Cortical activity is more stable when sensory stimuli are consciously perceived

Aaron Schurger[a,b,c,1], Ioannis Sarigiannidis[a,b,d], Lionel Naccache[e,f,g,h], Jacobo D. Sitt[a,b,e,f,g], and Stanislas Dehaene[a,b,d,i]

[a]Cognitive Neuroimaging Unit (U992), INSERM, Gif-sur-Yvette 91191, France; [b]Direction des Sciences du Vivant, I2BM, NeuroSpin Center, Commissariat à l'Energie Atomique, Gif sur Yvette 91191, France; [c]Laboratory of Cognitive Neuroscience, Department of Life Sciences, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland; [d]Université Paris-Sud 11, 91405 Orsay, France; [e]Institut du Cerveau et de la Moelle Épinière, INSERM U1127, 75013 Paris, France; [f]Institut du Cerveau et de la Moelle Épinière, CNRS UMR 7225, 75013 Paris, France; [g]Sorbonne Universités, Université Pierre et Marie Curie, Université de Paris 6, 75013 Paris, France; [h]Department of Neurophysiology, Groupe Hospitalier Pitié-Salpêtrière, 75013 Paris, France; and [i]Collège de France, 75005 Paris, France

Edited by Michael E. Goldberg, Columbia University College of Physicians and Surgeons and the New York State Psychiatric Institute, New York, NY, and approved March 11, 2015 (received for review October 14, 2014)

According to recent evidence, stimulus-tuned neurons in the cerebral cortex exhibit reduced variability in firing rate across trials, after the onset of a stimulus. However, in order for a reduction in variability to be directly relevant to perception and behavior, it must be realized within trial—the pattern of activity must be relatively stable. Stability is characteristic of decision states in recurrent attractor networks, and its possible relevance to conscious perception has been suggested by theorists. However, it is difficult to measure on the within-trial time scales and broadly distributed spatial scales relevant to perception. We recorded simultaneous magneto- and electroencephalography (MEG and EEG) data while subjects observed threshold-level visual stimuli. Pattern-similarity analyses applied to the data from MEG gradiometers uncovered a pronounced decrease in variability across trials after stimulus onset, consistent with previous single-unit data. This was followed by a significant divergence in variability depending upon subjective report (seen/unseen), with seen trials exhibiting less variability. Applying the same analysis across time, within trial, we found that the latter effect coincided in time with a difference in the stability of the pattern of activity. Stability alone could be used to classify data from individual trials as "seen" or "unseen." The same metric applied to EEG data from patients with disorders of consciousness exposed to auditory stimuli diverged parametrically according to clinically diagnosed level of consciousness. Differences in signal strength could not account for these results. Conscious perception may involve the transient stabilization of distributed cortical networks, corresponding to a global brain-scale decision.

correlated variability | consciousness | dynamical systems | pattern similarity | directional variance

Recent evidence suggests that population neuronal activity in stimulus-tuned regions of cortex is reliably drawn to the same region of state space in response to a stimulus (1). Churchland et al. (1) analyzed recordings from many different areas of monkey cortex, under a variety of different task conditions, and found a common decrease in intertrial firing-rate variability beginning just after stimulus onset. At the macroscopic level, intertrial variability of the pattern of activity, measured using functional MRI (fMRI), is reduced immediately following stimulus onset (2) and has been shown to correlate with perception (3) and explicit recognition memory (4). However, to be relevant for decision making and behavior, a reduction in variability must be realized within trial, in the form of a stable pattern of activity.

Both stability (within episode) and reproducibility (between episodes) are characteristic of decision states in recurrent attractor networks: When presented with a learned input pattern of sufficient intensity, the activity within the network will evolve from an arbitrary initial state toward a relatively low-energy stable state that can be read out as the network's decision (completion, recognition, or grouping) given the input (5, 6). Theorists have proposed that recurrent network dynamics in general, and stability in particular, might be relevant to conscious perception (7–11), with nonlocal recurrent interactions selecting the information that becomes consciously perceived (12–14). The dynamic selection of a stable global network would tend to further stabilize local populations via feedback connections. Thus, we predicted that the global pattern of neural activity in response to a threshold-level sensory stimulus will be more stable when subjects report having a conscious percept, with potentially a slight decrease in the net global level of activity. A prior computational model (9, 15) predicted that a period of stability should appear immediately following the onset of the late-positive potential (LPP, often referred to as the P300), and lasting ∼200–300 ms.

We recorded simultaneous magnetoencephalography (MEG) and electroencephalography (EEG) measurements while subjects performed a category discrimination task on threshold-level visual stimuli rendered nearly invisible by dichoptic color fusion. We used a measure of angular dispersion over the entire array of MEG gradiometers to test our prediction that the global pattern of neuronal activity evoked by a stimulus will tend to be more stable when the subject reports having consciously seen the stimulus.

In addition to covarying with subjective reports in response to threshold-level stimulation, a good signature of conscious perception should also covary with conscious state (e.g., wakefulness versus deep sleep) in response to suprathreshold stimulation. Thus, to test the generality of our hypothesis we applied our analysis to independent EEG data from a large cohort of patients with disorders of consciousness (DOC), who were exposed to auditory stimuli (16–18). This analysis also tests whether stability, as a signature of conscious perception, generalizes to a dif-

Significance

Recently, a reduction in the variability of neural activity across trials has been proposed as a general property of sensory perception. However, in order for this increased reliability in the neuronal response to be relevant for perception it must be present at the single-trial level. Here we show that the within-trial stability of the activity evoked by a threshold-level visual stimulus is a reliable and specific indicator of whether or not the stimulus was reported as "seen." This within-trial difference in stability coincides in time with a difference in variability across trials. We also show that the same neural stability can be used to discriminate the conscious state of brain-injured patients. These findings validate the relevance of transient neural stability for conscious perception.

Author contributions: A.S. and S.D. designed research; A.S. performed research; A.S. contributed new reagents/analytic tools; A.S., I.S., J.D.S., and S.D. analyzed data; and A.S., I.S., L.N., J.D.S., and S.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: aaron.schurger@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1418730112/-/DCSupplemental.

ferent sensory modality (audition) and links research on conscious state (e.g., coma versus awake) with research on conscious perception (e.g., seen versus unseen stimuli in an awake individual).

Measuring the Variability of Patterns of Activity

A given macroscopic pattern of neural activity in the brain will produce a distinct, albeit noisy, spatial pattern of activity across an array of M/EEG sensors. Spatiotemporal variability in the underlying pattern of neural activity will result in variability in the mutivariate pattern of activity at the sensor level. There are a number of ways that one might measure variability in a multivariate (multidimensional) space, and all are essentially measures of representational similarity (19): *t*he degree to which the different patterns resemble one another. Treating each sample across N sensors as a vector in an N-dimensional space, similarity can be computed as the degree to which a set of such vectors are all pointing in the same direction (Fig. 1), called the "directional variance" [or "circular variance" (20)]: With the norm (or length) of each vector set to 1, the norm of the mean of the vectors gives the degree to which the vectors are all pointing in the same direction, called the directional coherence ($\overline{R}$, where $0 < \overline{R} < 1$). This gives the degree to which the associated patterns collectively resemble one another. The directional variance is simply 1 minus the directional coherence ($1 - \overline{R}$). "Similarly to the variance of linear data, the smaller the value of the circular variance, the more concentrated the distribution" (ref. 20, p. 32).

