exhibit choice-related activity not confounded by motion direction. Ideally, the representation of the DV would be directly assessed in one step, separately for each direction of motion. However, sorting trials based on both motion direction and confidence would lead to an insufficient number of trials for multivariate analyses.

Prior to decoding, we first ran a GLM analysis in which we modeled the subject's choices separately for both directions of motion ("motion up, choice up," "motion up, choice down," "motion down, choice up," and "motion down, choice down"). As regressors of no interest, we additionally modeled missed trials in 1 regressor, movement in 6 regressors, and added a baseline regressor, leading to 12 regressors per run. Next, we used the parameter estimates of the GLM reflecting the choices of the subject in two separate searchlight classification analyses (Kriegeskorte et al. 2006; Haynes et al. 2007), one for each stimulus motion direction. This was performed within each subject, using a leave-one-run-out cross-validation scheme. All multivariate analyses were carried out with The Decoding Toolbox (Görgen et al. 2012) and other in-house software.

For the searchlight classification, a sphere of voxels was selected around a given voxel with a radius of 10 mm (139 voxels). From these voxels, the GLM parameter estimates were extracted, separately for all 5 runs and for the 2 regressors reflecting the two choices for only one of the motion directions (e.g., "motion up, choice up" and "motion up, choice down"). These served as 10 pattern vectors that were used for multivariate pattern classification of choices. Then, we used a linear support vector classification model in the implementation of LIBSVM (Chang and Lin 2011), with a standard cost value of $c = 1$. We assigned each vector a label corresponding to the choice of the subject ("choice up" vs. "choice down"). The pattern vectors of all but one run were then used to train a support vector machine (SVM) to predict the categories of both left-out patterns in each run. After training, we validated the model by comparing the true labels of the left-out patterns with the labels predicted from the model. We repeated this train-test approach iteratively for each run and calculated a mean cross-validated accuracy across all runs for this searchlight. The center voxel of the searchlight sphere was assigned this accuracy, and the whole searchlight procedure was repeated for all voxels in the brain. This generated a continuous map of mean cross-validated accuracies for each subject, representing the distributed information (within the extent of the searchlight) about the choice of the subject for one direction of motion. This searchlight analysis was then repeated for the other direction of motion (in our example "motion down, choice up" vs. "motion down, choice down"). Both maps of each subject were averaged, and this combined map was spatially transformed to MNI space and smoothed for group-level analyses. The chance-level accuracy for classification was set to 50%, because classification was based on an equal number of regressors for both categories. We were interested to find searchlight voxels where across the group the mean classification accuracy of choices was significantly above chance. For this, we subjected all accuracy maps to a group-level t-test (P-cutoff < 0.0001, corrected family-wise for cluster size at $P < 0.05$).

Please note that for classification with a total of 10 patterns, each cross-validation iteration involves only 4 training samples per class. This is a comparably small number for supervised classification, and one might be worried that this could compromise the reliability of the present results. However, the fact that run-wise beta estimates are more reliable than single-trial beta estimates should cancel the effect of small training sample size, at least for classifiers such as SVMs, which make little or no use of the variability of estimates. In practice, the approach can lead to higher accuracies (Ku et al. 2008), slightly improved power (Allefeld and Haynes 2014) and is computationally much less demanding than trial-wise classification. This makes the approach a useful method for inferential purposes in fMRI decoding studies (Poldrack et al. 2011, p. 164), which has successfully been applied in a number of previous group studies (e.g., Haynes et al. 2007; Kahnt et al. 2010; Christophel et al. 2012; Hebart et al. 2012).

Searchlight-Based Multivariate Regression Analysis: Continuous Decision Variable

In order to investigate which brain regions changed their patterns of activity depending on the confidence of the subject, we employed a multivariate regression analysis. The decision value is a theoretical variable that cannot be directly observed, so we operationalized it in accordance with signal detection theory. This theory postulates that, on each trial, the subject's confidence rating reflects the absolute distance of an internal decision variable from a criterion, and the subject's binary choice reflects the sign of that difference (Fig. 1, Green and Swets 1966; Macmillan and Creelman 2005). Under this assumption, a proxy of the DV (more specifically: of the difference between the DV and the criterion) on each trial can be constructed from the subject's behavioral reports, as follows: the binary choice (coded as −1 or 1 for "up" and "down," respectively) multiplied with the binned confidence rating (here coded as 0.5, 1.5, and 2.5 for equidistant labels). Binning the trials in 3 confidence levels was necessary to achieve sufficient statistical power; not binning by confidence would have yielded too few repetitions of each particular confidence rating across the experiment. For simplicity, we will refer to the resulting variable as "decision variable" in the following. Please note that the multivariate regression analysis does not depend on the criterion used by the subjects: in terms of detection theory, the criterion represents only a constant added to each of the labels used in the regression analysis, which is not used by the SVM and would thus leave the results of this analysis unchanged (see also Formula [3] in Fig. 1D).

We searched for brain regions that exhibited a linear relationship between local patterns of brain activity and the DV. For that purpose, we first ran a GLM analysis in which we modeled the 3 DVs separately for the 2 choices (up/down). This led to 6 regressors per run. As in the univariate analysis missed trials, motion regressors and a baseline regressor were added to the model.

The parameter estimates of the GLM reflecting the DV were then used for a searchlight regression analysis. This was performed similar to the "multivariate classification analysis" above, using a leave-one-run-out cross-validation approach within each subject (Kahnt et al. 2010) and a searchlight radius of 10 mm. First, we extracted the GLM parameter estimates of all voxels within a sphere around a given voxel for all 6 regressors reflecting the choices and associated confidence. This resulted in 30 pattern vectors that entered our multivariate regression analysis. Second, we used a linear support vector regression (SVR) model in the implementation of LIBSVM (nu-SVR, Chang and Lin 2011), with parameters $nu = 0.5$ and the cost value $c = 1$. Third, each vector was assigned a label ranging in equal steps from −2.5 to 2.5, depending on the DV (corresponding to the range from "choice down, high confidence" to "choice up, high confidence"). An SVR was then trained on the pattern vectors of all but one run to make continuous predictions about the labels of the 6 pattern samples of the left-out run. Fourth, we obtained the predictive accuracy as the correlation between the true labels of the left-out samples and the labels predicted from the regression model. This train-test approach yielded one correlation coefficient per run that was subsequently Fisher-z-transformed and averaged across runs. This mean Fisher-z-transformed correlation coefficient served as a measure of cross-validation performance, with higher positive values reflecting more DV-related information in the patterns of brain activity. The center voxel of the sphere was assigned this z-transformed correlation coefficient. This searchlight procedure was repeated for all voxels throughout the brain. This generated a continuous map of mean cross-validated correlation coefficients for each subject, which represents the locally distributed information that could be extracted from pattern of brain activity about the DV. The map of each subject was then spatially transformed to MNI space and smoothed for group-level analyses. These results were further masked by regions that were shown to carry choice-related information, controlled for motion direction (see above). We were interested to find searchlight voxels where the mean Fisher z-transformed correlation coefficient was larger than 0. For this, we subjected all correlation maps to a group-level t-test (P-cutoff < 0.0001, corrected family-wise for cluster size at $P < 0.05$).