A set of vectors pointing in the same direction, even if of different lengths, are all considered to have the same pattern, because the correlation coefficient between any two of them will be 1.0. A stable pattern is one that remains the same over time. A reproducible pattern is one that is the same every time it is instantiated (i.e., across trials) (3). Directional variance is computed in the same way in either case (Fig. 1*B*). We use the label *dva* to stand for directional variance, and 1 – *dva* is the directional coherence (*stability* within trial or *reproducibility* between trials). Note that stability/reproducibility is high when *dva* is low. We report directional variance (*dva*) in all figures, rather than stability or reproducibility (1 – *dva*), adopting the same convention as prior work using the Fano factor (1), which is also a measure of dispersion.

One important caveat regarding *dva* (and possibly all measures of representational similarity) concerns the role of noise. Two sets of equally stable vectors will score differently if one set is weak and the other strong relative to the noise (called the signal-to-noise ratio, or SNR): For very weak patterns, even if perfectly stable, the angle of the vectors will be mostly determined by the noise. Unfortunately, the relationship between *dva* and SNR is highly nonlinear, especially in the range where the signal amplitude is not much larger than the noise amplitude (Fig. 1*C*, *Inset*), as is the case for EEG, MEG, and fMRI. So, in this context, as in ref. 1, differences in signal amplitude are a potential confound. We address this in two ways. One is to measure the average norm of the vectors and make sure that lower variance is not associated with a larger norm. The other is to apply the mean-matching procedure described by ref. 1 to the vector norms (*Materials and Methods*) and verify that the result holds even when the norms are made to be equal.

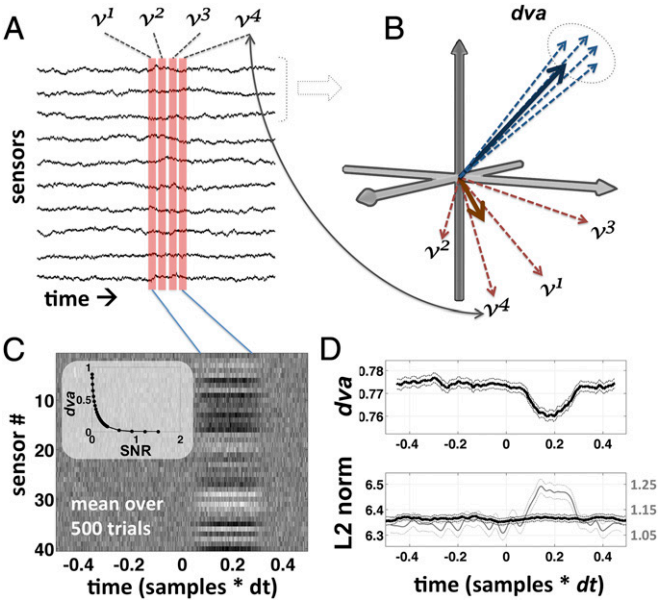

Fig. 1. Schematic illustration of *dva*. *dva* is explained graphically in *A* and *B*; *C* and *D* show the results of a simulation used to illustrate its properties. v^1, v^2, v^3, and v^4 are samples taken at four successive time steps in a single trial epoch, across a multichannel sensor array (*A*; time is on the horizontal axis and sensor on the vertical). v^1, …, v^4 can also be treated as vectors in n dimensions, where each element (dimension) carries a measurement from one of the n channels (*B*). *dva* is a measure of dispersion in the directionality of the vectors (dashed circle in *B*). *dva* can be computed for any number of channels, but in *B* we illustrate a hypothetical subset of three channels. *dva* is mathematically independent of the length of the vectors (L2 norm or spatial power), and depends only on their orientation with respect to one another. According to our hypothesis successive patterns of activity on seen trials, within a certain window of time, are more like the blue vectors—consistently pointing in the same direction. (*C*) The average over 500 simulated trials where a random stable pattern emerges from 100 to 300 ms (with time on the horizontal axis and sensor on the vertical; *Materials and Methods*). (*D*) The average *dva* and L2 norm over this set of simulated trials. Note that *dva* is sensitive to the presence of the pattern even though there is no difference in the mean norm in this simulation (*Materials and Methods*). Note also that the total power, or norm, of the mean (i.e., the norm over each column in *C*; gray line in *D*) is not equivalent to the mean norm (black line in *D*). This can explain why conscious perceptions is commonly associated with larger-amplitude evoked potentials: Trial averaging highlights stable/reliable patterns and suppresses unstable/unreliable ones, even if the spatial energy is the same on single trials. The relationship between *dva* and the SNR is highly nonlinear (*C*, *Inset*), which must be taken in to account when analyzing the data.

Results

Summary of Experiment and Task. Stimuli were line drawings of faces and houses, rendered difficult to see by presenting them dichoptically in two opposing isoluminant colors (dichoptic color masking; *SI Appendix*, Fig. S1). Visibility was manipulated by varying the color contrast over a range of five different perithreshold levels randomly interleaved across trials. Stimuli were always presented with the opposite color assignment in the two eyes. One-third of stimuli were uniform color patches with no object, subjectively indistinguishable from low-contrast objects. Subjects were not told about the control stimuli. Stimulus duration was 50 ms, and then at 1,000 ms the fixation point flickered, cueing the subject to respond. The task was to discriminate the object category (face/house), guessing if necessary, and then to report whether or not the object had been "seen." Subjects were explicitly instructed to respond "unseen" only if they saw nothing at all inside the yellow square, and to respond "seen" even if what they saw was not clearly recognized. This choice of task should not be taken to imply an assumption that perception is dichotomous in the brain. All seen-versus-unseen data analyses were restricted to a single threshold level of color contrast chosen individually for each subject, summarized in the horizontal bar at the bottom of *SI Appendix*, Fig. S2*A*.

Behavioral Responses. *SI Appendix*, Fig. S2 summarizes the behavioral results. Mean accuracy on the face–house identification task (*SI Appendix*, Fig. S2*A*; computed exclusive of control trials), from lowest to highest color contrast, ranged from chance level (53.8% correct, $P > 0.5$) to near ceiling (91.0% correct, $P << 0.01$). The mean proportion of face and house stimuli reported as seen (*SI Appendix*, Fig. S2*B*) remained constant at ~0.3 across contrast levels 1 and 2, and then increased linearly to ~0.8 at contrast level 5. The proportion seen for blank control stimuli

(*SI Appendix*, Fig. S2*E*) was roughly constant across all contrast levels, at ∼0.25, which was not significantly different from the proportion seen for object stimuli at contrast levels 1 and 2 ($P > 0.1$, signed rank test). This suggests that object stimuli at contrast levels 1 and 2 were not perceived differently from blank control stimuli. This latter observation is further supported by the fact that detection *d-prime* (d') (*SI Appendix*, Fig. S2*C*; the sensitivity of seen responses to the presence of an object; *Materials and Methods*) was not significantly different from zero for either contrast level 1 or 2. At contrast level 2, discrimination d' (i.e., for faces versus houses; *SI Appendix*, Fig. S2*D*) was significantly greater than zero ($P < 0.01$, signed rank test; *Materials and Methods*), but the difference between detection and discrimination d' was not significant ($P < 0.1$, signed rank test).

Evoked Responses and Time-Frequency Analyses. We found no significant differences in the evoked responses to the control stimuli across the five levels of color contrast in either the EEG or MEG data (*SI Appendix*, Fig. S3 *A* and *B*). However, when we examined the responses to objects rather than blank control stimuli (*SI Appendix*, Fig. S3 *C* and *D*) we found a clear and significant modulation of the amplitude of the LPP EEG component (400–700 ms after stimulus onset; *SI Appendix*, Fig. S3*C*) as a function of color contrast, and a significant, although not as pronounced, modulation in roughly the same time window in the MEG data (*SI Appendix*, Fig. S3*D*).