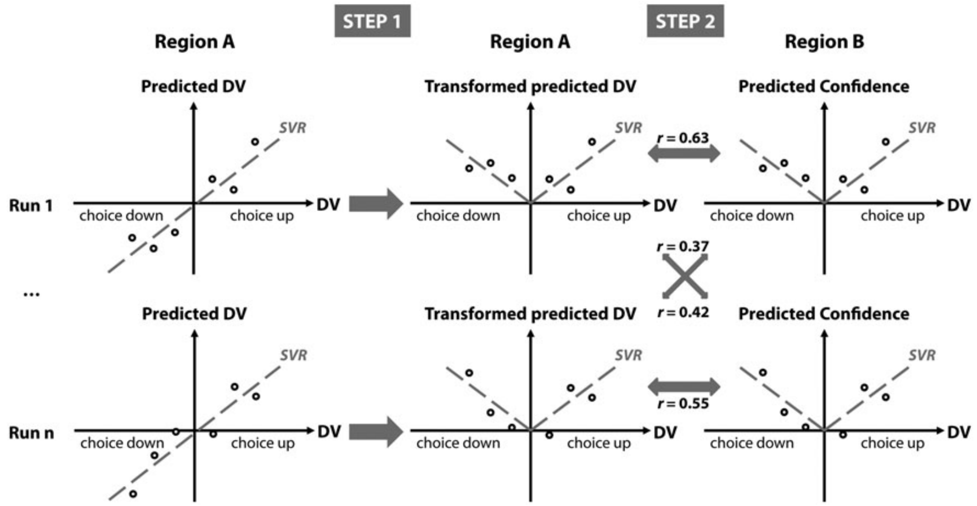

Figure 3. Rationale of the multivariate connectivity analysis. This analysis investigates similarities in the representation of the DV in region A and the representation of confidence in region B, which are specific to each run. In a first step, in region A, the predicted DV of each run (which resulted from the "multivariate regression analysis," see Materials and Methods section) is multiplied with the behavioral choices related to each of these predicted labels (see Fig. 1*D*), creating a transformed predicted DV. In a second step, the transformed predicted DV of each run is correlated with the predicted confidence of the same run. The correlation between directly decoded confidence and DV-based confidence across runs is expected to be non-zero even if there is no interdependence between the 2 regions. For that reason, it is important to compute a distribution of baseline correlation levels in the absence of interaction. All possible permutations of pairs of runs (e.g., run 1 and run *n*) are calculated, and the distribution of correlations is obtained, which is used as a basis for a permutation test.

Multivariate Connectivity: Predicting Confidence Signals in One Brain Region from Decision Variable Signals in another Region

Finally, we tested for more direct links between the neural representation of confidence signals and DVs. The idea was to see whether there is any evidence that the neural confidence signal in one area A might be computed from a neural DV in another area B. If that were the case, one would expect the across-run variation in DV signals in B to result in a corresponding across-run variation in confidence signals in A. Figure 3 illustrates how the transformation would take place. To anticipate the results of the "univariate analysis," we found a signal in bilateral ventral striatum that reflected confidence independent of the choice of the subject (the same result was found using a multivariate version of this analysis, see Results and Supplementary Fig. 2). The following analysis assessed whether it is possible to explain run-wise changes in confidence representations in this ventral striatal region by run-wise changes in DV coding in any other region. As candidate seed regions, we selected the 12 regions that were identified in the multivariate regression analysis reflecting the DV (see Fig. 5*B* and Table 2).

We selected a sphere of voxels around each region's peak (radius 10 mm), providing us with one searchlight per region. The peaks were defined for each subject by inverting the warping carried out by the DARTEL procedure for each subject. From each seed searchlight, we looked up the DV predicted from the multivariate regression analysis, providing us with 6 predicted DVs per run. Next, we transformed these predicted DVs to predicted confidence values according to the assumed relationship between the DV and confidence (Fig. 1*C*): for each run, we removed the mean across all labels (to compensate for potential offsets) and multiplied the labels with the vector [−1 −1 −1 +1 +1 +1], which represented the sign of the choices associated with each predicted label (Fig. 1*B*; please note that a simple rectification of labels would distort the expected systematic deviations). To conduct the analysis of interdependence, these transformed values were correlated with the values predicted from the "multivariate regression analysis" of confidence in the ventral striatum, separately for each run. These correlation coefficients were Fisher-z-transformed and averaged across runs. The whole process was repeated for each seed region separately. Please note that the results of the univariate analysis and the multivariate regression analysis of confidence are equivalent, but showing this direct link is much simpler between two multivariate regression analyses: first, the same number of voxels are used when comparing multivariate analyses, making the results more equivalent in terms of informational content. Second, it would be difficult to formulate this kind of link between a multivariate decoding model and a univariate encoding model (Naselaris et al. 2011), but this link is much simpler for two multivariate decoding models.

Since we selected regions that exhibited a good fit with the DV, the expected correlation between these regions and the ventral striatum is above 0 even in the case of no run-wise interdependence. For this reason, we implemented a permutation test: for each subject, we repeated the above-mentioned connectivity analysis, but this time permuted the runs of the predicted labels. This step should destroy systematic deviations of predicted labels that are specific to each run, but it preserves correlations that may be present in the average across runs. There were 120 possible combinations per subject. For group analysis, we picked one of these 120 combinations per subject and averaged them (yielding a total of $120^{18} = 2.66 \times 10^{37}$ possible combinations). This process was repeated 10^6 times to generate a distribution of correlation coefficients for each region of interest. Significance was determined by permutations that exceeded the non-permuted runs less than 0.0042 times (reflecting Bonferroni correction for $P < 0.05$ across 12 regions).

Results

Behavior

Results are reported with the 95% confidence interval in brackets, unless denoted explicitly. Subjects reported both their choice and confidence on 95.51% (±1.29) of trials, and only trials with both reports were used for later analyses. In agreement with the individually adjusted coherence levels, participants were correct on 75.06% (±2.19) of trials, with no differences between choices for upward or downward motion (Fig. 2*B*, M_{up}: 74.77%, M_{down}: 76.26%, $t_{(17)} = -0.99$, $P = 0.1685$). The tertiles on the confidence scale lay at 41.05% (±5.79) and 66.58% (±4.88), indicating that subjects were sampling confidence quite evenly (for a histogram of raw confidence responses, see Supplementary Fig. 1). Although there was a time gap of at least 1 s between stimulus offset and responses—which should be sufficient for the read-out of potential motion-related information (Roitman and Shadlen 2002) from a visual buffer—response time was

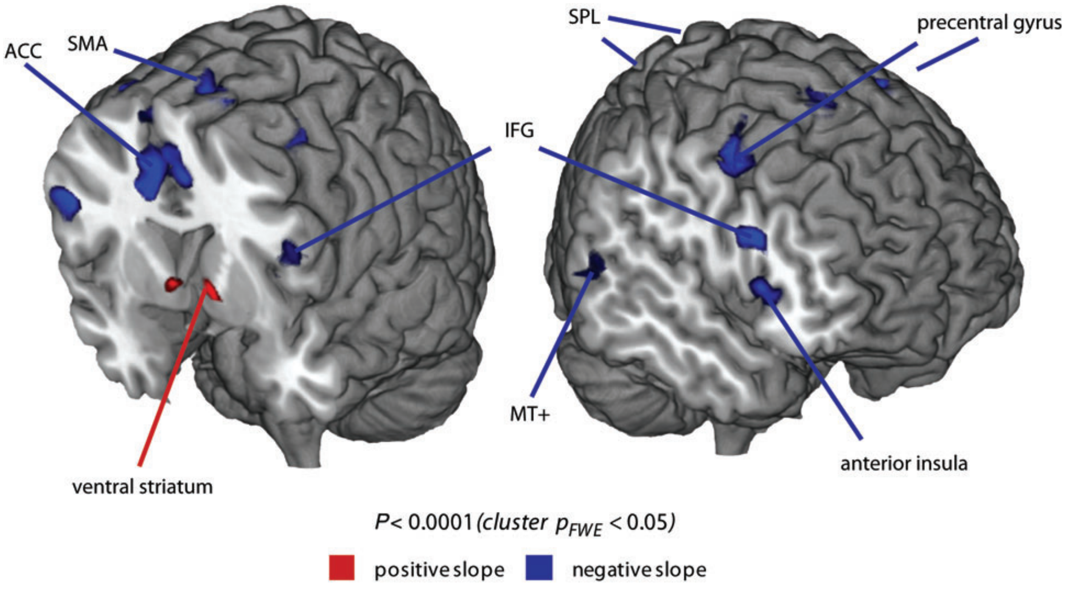

Figure 4. Brain regions exhibiting linear increases and decreases in the BOLD signal in relationship to confidence. A positive slope reflects a positive correlation of confidence and BOLD signal, whereas a negative slope corresponds to a negative correlation.

correlated negatively with accuracy (mean Fisher $z = -0.073$, $P = 0.001$) and confidence (mean Fisher $z = -0.248$, $P < 0.001$).

A central assumption underlying our approach for reconstructing the DV from binary choices and confidence ratings was the existence of a close link between decision confidence and accuracy. Such a link has commonly been observed in analogous perceptual choice tasks (Peirce and Jastrow 1884; Green and Swets 1966; Vickers 1979; Baranski and Petrusic 1998; Macmillan and Creelman 2005). To establish this link for the present data, we calculated the accuracy for each level of confidence on a continuous range, using 5% of all trials for each bin. Accuracy increased monotonically with confidence, approaching chance at lowest levels of confidence, but not ceiling at highest confidence levels (Fig. 2*C*). After binning confidence at the tertiles, the average performance was 61.04% (±2.96), 76.06% (±3.07), and 88.08% (±2.86) correct for the 3 confidence bins, respectively (dashed horizontal lines in Fig. 2*C*). These results are consistent with the notion that confidence may act as a direct function of the DV. It is possible that other cognitive variables also contributed to confidence judgments (Fleming et al. 2012), but the linear relationship between performance and confidence suggests no systematic deviations from this relationship due to such variables (see Discussion for a more detailed treatment of this topic).

Any relationship we find between the neural representation of DV and confidence could possibly stem from a positive behavioral correlation of DV and general confidence. Such a bias would only be possible if there was a correlation of performance with one choice (e.g., motion up), but not the other. To test for this, we applied equation (3) from Figure 1*D* to each confidence bin to derive DVs for each subject. The correlation of DV and confidence was not significant (mean Fisher z: 0.002, $t_{(17)} = 0.118$, $P = 0.4537$), demonstrating that the comparison of both variables is not biased. Similarly, we correlated the binned DV and response time and found no linear relationship between the two (mean Fisher z: −0.059, $t_{(17)} = 1.237$, $P = 0.2331$).

Neural Correlates of Perceptual Confidence

Whereas several brain regions exhibited decreases in BOLD signal with increasing confidence (Fig. 4 and Table 1), only one region displayed a positive relationship with confidence: the ventral striatum around the nucleus accumbens. Only at a much lower threshold, a small cluster in medial orbitofrontal cortex was apparent ($P < 0.01$ uncorrected); however, this cluster did not survive any correction for multiple comparisons; for that reason, we do not discuss it further. For illustrative purposes, we also computed a multivariate regression analysis (see Materials and Methods) in which we decoded choice confidence, with very similar results albeit at a higher level of significance and thus a larger spatial extent (Supplementary Analyses and Supplementary Fig. 2). However, the directionality of the BOLD signal cannot be interpreted in multivariate pattern analysis which is why we focus on the results from the mass-univariate analysis (Jimura and Poldrack 2012).