We compared the amplitude of the late-positive evoked potential (LPP) on seen and unseen trials. Recall that for all seen–unseen analyses we chose a single, threshold level of color contrast for each subject so that the physical properties of the stimuli were fixed and only the subjective report varied. Consistent with prior studies (21) we found a significant difference in the amplitude of the LPP for seen versus unseen stimuli at parietal EEG electrodes ($P < 0.001$, cluster-based permutation test), from ∼500–700 ms (*SI Appendix*, Fig. S4 *A* and *B*). We found no time point at which the LPP evoked response to blank control stimuli was significantly different from the response to unseen object stimuli ($P > 0.2$ signed-rank test).

Previous studies comparing trials with seen and unseen subjective reports have found that reports of a conscious percept are accompanied by a relative decrease in power in the alpha band and low beta band (22–24) and a relative increase in power in the gamma band (22–26). We replicated these prior observations by performing a time-frequency decomposition on the output of spatial filters derived (for each subject) based on the difference between seen and unseen trials (27) among the MEG gradiometers. This analysis revealed an increase in high-gamma power (∼60–80 Hz), coincident in time with the LPP (at ∼500 ms), and a later suppression of alpha and low-beta power between 500 and 1,000 ms ($P < 0.01$ corrected; *Materials and Methods* and *SI Appendix*, Fig. S4*C*). We also found a decrease in alpha power, within this same time window, in the output of spatial filters derived based on the difference between face and house trials ($P < 0.01$, corrected; *Materials and Methods* and *SI Appendix*, Fig. S4*D*), but with no significant increase in gamma power.

Norm (Overall Intensity) of the Patterns. For both object and control stimuli, where there was a difference in the norm, the norm was greater for unseen stimuli (Figs. 2*E* and 3*E*). This is the opposite of what might indicate an artifactual difference in variability, and thus the results without mean matching are conservative. Note also that for a biologically plausible recurrent network (i.e., with a mix of excitatory and inhibitory connections), convergence toward a stable state predicts a decrease in the overall (free) energy in the network, and hence a decrease in the norm, which is what we observe overall in the trial epochs: an initial small increase in the norm just after stimulus onset, followed by a drop in the norm reaching a minimum at about +450 ms (Figs. 2*E* and 3*E*). For blank control stimuli, the norm was significantly lower on trials where the stimulus provoked an illusory conscious percept (∼600–800 ms; Fig. 3*E*), which is both counterintuitive and consistent with our hypothesis. A similar pattern appears in the patient data (discussed below).

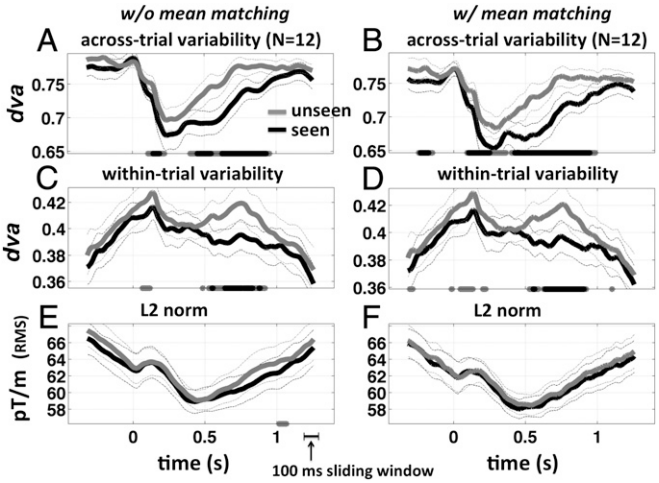

Fig. 2. Reproducibility, stability, and vector norm for object trials. Across-trial and within-trial *dva* and spatial L2 norm, for noncontrol (object) trials, are shown both with (left column) and without (right column) mean matching. (*A*) Across-trial directional variance (1 – reproducibility) as a function of time for unseen (gray) and seen (black) target-present trials. Stars at the bottom of each panel mark time points where the difference unseen – seen is significantly greater than chance (gray, $P < 0.05$; black, $P < 0.01$, both corrected for temporal nonindependence using a cluster-based permutation test). (*C*) Within-trial directional variance (1 – stability). (*E*) L2 norm of the mean vector within the sliding window. (*B*, *D*, and *F*) Same as *A*, *C*, and *E* after performing the mean-matching procedure (1) on the L2 norm. Analyses were performed using a 100-ms sliding window.

Across-Trial Pattern Variability (Reproducibility). We tested the across-trial directional variance of stimulus-evoked patterns of activity (*Materials and Methods*), by computing *dva* independently at each time point, across trials. This analysis revealed a significant decrease in variability (increased reproducibility) beginning ∼150–200 ms after stimulus onset ($P < 0.01$ corrected, signed-rank test, for both seen and unseen trials; Fig. 2*A*), remarkably similar to prior single-unit data in monkeys (1). This was followed by a significant difference in variability between seen and unseen trials from ∼400–900 ms poststimulus, with seen trials exhibiting less variability across trials ($P < 0.01$ corrected, cluster-based permuted signed-rank test, $P < 0.01$ samplewise threshold; Fig. 2*A*). We also found a significant difference at ∼100–200 ms ($P < 0.01$ corrected). The results remain significant after performing the mean matching procedure of ref. 1 (Fig. 2 *B* and *D*). Thus, seen trials are more reproducible compared with unseen trials at the same threshold level of color contrast.

A significant difference in *dva* was also found for seen versus unseen blank control stimuli throughout the trial epoch (Fig. 3*A*). Although the temporal extent of this difference was greatly reduced by the mean-matching procedure, the difference in the norm was in the opposite direction of what might indicate an artifactual difference in *dva*. Thus, a decrease in energy (norm) was associated with an increase in stability (lower *dva*). This pattern is inconsistent with the notion that perceptual decisions are heralded by an increase in activity but is consistent with the settling of a recurrent network into a decision state. In addition, even with mean matching the difference remained significant at ∼100–200 ms poststimulus and ∼950–1,050 ms (Fig. 3*B* and cf. Fig. 2*B*). Interestingly, mean matching also revealed a difference at and slightly before the time of stimulus onset for blank control stimuli, consistent with prestimulus activity playing a role in nonveridical perception (28).

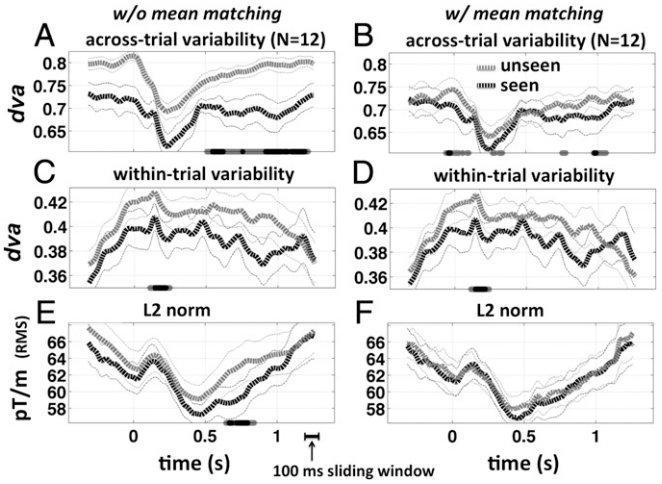

Fig. 3. Reproducibility, stability, and vector norm for blank control trials. *A–F* are the same as in Fig. 2, but for target-absent (control) trials. In this case, reports of "seeing" an object reflect endogenously generated perceptual false positives. Data were noisier because there were fewer trials of this type. Because no object was present on the screen, these trials count as instances where perceptual decision making was decoupled and deconfounded from bottom-up sensory information processing. Note that the significant difference in the norm at ∼600–800 ms (*E*) is in the opposite direction of what might otherwise lead to a lower average *dva* for seen trials (*A*), and thus the observed difference in across-trial variability (reproducibility) cannot be tied to a simple difference in signal strength. The difference in the norm is also the opposite of what one might intuitively predict for seen versus unseen trials. Recurrent network models, however, allow for higher energy in a network that fails to settle into a decision state. The timing of the early difference in within-trial variability (stability) is consistent with that of the visual-awareness negativity (88).