Since response time correlated with confidence, we were interested to see whether the relationship between confidence and the BOLD signal was mediated by response time. We repeated the same mass-univariate analysis as mentioned above but first removed the variance in BOLD signal amplitude explained by response time. The positive relationship between confidence and BOLD signal amplitude remained, whereas many of the regions previously displaying a negative relationship no longer survived the statistical threshold (see Supplementary Table 1, for all related results). At a more lenient threshold, the results were still present, but in contrast to the ventral striatum, all effects were strongly reduced by the inclusion of response time (interaction with ventral striatum: mean $z = 2.87$, $P = 0.002$). These analyses indicate that a significant portion of most negative correlations with confidence could be related to other cognitive variables leading to variations in response time, such as top-down attention, cognitive control (Miller and Cohen 2001; Corbetta and Shulman 2002; Botvinick et al. 2004). In this study, we focused on regions displaying a "positive" relationship between confidence and BOLD signal amplitude because these could be more unambiguously treated as candidates for a confidence signal (see also Discussion).

Table 1
Brain regions where activity is related to confidence

Region	*X*	*Y*	*Z*	*Z*-value	Cluster size
Positive parametric effect					
Left ventral striatum	−9	9	−3	4.42	17
Right ventral striatum	9	9	−3	4.59	15
Negative parametric effect					
Supplementary motor area/ACC	−9	15	54	5.27	436
Right anterior insula	33	27	3	4.99	103
Right frontal eye field	45	6	45	5.36	89
Right inferior frontal gyrus	51	18	21	5.19	88
Left frontal eye field	−45	6	48	4.60	70
Left inferior frontal gyrus	−45	21	12	4.19	59
Left superior parietal lobule	−21	−54	54	4.38	29
Right middle temporal gyrus	63	−48	0	3.95	20
Left anterior insula	−39	21	−3	3.94	17
Right superior parietal lobule	30	−60	60	4.00	16

Note: A complete list of brain regions with parametric increases and decreases in brain activity in relationship to confidence. The coordinates and *z*-values refer to the peak within each cluster. Results are reported at $P < 0.0001$, corrected family-wise for cluster size at $P < 0.05$.

Multivariate Searchlight Regression: Neural Correlates of Choices and of the Decision Variable

Having identified a subcortical region carrying a general confidence signal, we next searched for brain regions carrying representations of choices and of the DV. We restricted the search for the DV to regions encoding choices while controlling for motion direction to ensure that our results of the DV truly reflected decision-related information rather than the motion direction presented on the screen. Choice-related information was found mostly in prefrontal and posterior parietal regions (Fig. 5*A*), with a mean accuracy across all voxels of 56.83% (± 1.59 SD). Additional analyses confirmed that these regions were choice-related rather than only reflecting the correctness of the subjects (see Supplementary Analyses for additional control analyses testing for further alternative explanations than choice-related information). Within these regions, we searched for the representation of the DV.

Information about the DV was found in activity patterns in several brain regions, most prominently the left middle frontal gyrus, the left posterior parietal cortex including the precuneus, the right ventrolateral prefrontal cortex encompassing inferior frontal cortex and the anterior insula, and the left middle

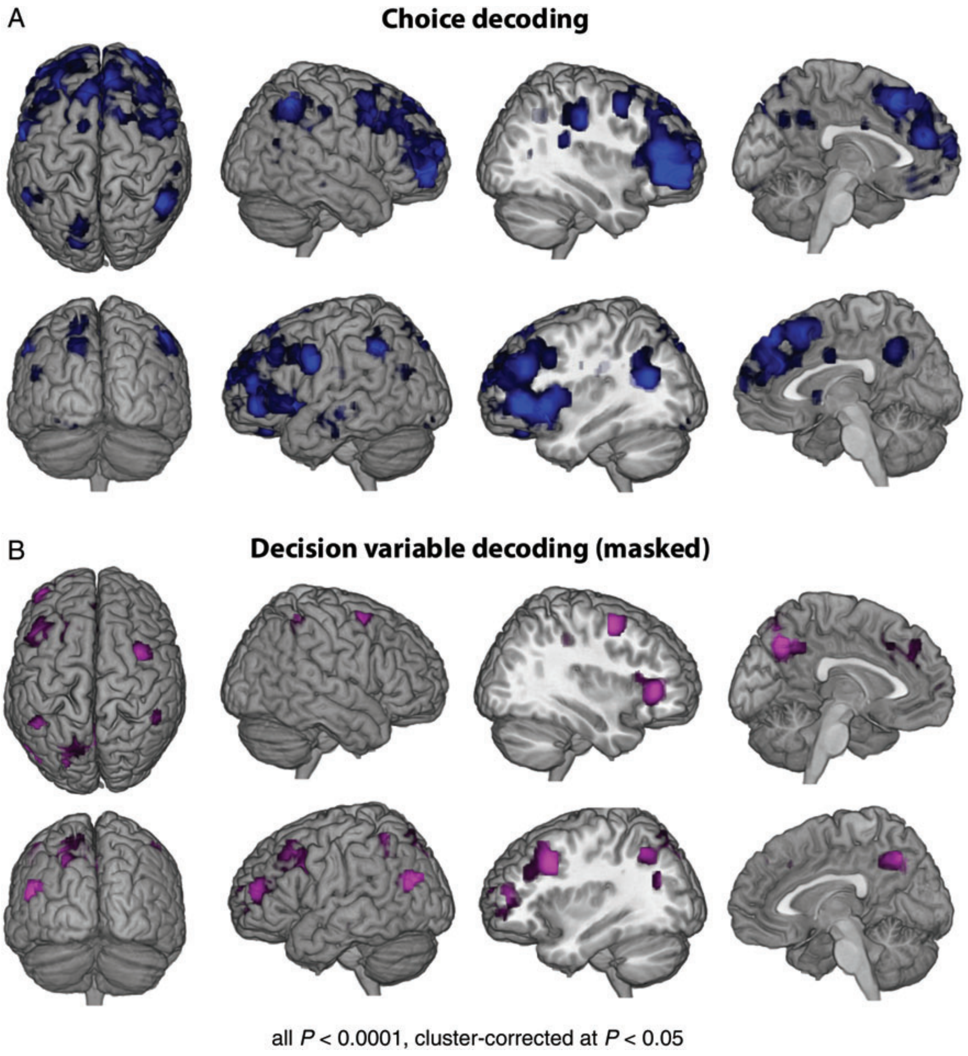
A
Choice decoding
B
Decision variable decoding (masked)
all $P < 0.0001$, cluster-corrected at $P < 0.05$

Figure 5. (*A*) Brain regions encoding the choice independent of the direction of motion of dots (Britten et al. 1996). (*B*) Limited to these regions signals representing the DV as defined by the choices of subjects and the associated confidence.

temporal gyrus (Fig. 5B, for a complete list, see Table 2, for unmasked results, see Supplementary Fig. 3). It should be noted that due to the response-mapping cue, choices were decorrelated from motor responses: therefore, the DV was represented independent of motor responses (Hebart et al. 2012). Also the criterion that possibly varied slightly between the highly trained subjects could not affect these results, because it represents only a constant that would leave the results of our multivariate analyses unchanged (see Formula [3] in Fig. 1D). Since response time showed no linear relationship with the DV, we did not run a similar control analysis as was done for perceptual confidence (see previous section).

Table 2
Brain regions informative about the DV

Region	*X*	*Y*	*Z*	Mean fisher *z*	*Z*-value	Cluster size
Left middle frontal gyrus	−36	18	36	0.37	4.82	964
Left superior frontal gyrus	−18	60	9	0.32	4.51	
Superior medial frontal lobe	−12	36	36	0.31	4.47	
Precuneus	−6	−63	42	0.39	5.27	454
Left superior parietal lobe	−27	−69	48	0.37	4.70	
Left inferior parietal lobe	−42	−51	51	0.43	4.64	
Right ventrolateral prefrontal cortex (inferior frontal gyrus /right anterior insula)	33	36	3	0.32	4.87	119
Left Middle temporal gyrus (posterior)	−51	−72	21	0.31	5.34	79
Right inferior parietal lobe	45	−45	57	0.29	4.33	71
Right middle frontal gyrus	33	3	54	0.39	4.91	67
Left inferior frontal gyrus /left orbitofrontal gyrus	−18	33	−9	0.30	4.39	60
Left middle temporal gyrus (central)	−51	−27	−9	0.38	4.47	45

Note: Brain regions carrying information about the DV reconstructed from choices and confidence ratings. Mean Fisher *z*-values reflect the *z*-transformed size of the correlation between predicted and true values in the multivariate regression analysis (see Materials and Methods), averaged across subjects. The coordinates, mean Fisher *z*-values, and statistical *z*-values of the normal distribution refer to the peak within each cluster. For clusters spanning several brain regions, multiple peaks are shown. Results are reported at $P < 0.0001$, corrected family-wise for cluster size at $P < 0.05$.

Interdependence between Neural Representation of Decision Variable and Confidence

We investigated which of the brain regions informative about the DV (Fig. 5) predicted the confidence-associated signal in the ventral striatum (Fig. 4). Only the activity patterns in the right ventrolateral prefrontal cortex exhibited the specific correlation with the activity patterns in the ventral striatum (see Fig. 6, for raw Fisher *z*-values, see Supplementary Fig. 4) as expected for the transformation from DV to confidence (Fig. 1C, D). Please note that this analysis accounted for across-run average correlations between DV and confidence.

It is possible that a common input drove both signals in vlPFC and the ventral striatum, rather than these two regions interacting directly. In this scenario, sensory evidence from motion-sensitive cortex should feed into vlPFC, and the same sensory evidence would be transformed to confidence in the ventral striatum. To test for this, we took the signals from motion-sensitive area MT+ and ran the same multivariate connectivity analysis as mentioned above, but this time between MT+ and vlPFC as well as between MT+ and the ventral striatum. According to this view, the predicted labels in MT+ would correlate with the predicted labels in vlPFC, and the transformed predicted labels in MT+ would correlate with the predicted labels of confidence in the ventral striatum. We found a correlation between MT+ and vlPFC (Fisher $z = 0.18$, permutation $P = 0.0100$), but not between MT+ and the ventral striatum (Fisher $z = -0.04$, permutation $P = 0.5793$). This result is in line with the idea that a DV in vlPFC is constructed from signals in motion-sensitive cortex and confirms the suggestion that this DV is transformed to a confidence signal in the ventral striatum, rather than both signals being mediated by sensory evidence in MT+.

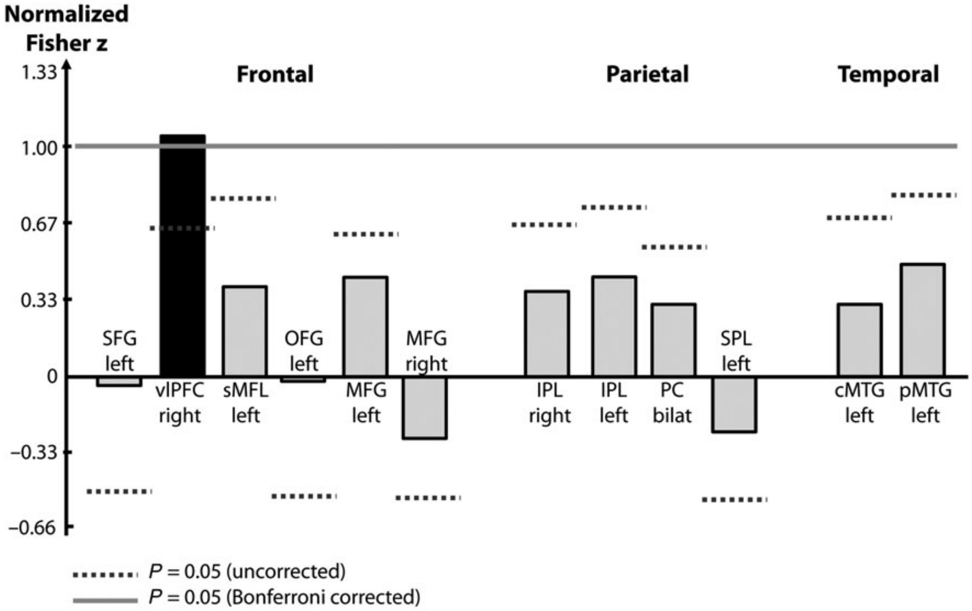

Figure 6. Normalized multivariate connectivity between several seed regions and the ventral striatum as target region. Normalization was carried out by dividing the Fisher *z*-value by the significance cutoff of the permutation test, but preserving the sign. A value exceeding 1 (solid line) would be deemed significant at the Bonferroni corrected level. MFG, middle frontal gyrus; SFG, superior frontal gyrus; sMFL, superior medial frontal lobe; PC, precuneus; SPL, superior parietal lobe; IPL, inferior parietal lobe; vlPFC, ventrolateral prefrontal cortex; pMTG, posterior middle temporal gyrus; OFG, orbitofrontal gyrus; cMTG, central middle temporal gyrus.