Within-Trial Pattern Variability (Stability). We measured *dva* within a 100-ms-wide sliding window over the entire array of 204 planar gradiometers at a fixed threshold level of color contrast chosen individually for each subject (*Materials and Methods*; unless otherwise indicated, all figures present results for a 100-ms-wide sliding window). We found a short-lived but significant difference in *dva* between seen and unseen trials at ∼100 ms poststimulus (Fig. 2*C*, $P < 0.01$, corrected), followed by a general peak in variability at ∼150 ms. This was followed by a sustained significant difference in *dva* between seen and unseen trials from ∼500–800 ms, with seen trials exhibiting greater stability (lower variability) consistent with our prediction. This was mainly accounted for by a destabilization of the pattern on unseen trials ($P < 0.05$, *dva* at 750 ± 10 ms versus *dva* at 370 ± 10 ms).

As an independent verification of our implementation of directional variance, we repeated the analysis using a singular-value decomposition (SVD) to measure stability: The variance accounted for by the first singular vector is higher when the pattern of activity within a given sensor × time-sample window is more stable (see *SI Appendix*, Fig. S5 for details). We obtained a qualitatively similar result using the SVD (*SI Appendix*, Fig. S5), although *dva* seems to be a more sensitive measure. Importantly, there were no significant differences in the norm of the response vectors that could account for the difference in stability. As a more stringent test, the analysis was repeated using the mean-matching procedure of ref. 1, and the effect remained significant (Fig. 2*D*), effectively ruling out even a small difference in signal strength as a potential explanation. Although the main difference in stability overlapped in time with the peak of the LPP, the peak difference in stability lagged behind the peak difference in the LPP by ∼140 ms ($P < 0.02$, signed-rank test; Fig. 4 *C* and *D*).

As mentioned previously, we found a significant reduction in the power of alpha oscillations for seen versus unseen stimuli during the same time period as the difference in stability (*SI Appendix*, Fig. S4). Because alpha oscillations tend to have a large amplitude, it is conceivable that their suppression might create an artifactual difference in stability, especially if the oscillations vary from above to below the noise floor. To rule out this possible explanation we reran the stability analysis after applying a fifth-order Butterworth bandstop filter (7–13 Hz stop band) to the data. The results remained unchanged (*SI Appendix*, Fig. S6).

Although the level of color contrast was fixed for these analyses (*Materials and Methods*), the probability of a correct response and the probability of a seen response are tightly coupled in general across the range of color contrasts used (*SI Appendix*, Fig. S2 *A* and *B*). Thus, any effects associated with subjective report might be confounded by a difference in task performance. To control for this, we compared the stability associated with seen and unseen trials for correct responses only and found that the main results survived (*SI Appendix*, Fig. S7 *A* and *B*). In addition, we found no difference in stability for correct versus incorrect unseen trials (*SI Appendix*, Fig. S7 *C* and *D*). The number of incorrect seen trials was too small to allow for the same analysis to be applied only to seen trials.

We chose to use a 100-ms sliding window to capture the slow sustained dynamics associated with conscious processing. However, we also performed the stability analysis across a range of different sliding-window widths (*SI Appendix*, Fig. S8). The effect is not apparent for windows below ∼80 ms in width and has a consistent ∼300-ms duration across a range of longer window sizes. A prior computational model (9, 15) predicted that this period of stability should appear immediately following the onset

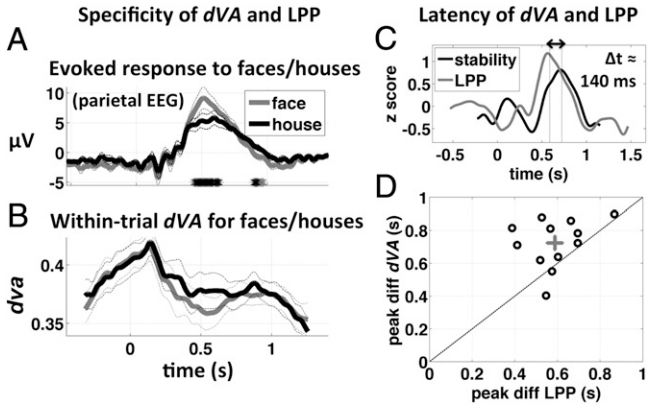

Fig. 4. Specificity and latency of *dva* and the LPP. In this figure we compare the specificity of *dva* and the LPP to the "seen – unseen" dimension and also compare the latency of the peak difference in *dva* (seen – unseen) with that of the LPP. (*A*) Evoked potential at parietal EEG electrode Pz and (*B*) within-trial directional variance for seen face (gray) and seen house (black) stimuli at maximum color contrast. The LPP, which is higher in amplitude for seen versus unseen subjective reports, is also significantly higher in amplitude for face versus house stimuli (*A*). Gray and black stars at the bottom of the panel indicate time points where the difference is greater than chance at $P < 0.05$ and $P < 0.01$, respectively (corrected for temporal nonindependence). No significant differences between face and house stimuli were found for within-trial directional variance (stability, *B*), suggesting that stability is a more specific indicator of a positive subjective report. (*C*) Amplitude-normalized time course of the difference potential at electrode Pz (P300, gray) and the difference in directional variance, seen – unseen object. The difference in the latency of the peaks was 140 ms ($P < 0.01$, two-sided signed-rank test; *Materials and Methods*). (*D*) Scatter plot of the time of peak difference between evoked potentials at Pz, versus the time of peak difference in directional variance (each circle is one subject, $n = 12$). Gray cross shows the mean and SE of both variables. The peak stability effect follows the peak difference in the LPP, consistent with the notion of the LPP as an "ignition" event, the outcome of which is a transient period of relative stability of perceptual information.

of the LPP and last ~200–300 ms, which is very close to what we found.

It might seem surprising that the norm of the activation vectors did not differ significantly as a function of subjective report (Fig. 2*E*), given the difference in amplitude of the LPP. However, we point out that trial averaging suppresses activity whose spatial distribution is variable across trials, even if the net activity across sensors (L2 norm or rms) is high on individual trials (Fig. 1*D*). Also, although individual populations captured by one or more sensors might increase their activity when their preferred stimulus is consciously perceived (29), the spatial norm across the sensor array may or may not differ, depending on the change in activity at other sensors.

Single-Trial Classification of Subjective State. A nearest-mean classifier was able to correctly predict the subjective response (seen or unseen) on individual trials based only on the time course of directional variance (63.7% correct, $\mathbf{d}' = 0.69 \pm 0.1$ SEM, $P < 0.001$ two-sided signed-rank test). As with all of the seen/unseen analyses, this analysis was restricted to a fixed threshold-level of color contrast chosen for each subject. The performance of the same classifier when using a matched spatial filter to reduce the data to one dimension (27) was not significantly different from chance (51.3% correct, $\mathbf{d}' = 0.066 \pm 0.053$ SEM, $P = 0.3$ two-sided signed-rank test).