Although our behavioral results showed no significant correlation between DV and confidence, a possible concern is that a brain region could be biased to selectively respond only to one choice (e.g., "motion up") and not the other. In the same manner as the behavioral results, this could explain a positive correlation between DV and confidence. To test for this possibility, we looked up the labels predicted by the SVR for the DV, transformed them according to the assumed relationship between DV and confidence (Fig. 1*C*), and correlated these transformed labels with those that had been predicted for confidence in the same region. This analysis revealed no significant positive correlation in vlPFC (mean Fisher z: 0.09, $t_{(17)} = 0.8470$, $P = 0.4088$). Finally, to test whether such a bias was present anywhere in the brain, we repeated the same procedure using a searchlight approach and the same threshold applied to all whole-brain analyses. The analysis revealed no significant voxels, demonstrating that this alternative cannot explain the results of the multivariate connectivity analysis.

Discussion

It has long been postulated that human observers use an internal DV reflecting the amount of evidence in favor of one perceptual interpretation over another for determining their binary choice and their absolute level of confidence in that choice (Peirce and Jastrow 1884; Vickers 1979; Baranski and Petrusic 1998). The former requires a comparison of a continuous DV with a decision criterion, whereas the latter requires a transformation of the DV to the estimation of the probability of being correct, independent of the type of choice. Here, we demonstrate that activity in the ventral striatum increased with the level of confidence and that activity patterns in a wide network of cortical brain regions reflected the perceptual DV that we reconstructed from choices and confidence ratings according to signal detection theory. The DV-related activity patterns specifically in the right ventrolateral prefrontal cortex co-varied with the confidence-related activity patterns in the ventral striatum, suggesting a mechanistic link for the transformation of neural representations of the DV into neural representations of confidence.

Our findings complement recent studies on perceptual confidence in animals. An estimate of confidence that depended on the choice was found in macaque lateral intraparietal area (Kiani and Shadlen 2009), and single unit activity in rat orbitofrontal cortex reflected a more general perceptual confidence signal (Kepecs et al. 2008), which has also been suggested for macaque striatum (Ding and Gold 2012). We extend these findings by demonstrating a specific relationship between these two types of decision-related neural signals. Specifically, our results suggest that the transformation of DV to confidence entails a non-selective pooling across responses of choice-selective populations of prefrontal neurons and that this pooling takes place in the ventral striatum: the bigger the activity levels of "any" of the choice-selective prefrontal populations, the bigger the response of the striatal neurons encoding choice-independent confidence.

For identifying the DV, we were careful to focus only on brain regions that encoded choices while controlling for the stimulus shown (Li et al. 2009; Hebart et al. 2012). In addition, we kept sensory evidence constant for each participant to ensure that we would not confuse representations of the stimulus with representations of the DV. Activity patterns in a number of predominantly frontal and parietal brain regions were consistent with a representation of the DV (Kim and Shadlen 1999; Gold and Shadlen 2000; Shadlen and Newsome 2001; Roitman and Shadlen 2002). The ventrolateral prefrontal cortex, exhibiting the specific covariation with the ventral striatum reported in our study, has previously been shown to encode perceptual choices in studies using multivariate pattern analysis (Pessoa and Padmala 2007; Li et al. 2009) and shows a direct anatomical connection to regions of the basal ganglia (Aron et al. 2007). Outside of the signal detection theory framework, these patterns of activity can reflect the amount of sensory evidence available to the subject, the accumulated sensory evidence, or a high-level monitoring signal keeping track of the accumulated evidence. Different or additional brain regions may be involved in encoding this DVs when choices are not by design decoupled from motor plans as in the present study (Bennur and Gold 2011; Hebart et al. 2012; O'Connell et al. 2012; de Lange et al. 2013; Filimon et al. 2013).

The ventral striatum exhibited a robust positive correlation with confidence and was the primary candidate region identified in our study for encoding perceptual confidence. The ventral striatum seems to play a general role in encoding decision certainty (Preuschoff et al. 2006), reward prediction error (O'Doherty et al. 2004), motivational salience (Zink et al. 2004), and motivation in general (Talmi et al. 2008). While speculative, this signal could potentially correspond to the rewarding feeling associated with being confident, which could be used to reinforce behaving in a similar manner again (e.g., paying attention to the stimulus). Perceptual confidence signals encoded in this region could be used to evaluate present and fine-tune future choices (Ding and Gold 2012) even in the absence of feedback as was the case in our study, which is consistent with recent findings demonstrating a confidence prediction error signal in the ventral striatum using a similar perceptual discrimination task (Daniel and Pollmann 2012). Expanding on our findings, we have recently conducted a perceptual learning study in which we showed that subjects whose striatal responses are modulated more strongly by confidence exhibit better learning, even in the absence of external feedback (unpublished data). In that respect, the internal monitoring of the quality of evidence and this self-generated feedback may also participate in more general forms of learning and would be particularly useful in contexts where external feedback is not available. In order for a confidence signal to affect other neuronal responses or behavior irrespective of the choice, it requires an explicit representation, rather than an implicit representation in the form of an unsigned DV.

The ventral striatum is also anatomically closely connected to the orbitofrontal cortex, and it has been reported in rats that the latter region may also carry a general confidence signal (Kepecs et al. 2008). The absence of an effect in orbitofrontal cortex in our study may be due to one of the following factors: possibly reduced or distorted fMRI signal due to air-tissue boundaries in the nearby sinuses (Ojemann et al. 1997), the spatial proximity of neuronal signals positively "and" negatively correlated with confidence that cannot be resolved with the spatial resolution of fMRI (Kepecs et al. 2008), or the encoding of confidence as an outcome signal in the OFC only when explicit feedback is provided (Mainen and Kepecs 2009).

The brain might also compute signals encoding that the expected outcome of the behavior is negative, that is reflecting uncertainty about the previous choice (Bach and Dolan 2012). Uncertainty has been investigated for both value-based (Hsu et al. 2005; Huettel et al. 2005) and perceptual or categorical decision-making (Fleck et al. 2006; Grinband et al. 2006; Fleming et al. 2012). One study identified human brain regions, in which activity correlated with confidence ratings (Fleming et al. 2012); of those regions, the anterior cingulate and the rostrolateral prefrontal cortex showed a negative correlation of perceptual confidence and the BOLD response, in line with a neural representation of perceptual uncertainty. The regions in which activity correlated negatively with confidence in our data are also involved in cognitive control and top-down attention (Miller and Cohen 2001; Corbetta and Shulman 2002; Botvinick et al. 2004). This might suggest that uncertainty leads to brain activations signaling the need to adapt future behavior. At the same time, we showed that activity in many of these regions showing a negative relationship with confidence might at least in part be explained by variations in response time. This indicates that either some of these brain responses were not directly related to confidence or that the effects of confidence and response time could not be distinguished, which makes it difficult if not impossible to uniquely attribute them to confidence. Whether these signals directly reflect representations of uncertainty, top-down control processes that follow from uncertainty, or other processes that contribute to confidence is a prospect for future studies. Future studies might also elucidate whether the relationship between DV and confidence signals can also be demonstrated between DV and uncertainty signals.

We defined the DV as the amount of information available to the observer in favor of a particular choice. This definition of the DV differs slightly from other uses of this term where it is equated with the temporal accumulation of evidence (e.g., Kiani and Shadlen 2009). The model derived from signal detection theory used here is agnostic to the temporal evolution of this signal and is limited to the end product of this purported accumulation process. In addition, we did not operationalize the DV independent of confidence based on neuronal response profiles, as was done for example in Kiani and Shadlen (2009). However, it was not the goal of the present study to confirm signal detection theory as a model that can link choices, the DV and confidence. Rather, we merely used this model to reconstruct the hypothetical DV from choices and confidence and link the representation of this variable to the representation of confidence. In that way, the reconstructed DV in this study could alternatively be described as "signed confidence," referring to the amount of confidence associated with a given choice.

While the framework that guided our study is well supported by a wealth of behavioral and neuronal data, we do not claim that signal detection theory is the only model to explain the relationship between the DV and confidence (Zylberberg et al. 2012). In addition, our findings do not imply that the DV is the "only" factor that governs confidence. Indeed, it has been reported that objective decision performance and subjective judgments can (at least partially) be decoupled by means of additional processes, which may or may not depend directly on the evidence used for the decision (Lau and Passingham 2006; Persaud et al. 2007; Hesselmann et al. 2011). Additional evidence-dependent processes after the choice (Resulaj et al. 2009) that could affect the confidence level (Pleskac and Busemeyer 2010) are unlikely, because our interrogation task entailed an imposed 1-s delay between stimulus offset and response prompt. Any additional sensory processing would have been completed by the time of the choice (Roitman and Shadlen 2002). However, "evidence-independent" processes during and after the choice may have contributed to the confidence ratings, and indeed the relationship between performance and confidence in the present study was not perfect (Fig. 2*C*). Such additional processes may not only translate the DV into confidence judgments but encompass additional dedicated decision-making mechanisms evaluating the quantity and quality of evidence (Fleming and Dolan 2012; Yeung and Summerfield 2012). However, a non-perfect relationship of behavioral performance and confidence could also be explained by different sources of noise, for example, in the transformation process of DV to confidence, or by a decay of the DV in the time window between the choice and confidence ratings. Such noise sources would only weaken the relationship between decision performance and confidence ratings but do not disagree with the suggested transformation process of DV to confidence.

The approach employed in this study could be used to investigate related questions: To what degree do the neural computations underlying perceptual confidence generalize to other forms of confidence, such as categorical confidence (Grinband et al. 2006), memory confidence (Henson et al. 2000) or value-based confidence (De Martino et al. 2013), and to what degree do they differ (Baird et al. 2013)? The temporal separation between the choice and confidence rating could be varied, thus providing stronger variation in post-decision time, which could be used to specifically target these post-decisional processes. Finally, in a reaction time version of the task, our approach could be generalized to incorporate the influence of variability in response criteria on confidence judgments (Ratcliff and Starns 2009; Pleskac and Busemeyer 2010). Targeting the relationship between the DV and confidence as has been done in the present study can be seen as an important step toward a more complete understanding of the neuronal processes underlying human perceptual choice behavior.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

This work was supported by the German Research Foundation (DFG Grant HA 5336/1-1), the Bernstein Computational Neuroscience Program of the German Federal Ministry of Education and Research (BMBF Grants 01GQ0411, 01GQ1001C), and the Excellency Initiative of the German Federal Ministry of Education and Research (DFG Grant GSC86/1-2009). MNH was supported by the German National Merit Foundation (Studienstiftung des deutschen Volkes) and the Berlin School of Mind and Brain.

Notes

Conflict of Interest: None declared

References

Allefeld C, Haynes J-D. 2014. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. NeuroImage. 89:345–357.

Aron AR, Behrens TE, Smith S, Frank MJ, Poldrack RA. 2007. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. J Neurosci. 27:3743–3752.

Ashburner J. 2007. A fast diffeomorphic image registration algorithm. Neuroimage. 38:95–113.

Bach DR, Dolan RJ. 2012. Knowing how much you don't know: a neural organization of uncertainty estimates. Nat Rev Neurosci. 13:572–586.

Baird B, Smaalwood J, Gorgolewski KJ, Margulies DS. 2013. Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. J Neurosci. 33:16657–16665.

Baranski JV, Petrusic WM. 1998. Probing the locus of confidence judgments: experiments on the time to determine confidence. J Exp Psychol Hum Percept Perform. 24:929–945.

Bennur S, Gold JI. 2011. Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. J Neurosci. 31:913–921.

Botvinick MM, Cohen JD, Carter CS. 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci. 8:539–546.

Braddick O. 1974. A short-range process in apparent motion. Vision Res. 14:519–527.

Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA. 1996. A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci. 13:87–100.

Chang CC, Lin CJ. 2011. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol. 2:1–27.

Christophel TB, Hebart MN, Haynes J-D. 2012. Decoding the contents of visual short-term memory from human visual and parietal cortex. J Neurosci. 32:12983–12989.

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci. 3:201–215.

Daniel R, Pollmann S. 2012. Striatal activations signal prediction errors on confidence in the absence of external feedback. Neuroimage. 59:3457–3467.

de Lange FP, Rahnev DA, Donner TH, Lau H. 2013. Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. J Neurosci. 33:1400–1410.

De Martino B, Fleming SM, Garrett N, Dolan RJ. 2013. Confidence in value-based choice. Nat Neurosci. 16:105–110.

Ding L, Gold JI. 2012. Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. Cereb Cortex. 22:1052–1067.

Filimon F, Philiastides MG, Nelson JD, Kloosterman NA, Heekeren HR. 2013. How embodied is perceptual decision making? Evidence for separate processing of perceptual and motor decisions. J Neurosci. 33:2121–2136.