Selectivity. The LPP (often referred to loosely as the P300) is a well-known correlate of conscious perception (12, 30). However, the LPP is also significantly modulated by the relative frequency of different stimuli, independent of conscious perception, as in the well-known oddball paradigm (30). In our data, we found that the LPP was significantly larger in response to seen face stimuli, compared with its response to seen house stimuli ($P < 0.01$ corrected, at maximum color contrast, correct responses only; Fig. 4*A*). Within-trial directional variance (stability) was relatively insensitive to the stimulus category (Fig. 4*B*) but highly sensitive to the seen–unseen contrast (Fig. 2 *C* and *D*), suggesting that stability is a comparatively selective signature of conscious perception.

Estimation of Conscious State in DOC Patients. We measured within-trial directional variance (100-ms sliding window) and L2 norm in scalp EEG data previously acquired from a cohort of 165 patients with DOC and 12 healthy control subjects (CTRLs) using 256-electrode geodesic sensor nets (EGI Inc.). We used a subset of 166 of the 256 electrodes, excluding electrodes around the periphery of the nets that are often noisy in bedside recordings. Details of this dataset acquired from patients at the Hôpital Pitié Salpêtriere have been reported elsewhere (16–18). Patients were evaluated by trained neurologists using the French version of the Revised Coma Recovery Scale (31). This scale allows for the behavioral classification of the patient as being in a vegetative state (VS), minimally conscious state (MCS), or (recovered) conscious state (CS). As part of an auditory stimulation protocol, subjects were repeatedly exposed to trains of five 50-ms auditory beeps with a 150-ms interstimulus interval (ISI) (32). We treated each tone sequence as a "ping" of the brain, allowing us to measure how the brain responds, similar to the approach that has recently been taken up using transcranial magnetic stimulation (33).

Fig. 5 summarizes the results of our analysis of these data. The mean time course of *dva* and norm diverged parametrically according to conscious state throughout the period of stimulation, with maximal divergence roughly 150 ms after the onset of the last beep (*SI Appendix*, Fig. S9). We computed the mean directional variance and the mean vector norm over the period from 600 to 1,000 ms after the onset of the train of beeps (the final beep in each train was at 600 ms) separately for each patient category. For each variable we also computed the difference between its mean value in this time window and its value at the time of stimulus onset (i.e., change relative to baseline; Fig. 5*A*). The L2 norm was highest for the VS patients and lowest for the

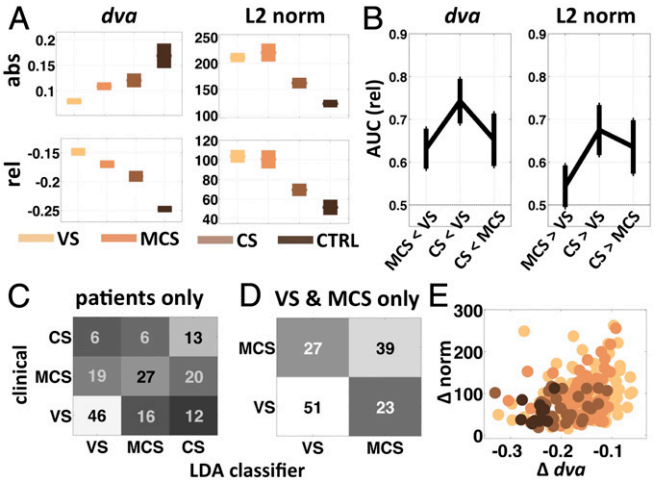

Fig. 5. Stability and norm in DOC patients. (*A*) Mean (± SEM) *dva* and L2 norm over the time period from 600 to 1,000 ms after the onset of the first tone in the tone sequence (0–400 ms after onset of the last tone; *SI Appendix*, Fig. S9) for each of the three patient categories and the healthy control subjects. Bottom row of *A* shows the same two measures, each relative to its value at the onset of the first tone (t_0). (*B*) The area under the receiver operating characteristic (AUC) gives an unbiased measure of the degree to which each of the three pairs of patient categories can be distinguished from each other based on *dva* (*Left*) and the spatial L2 norm (*Right*). (*C* and *D*) Confusion matrices summarizing the results of a linear discriminant analysis (LDA) applied to the patient data. The analysis was run once on all of the patient data (*C*; 50% correct, $P < 0.01$, chance = 33%) and again on only the data from MCS and VS patients (*D*; 65% correct, $P < 0.01$, chance = 50%), which are more challenging to distinguish. (*E*) A scatter plot of the *dva* versus norm, both relative to t_0 (*A*, bottom row), with each patient category color coded as in *A*. abs, absolute; rel, relative; Δ, change relative to baseline.

controls, both in absolute terms and relative to the value at stimulus-train onset (both $P < 0.01$, one-way ANOVA, df = 3). Directional variance, however, increased in absolute terms with increasing conscious state but decreased relative to baseline (both $P < 0.0001$, one-way ANOVA, df = 3) (Fig. 5*A*). This suggests that the healthy brain maintains a higher overall level of variability, but a greater reduction in variability in response to a stimulus, consistent with the idea that the brain normally operates at or near the point of criticality (34–36)—capable of a much greater number of different states, but also a much greater reduction in uncertainty when one of those states is attained (37).

The area under the receiver operating characteristic curve (AUC, for area under the curve) for baseline-subtracted *dva* shows that this variable can discriminate between all pairs of the three patient classes (MCS-VS 0.63 ± 0.05 SEM, $P < 0.005$; CS-VS 0.74 ± 0.05 SEM, $P < 0.0001$; CS-MCS 0.65 ± 0.06 SEM, $P < 0.02$, permutation test; Fig. 5*B*). The AUC for the L2 norm was not significantly different from chance (0.5) for MCS versus VS (0.54 ± 0.048 SEM, $P < 0.2$, consistent with Fig. 5*A*), but the norm could discriminate between CS and VS, and between CS and MCS patients (CS-VS 0.68 ± 0.058 SEM, $P < 0.01$; CS-MCS 0.64 ± 0.062 SEM, $P < 0.03$; Fig. 5*B*). A linear-discriminant analysis classifier applied to the combined *dva* and L2 norm from all patients and control subjects was able to correctly classify the conscious state from among the four possibilities (VS, MCS, CS, and CTRL) far better than expected by chance (49.6% correct, $P < 0.01$, permutation test, bootstrap mean = 26.5%). When the classifier assigned the incorrect label, it tended to assign one of the neighboring states (e.g., mistaking VS for MCS; Fig. 5 *C* and *D*). Importantly, the classifier was able to correctly separate the patient categories, exclusive of the normal controls (50% correct, $P < 0.01$, permutation test, with bootstrap mean = 33%) and most importantly was able to correctly distinguish MCS from VS

patients (65% correct, $P < 0.01$, permutation test, bootstrap mean = 49.6%).

Taken together these results demonstrate that a very simple transformation of the data into the total magnitude of activation (norm) and the net change in temporal variability (within-trial *dva*) in response to a stimulus is highly sensitive to the conscious state of human subjects (Fig. 5*F*). Counter to intuition, loss of consciousness and greater degree of brain damage was associated with a stronger overall neuronal response to stimulation. Interestingly, this latter observation confirms recent reports of consistently and significantly higher-amplitude transient EEG responses to cortical magnetic stimulation in VS compared with MCS or locked-in patients (38) and in non-REM sleep compared with REM sleep or wakefulness (39). This pattern is suggestive of "runaway" activation that, unlike runaway activation in epilepsy, is not self-reinforcing, perhaps owing to a general disruption of both excitatory and inhibitory feedback pathways.