Fleck MS, Daselaar SM, Dobbins IG, Cabeza R. 2006. Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. Cereb Cortex. 16:1623–1630.

Fleming SM, Dolan RJ. 2012. The neural basis of metacognitive ability. Philos Trans R Soc B Biol Sci. 367:1338–1349.

Fleming SM, Huijgen J, Dolan RJ. 2012. Prefrontal contributions to metacognition in perceptual decision making. J Neurosci. 32:6117–6125.

Gold JI, Shadlen MN. 2007. The neural basis of decision making. Annu Rev Neurosci. 30:535–574.

Gold JI, Shadlen MN. 2000. Representation of a perceptual decision in developing oculomotor commands. Nature. 404:390–394.

Görgen K, Hebart M, Haynes JD. 2012. The Decoding Toolbox (TDT): A new fMRI analysis package for SPM and Matlab. Poster presented at 18th Annual Meeting of the Organization for Human Brain Mapping (OHBM) 2012 10–14 Jun 2012:378MT.

Green DM, Swets JA. 1966. Signal detection theory and psychophysics. New York: Wiley.

Grinband J, Hirsch J, Ferrera VP. 2006. A neural representation of categorization uncertainty in the human brain. Neuron. 49: 757–763.

Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. 2007. Reading hidden intentions in the human brain. Curr Biol. 17:323–328.

Hebart MN, Donner TH, Haynes J-D. 2012. Human visual and parietal cortex encode visual choices independent of motor plans. Neuroimage. 63:1393–1403.

Heekeren HR, Marrett S, Ungerleider LG. 2008. The neural systems that mediate human perceptual decision making. Nat Rev Neurosci. 9:467–479.

Henson RNA, Rugg MD, Shallice T, Dolan RJ. 2000. Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. J Cogn Neurosci. 12:913–923.

Hesselmann G, Hebart M, Malach R. 2011. Differential BOLD activity associated with subjective and objective reports during "blindsight" in normal observers. J Neurosci. 31:12936–12944.

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF. 2005. Neural systems responding to degrees of uncertainty in human decision-making. Science. 310:1680–1683.

Huettel SA, Song AW, McCarthy G. 2005. Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci. 25:3304–3311.

Jimura K, Poldrack RA. 2012. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia. 50:544–552.

Kahnt T, Grueschow M, Speck O, Haynes J-D. 2011. Perceptual learning and decision-making in human medial frontal cortex. Neuron. 70:549–559.

Kahnt T, Heinzle J, Park SQ, Haynes J-D. 2010. The neural code of reward anticipation in human orbitofrontal cortex. Proc Natl Acad Sci USA. 107:6010–6015.

Kepecs A, Uchida N, Zariwala HA, Mainen ZF. 2008. Neural correlates, computation and behavioural impact of decision confidence. Nature. 455:227–231.

Kiani R, Shadlen MN. 2009. Representation of confidence associated with a decision by neurons in the parietal Cortex. Science. 324:759–764.

Kim J-N, Shadlen MN. 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci. 2:176–185.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci USA. 103:3863–3868.

Ku SP, Gretton A, Macke J, Logothetis NK. 2008. Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys. Magn Reson Imaging. 26:1007–1014.

Lau HC, Passingham RE. 2006. Relative blindsight in normal observers and the neural correlate of visual consciousness. Proc Natl Acad Sci USA. 103:18763–18768.

Li S, Mayhew S, Kourtzi Z. 2009. Learning shapes the representation of behavioral choice in the human brain. Neuron. 62:441–452.

Macmillan NA, Creelman CD. 2005. Detection Theory: A User's Guide. Mahwah, NJ: Erlbaum.

Mainen ZF, Kepecs A. 2009. Neural representation of behavioral outcomes in the orbitofrontal cortex. Curr Opin Neurobiol. 19:84–91.

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 24:167–202.

Naselaris T, Kay KN, Nishimoto S, Gallant JL. 2011. Encoding and decoding in fMRI. Neuroimage. 56:400–410.

O'Connell RG, Dockree PM, Kelly SP. 2012. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat Neurosci. 15:1729–1735.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science. 304:452–454.

Ojemann JG, Akbudak E, Snyder AZ, McKinstry RC, Raichle ME, Conturo TE. 1997. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. Neuroimage. 6:156–167.

Peirce CS, Jastrow J. 1884. On small differences of sensation. Mem Natl Acad Sci. 3:75–83.

Persaud N, McLeod P, Cowey A. 2007. Post-decision wagering objectively measures awareness. Nat Neurosci. 10:257–261.

Pessoa L, Padmala S. 2007. Decoding near-threshold perception of fear from distributed single-trial brain activation. Cereb Cortex. 17:691–701.

Pleskac TJ, Busemeyer JR. 2010. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychol Rev. 117:864–901.

Poldrack RA, Mumford JA, Nichols TE. 2011. Handbook of Functional MRI Data Analysis. Cambridge, UK: Cambridge University Press.

Preuschoff K, Bossaerts P, Quartz SR. 2006. Neural differentiation of expected reward and risk in human subcortical structures. Neuron. 51:381–390.

Ratcliff R, Starns JJ. 2009. Modeling confidence and response time in recognition memory. Psychol Rev. 116:59–83.

Resulaj A, Kiani R, Wolpert DM, Shadlen MN. 2009. Changes of mind in decision-making. Nature. 461:263–266.

Roitman JD, Shadlen MN. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci. 22:9475–9489.

Shadlen MN, Newsome WT. 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol. 86:1916–1936.

Talmi D, Seymour B, Dayan P, Dolan RJ. 2008. Human Pavlovian-instrumental transfer. J Neurosci. 28:360–368.

Tootell RBH, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci. 15:3215–3230.

Vickers D. 1979. Decision Processes in Visual Perception. New York: Academic Press.

Watson AB, Pelli DG. 1983. QUEST: a Bayesian adaptive psychometric method. Percept Psychophys. 33:113–120.

Yeung N, Summerfield C. 2012. Metacognition in human decision-making: confidence and error monitoring. Philos Trans R Soc B Biol Sci. 367:1310–1321.

Zink CF, Pagnoni G, Martin-Skurski ME, Chappelow JC, Berns GS. 2004. Human striatal responses to monetary reward depend on saliency. Neuron. 42:509–517.

Zylberberg A, Barttfeld P, Sigman M. 2012. The construction of confidence in a perceptual decision. Front Integ Neurosci. 6:79.