Transient Stability Versus Low-Frequency Energy. A recent study found that conscious perception was correlated with a large slow event-related field and a significant increase in phase locking and spectral power in the 0.05–1.0 Hz range (40). This finding may also arise from the stability phenomenon reported here. A simple monophasic "on–off" episode during which a stable pattern of activity appears over the sensor array (as in Fig. 1*C*), seen from the point of view of a single sensor or local cluster of sensors, will produce both significant energy and phase locking in the frequency whose period is approximately twice the duration of the episode of stability (*SI Appendix*, Fig. S10). Using simulated data (as in Fig. 1*C*, see *Materials and Methods*) we show that a discrete episode of stability in the pattern of activity over an array of sensors of duration 1 s, results in both significant phase locking and significant spectral power at ∼0.5 Hz (*SI Appendix*, Fig. S10). The same also results in a large slow evoked potential at individual sensors, even though total spatial power (L2 norm) was constant throughout all of the simulated epochs (Fig. 1*D*). In addition, the preferred phase for sensors showing a significant effect is guaranteed to cluster around the peak and trough of the low-frequency "cycle" because significant effects have a high (positive or negative) amplitude, and the "phase" of a monophasic event is tied to its amplitude. Thus, stability can account for the low-frequency energy and phase locking associated with conscious perception (40) but provides a different perspective on the underlying cause. Note that whereas "<1 Hz energy" is ill-defined for data windows shorter than ∼500 ms, stability is well-defined for any window size (minimum two samples; *SI Appendix*, Fig. S8).

General Discussion

Using MEG and EEG in human subjects, we investigated changes in the variability of cortical responses to threshold-level stimuli and their implications for the encoding of sensory information. Although the study of variability in neural responses is a relatively recent line of inquiry in human neuroscience (2–4, 41), it has been a major area of research in the macaque, going back at least to the seminal study of Britten et al. (42) who showed, by combining single-unit data with theory, that correlated trial-to-trial variability is associated with reduced fidelity of neuronal signaling.* Subsequent work showed that presentation of a stimulus reduces the magnitude of slow fluctuations in correlated trial-to-trial variability (43) and that directing attention to a stimulus further reduces trial-to-trial variability in the neuronal response (44). Others have gone on to show that fluctuations in spontaneous cortical activity over time may be correlated between neurons separated by as much as 10 mm (45, 46), and that these correlations are cut in half when attention is directed toward a stimulus (47, 48). Collectively, these studies lead to a picture in which cognitive state modulates neuronal response variability, resulting in improved sensory processing.

The present study adds to our understanding of the relationship between neuronal variability and perception in humans by showing direct evidence that measures of response variability predict subjective report, and by examining these in patients with DOC. Our measure of variability was defined in a multivariate space, incorporating activity from broadly distributed areas of cortex, thus distinguishing our study from prior studies that have primarily used univariate measures such as the Fano factor. Most importantly, whereas previous studies have focused on variability across trials, we also examined variability within trial, which reflects the stability of the distributed neuronal response (Fig. 2 *A* and *C*). Variance across trials cannot play a causal role in cognitive and behavioral phenomena that manifest within each single trial (such as whether or not a subject reports having seen a visual object), but within-trial variability across a short span of time (stability) can.

Previous research has shown that the duration and level of activation of neural activity encoding a perceptual object or feature tend to be higher when that object or feature is reported as consciously perceived (29, 49). Dynamic long-range phase synchrony is also enhanced when visual stimuli are consciously seen (14, 22, 50). Importantly, recurrent rather than feed-forward interactions have been implicated in conscious perception (8, 51), and such interactions may reflect a process of convergence toward a transiently stable state (9). The notion of transient stability (or metastability) has been invoked in many theoretical accounts of conscious sensory perception (9, 11, 13, 15, 37, 52, 53). Here we asked whether reports of a conscious sensation in response to a sensory stimulus are preceded by a transient period of relative stability (reduced variability) of neuronal activity following the onset of the stimulus, compared with when no sensation is reported.

The presence of stable attractor states is a computationally useful property of recurrent networks (5, 6). Convergence to a stable state takes longer than feed-forward decisions but results in a sustained activity pattern that is robust to perturbation and provides a transient internal "memory" of the stimulus after it is gone. The stable state can be thought of as an optimal completion of the current input, and thus represents a "high-certainty" state. Temporal stability at the population or systems level will also increase the impact of a given pattern of activity on other neurons that are computing a weighted temporal integral over that pattern. The resulting slow sustained activation may be essential for conscious perception owing to the long-range corticocortical connections that are thought to be involved (54). Thus, the notion of stability is consistent with the properties that are thought to characterize conscious brain states: They take longer to develop (55, 56) than purely feed-forward processes, such as certain fear responses that can proceed outside of awareness (57); they arise late in the sequence of neural events following stimulus onset (58), are sustained over time (22, 29, 49, 54, 59), and are associated with high metacognitive certainty (60–62) and objective discrimination accuracy (3, 14, 21).

The P3 LPP is thought to reflect a domain-general processing event and has been directly associated with explicit perceptual decision making (63). For example, data from whole-cell recordings in mouse barrel cortex point to a causal role for sustained late excitation (∼300–400 ms) in sensory perceptual decision making (64). We replicated the prior observation of a significant difference between seen and unseen trials in the LPP (21), but we also found that the onset of the LPP (*SI Appendix*, Fig. S4*A*) seemed delayed compared with prior studies using achromatic stimuli (21). The late LPP response may have been due to the nature of the stimuli used in the present paradigm, with boundaries defined by color and not by luminance, which may have been processed primarily by the slower parvocellular pathway. Stimuli of low color contrast are perceived as having a longer duration (65) and reaction times are slower when edges are defined only by differences in color and not luminance (66).

*In the literature, "correlated variability" refers to variability across trials that is shared between neurons or populations of neurons.

In general the latency of the LPP might vary depending on the strength of the sensory evidence, which would in turn affect the rate at which it is accumulated, consistent with prior reports (63).

One might argue that our results can also be interpreted in terms of working memory (WM), owing to the imposed 1-s interval between stimulus onset and responding. However, in a typical WM study the task itself is designed to load WM, and WM effects are only observed when its capacity is loaded, and are largely absent when only a single item is maintained (67, 68). In our task the load placed on WM was trivial, because subjects were asked only to report whether the target was a face or a house. In our data, possible evidence of WM involvement [a sustained negative deflection in the evoked potential after the LPP (68)] appears well after the cutoff of the stability effect (*SI Appendix*, Fig. S3*C*, ∼1,000–1,200 ms). Also, given the relatively long response times associated with isoluminant chromatic stimuli (66), subjects may have taken close to 1 s to respond even without the imposed delay. In general it is difficult to separate WM from conscious perception, both conceptually and empirically (69). Even if subjects had made speeded responses in our task one could always argue that WM was engaged any time the subject was asked to produce a response and/or pay attention to and take note of stimuli.

A specific empirical challenge along these lines comes from the recent work of Pitts et al. (70), who argue that two of the canonical correlates of conscious sensory perception, the LPP and induced gamma-band activity, reflect postperceptual processes rather than conscious perception. They found that these correlates disappear when stimuli are task-irrelevant, but still consciously noticed. One limitation of their study is that subjective reports were given after each run rather than after each trial. Thus, there is no way to know precisely when processing of the irrelevant objects took place relative to the trial onsets, because a very demanding primary task was being performed in parallel. Central processing of the irrelevant stimuli may have been pushed back in time by the psychological refractory period because target and irrelevant stimuli were synchronous, but their data analyses only extended out to 600 ms posttarget. Further research will be required to elucidate the apparent disconnect between these results and prior evidence linking the LPP/induced gamma response with conscious perception.

A more general caveat has been raised recently concerning the contrastive method (71, 72), where trials are grouped by subjective report and contrasted, as we did in the present study: Differences in the average over trials with and without conscious perception confound the "core" neural correlate of consciousness with other neural events that reliably precede or follow it (71, 73). One limitation of this view is that it assumes that there exists a single "core" correlate of conscious perception in the first place. Trying to isolate this hypothetical rarefied conscious event (71) may be too restrictive, because it need not be the case that a single isolated parameter or event in the poststimulus time course constitutes *the* neural correlate of consciousness. Instead, the minimally sufficient condition (10) might be a conjunction of separate necessary events happening at different points in time, or a conjunction of factors such that the "minimally sufficient" level of one factor depends on the level of the others (11). Also, the view hinges on the degree to which conscious perceptual events can be said to occur at a specific moment (74), against which "before" and "after" can be defined. Instead, the present results suggest that consciousness may be tied to an extended period of neuronal activity.

We also show that within-trial stability is reflected in between-trial reproducibility, which has previously been proposed as a correlate of conscious perception (3). [Note that in the present study stimuli were always presented in the opposite color configuration to the two eyes. Therefore, prior criticisms of (3) pointing to a possible confound of pre-conscious inter-ocular rivalry (75, 76) are effectively countered.] Increased reproducibility of a pattern of activity across trials should be accompanied by a reduction in individual-neuron firing-rate variability across trials, which has recently been shown to be a general phenomenon following the onset of sensory stimuli (1). Here we show that the variability of the pattern of activity across trials is not only decreased at the onset of a stimulus (equivalent to increased reproducibility), but also, within a later time window, differentiates whether or not the stimulus is reported as seen (Figs. 2 *A* and *B* and 3 *A* and *B*). In addition, both reproducibility and stability might also be associated with more sharply tuned (77) and/or discrete (78) patterns of activity, which have also been proposed as correlates of conscious perception.

Note that stability does not change to nearly the same degree as reproducibility at any time during the trial epoch (Fig. 2). This highlights the partial independence of the two phenomena (variability within and across trials). A neuronal response could be relatively unstable within a certain window of time within each trial, but the average over that same window could still be relatively invariant across many trials. Hence, the appearance of the increase in reproducibility at ∼150–200 ms after stimulus onset (Fig. 2*A*) does not necessarily imply an equally pronounced increase in stability at the same time (Fig. 2*C*).

An important clarification regarding the notion of transient stability is that it should not be taken to imply a single brain-wide point attractor. Our proposal is that a widely distributed subset of ongoing cortical activity, involved in processing and integrating the information content of perception, is transiently stable compared with when the same information content is not consciously perceived. This kind of state might be described as a "chimera state" (52, 79) wherein a subset of a system is relatively integrated (synchronized) and the remainder of the system is relatively nonintegrated (desynchronized). Future theoretical work in this direction could be fruitful. Finally, we point out that our data do not address the question of whether stability is a cause or an effect of conscious perception: It may be both an effect and (through recurrent interactions) a cause (80). The formal definition of stability most commonly used in dynamical systems theory is couched in terms of robustness to perturbation (81) and predicts the persistence of patterns of activity (which we tested here). This definition could be used in future work to directly investigate the causal relationship between stability and perception. Whatever the causal relationship, stability seems to be, as predicted, a reliable and specific signature of conscious perception.

Materials and Methods

Human Subjects. Simultaneous MEG, EEG, and behavioral data were collected from a total of 16 human subjects (8 female, minimum/median/maximum age 19/21/34 y, all right-handed). All subjects had normal trichromatic vision and normal or corrected-to-normal visual acuity. All subjects gave written informed consent and were given a monetary compensation. Data from four of the subjects were not used: One subject was unable to identify the stimuli even at high contrast; one subject completed only 5 runs (out of 12) before complaining of discomfort; technical problems prevented head-position measurements for a third subject; and MEG data for a fourth subject were excessively noisy (inclusion of these subjects, while adding noise to the results, did not change the main conclusions of the study). Thus, unless noted otherwise $n = 12$. (A separate section below discusses DOC patients.)

Stimuli and Task.

Stimuli: Preparation and presentation. We used dichoptic color masking (*SI Appendix*, Fig. S1) (3, 82) to reduce the subjective visibility of visual objects while maintaining a physical stimulus strength that would otherwise be clearly visible (see *SI Appendix, Supporting Methods* for details). Stimuli were simple line drawings of faces and houses (*SI Appendix*, Fig. S1), 175 × 175 pixels, rendered in exactly two colors using error-diffusion dither and presented stereoscopically with the aid of a cardboard divider and prism lenses mounted in MEG-compatible frames (83). Stimuli were back-projected onto a translucent viewing screen at a distance of ∼60 cm, under low ambient lighting, and were always presented in the "opposite color" configuration (*SI Appendix*, Fig. S1). Stimuli always appeared within a pair of black square frames subtending ∼5.5° of visual angle and positioned on either side of the cardboard divider. The frames remained on the screen at all times during blocks of trials to aid the maintenance of stereoscopic fusion. When fused, the two frames were perceived as a single black frame. To confirm

interocular fusion of the two frames, at the beginning of the experiment a dual-image random-dot stereogram was presented within the frames and subjects had to name the symbol hidden in the stereogram. The background of the stimulus display was blue and adjusted so as to be similar in luminance to the stimuli, so that the onset of stimuli would be less likely to provoke blinks or pupillary reflexes, which might disrupt fusion of the two images. Control stimuli consisted of uniform colored squares, equivalent to the background of the object stimuli, but without the object. Subjects were not explicitly told about the control stimuli, but were only told that the stimuli would vary in salience and might sometimes be imperceptible.

Practice tasks. Before entering the testing room, subjects practiced maintaining steady fixation by trying to maintain the Troxler illusion (disappearance of a blurred peripheral annulus) for a few seconds at a time. Subjects were instructed to maintain steady fixation during the experiment, whenever the fixation point was present on the screen (i.e., for the first 2 s of each trial, beginning just before stimulus onset). Before beginning the experiment, subjects performed three "preview" blocks during which they practiced discriminating the category of the stimuli (face versus house) at three suprathreshold, successively lower levels of color contrast. During the preview blocks, if the subject responded incorrectly, the fixation point blinked and the same stimulus was presented again. The purpose of the practice session was to familiarize the subjects with the stimuli and also to precipitate perceptual-learning effects that might otherwise accrue during the experiment itself. Control stimuli were not used during the practice task.

Experimental task. The experiment was performed in 12 blocks of 60 trials each, with each block lasting ∼5 min. Time constraints and technical difficulties did not always allow completion of all 12 runs, but we imposed a minimum of 9 runs for inclusion in the data analyses (72 trials per contrast level = 36 face and 36 house). Each block had an equal number of face, house, and control stimuli at each of the five levels of color contrast, once for each assignment of background color to the two eyes (orange to the left eye, green to the right eye, and vice versa). A single visual stimulus was presented on each trial and the task was to identify the category (face or house) of the stimulus, guessing if necessary, and also to report on the subjective visibility of the object (seen or unseen). Both responses were made by button press. The assignment of buttons to responses and the order in which the responses were given were counterbalanced across subjects.

The sequence of events in each trial was as follows (*SI Appendix*, Fig. S1). Each trial began with only the black frame(s) on the screen. Then a small fixation point appeared in the middle of the black frame(s) cueing the subject to maintain steady fixation for the next 2 s. After a delay of 500 ± 117 ms the stimulus appeared at one of the five preselected levels of color contrast, for a duration of 100 ms. The fixation point remained on for another 900 ms and then blinked, cueing the subject to respond to the first question (either face/house or seen/unseen), and then blinked a second time, cueing the subject to respond to the second question, with a 2-s time limit for each response. After the second response was made, the fixation point disappeared for an intertrial interval of 2 s. Subjects were instructed to respond "unseen" only if they had the subjective impression of a uniform yellow square, and failed to detect the presence of any contours or edges within the patch of color. As part of a behavioral pilot test, 6 of the 13 subjects were given the option to respond "seen-but-not-recognized," in addition to "seen" or "unseen." For all data analyses, these were treated as seen responses.

Simultaneous MEG/EEG Data Acquisition. MEG and EEG data were recorded simultaneously in a magnetically shielded chamber using equipment built by Elekta Neuromag. The MEG recording apparatus has a total of 306 sensors, with 102 magnetometers and 102 pairs of orthogonal planar gradiometers. In addition the subject wore an MEG-compatible EEG cap with 60 ring-shaped electrodes filled with conductive electrode gel, with a reference electrode placed on the nose, and a ground electrode on the clavicle. The impedance between the reference and each of the other electrodes was kept below 15 kOhm. Separate electrode pairs connected to bipolar amplifier channels were used to monitor cardiac and ocular activity. Four small coils used for estimating the head position within the MEG helmet were taped to the EEG cap. The positions of the four coils with respect to three fiduciary points (nasion and just above the tragus in each ear) were digitized using a 3D Fastrak digitizer (Polhemus Inc.). The subject's head position was recorded at the beginning of each block of trials. EEG and MEG data were sampled at 1,000 Hz, with a 0.03- to 333-Hz analog band-pass filter. The MaxShield feature of the Neuromag MEG system, which actively compensates for external magnetic field fluctuations, was used during the recordings.

MEG and EEG Data Preprocessing. MEG data were preprocessed using the MaxFilter software application developed by Elekta Neuromag. This application automatically detects and interpolates bad MEG channels, applies the Signal Space Separation algorithm (for attenuation of magnetic field fluctuations originating outside of the MEG helmet), and transforms the MEG data to a head-centered coordinate system using the head-position measurements taken at the beginning of each block of trials. The MEG and EEG data were then imported into MATLAB (The MathWorks, Inc.) for further preprocessing and subsequent data analyses using the FieldTrip software toolbox for MatLab (fieldtrip.fcdonders.nl). At this stage, both the EEG and MEG data were down-sampled to 250 Hz. Data epochs were extracted spanning the time period from 0.5 s before stimulus onset to 1.5 s after stimulus onset. No baseline correction was performed. Ocular and cardiac artifacts were isolated using independent components analysis (84) and removed by projecting the data onto the pseudoinverse of the artifact components. Trials with noticeable artifacts remaining after this step were rejected based on visual inspection, performed separately for EEG and MEG data.

Data Analyses and Statistics.

Behavioral. "Proportion correct" and "proportion seen" were computed over noncontrol (face and house) trials. Detection *d-prime* (d') was computed by treating face and house trials as "target present," control trials as "target absent," and a seen response on a target-present trial as a "hit' (or true positive). d' was then computed as *zinv*(HIT) – *zinv*(FA), where HIT and FA are the true-positive and false-alarm rate, respectively, and *zinv* is the inverse normal distribution. d' for face versus house was computed on noncontrol trials, with a "face" response on a face trial treated as a true positive (hit) and a "face" response on a house trial treated as a false positive.

Seen versus unseen analyses of MEG and EEG data. For all analyses where we compared responses to seen and unseen stimuli, we chose a single fixed level of color contrast for each subject and restricted the analysis to trials with stimulation at that color contrast. The color contrast was chosen separately for each subject by finding the color contrast where the number of seen and unseen responses was closest to being equal. Of the 12 subjects included in the analysis, 6 had more seen than unseen trials, 5 had fewer seen than unseen trials, and 1 had exactly the same number of seen and unseen trials, at the chosen contrast level. The average number of trials per subject per condition (seen or unseen) at the single fixed contrast level was 42 (minimum 35, maximum 59).

Spatial filtering. For the time-frequency analyses, effect-matched spatial filtering (27) was used to summarize multisensor data in a single time course. Filters were applied only to the gradiometer data, in order for the results to be directly comparable to the measures of stability and reproducibility (Figs. 2 and 3) that were also computed from gradiometer data. Spatial filtering was performed separately at each time point in the trial and in a leave-one-out fashion by projecting each trial's data onto a filter estimated from all of the other trials. Filters used in *SI Appendix*, Fig. S4 *C* and *D* were derived by taking the difference, at each sensor, between the mean over seen trials and the mean over unseen trials, and between the mean over face trials and the mean over house trials, respectively.

dva. The operational definition and procedures for estimating stability described below were defined a priori. Stability was operationally defined as the relative constancy of the pattern of activity across a set of sensors within a given window of time, within a given trial. Stability was measured at the single-trial level using a 100-ms sliding window, and then averaged across trials. A range of window sizes ≥80 ms gave similar results (*SI Appendix*, Fig. S8). We chose (a priori) to perform this analysis on data from gradiometers because we defined stability in terms of the spatial pattern of activity, and planar gradiometers are known to have a higher degree of spatial selectivity than either magnetometers or scalp electrodes (85).

Directional variance can be defined in terms of the norm of the sum of a set of vectors. If each of the vectors is first normalized to unit length then the norm of the resultant vector sum will be proportional to the degree to which the vectors all point in the same direction (Fig. 1). The logic is similar to that applied in Rayleigh's statistical test of the distribution of phase angles. Directional variance (also called circular variance or angular variance), *dva*, is formally defined as follows (20):

For a set of vectors, v_1, v_2, v_3, …, v_n

$$dva = 1 - \overline{R},$$

where

$$\overline{R} = \frac{\left\| \frac{v_1}{\|v_1\|} + \frac{v_2}{\|v_2\|} + \frac{v_3}{\|v_3\|} + \dots + \frac{v_n}{\|v_n\|} \right\|}{n}$$

and

$$0 \leq dva \leq 1.$$

Stability is defined as $1 - dva$, or simply $\overline{R}$, which we refer to as "directional coherence." Similar metrics have been used for vector-based analyses in other domains, for example in measuring the persistence (over time) of ocean currents in a given direction (86).

In computing $\overline{R}$ each of the vectors in the numerator is divided by its own norm so that only the angle of the vectors in relation to one another will determine the norm of their sum. This means that weak vectors and strong vectors contribute equally. Because the angle of weak vectors will tend to be influenced by noise to a greater degree than the angle of strong vectors, differences in the signal-to-noise ratio can produce artifactual differences in stability. Therefore, it is essential to test for a possible difference in the norm (signal strength or power) of the vectors (3), which we do systematically for each analysis here.

Directional variance was also used to compute reproducibility across trials, except in this case the metric was computed across trials where each vector in the set of vectors $v_{1...n}$ was the mean vector within the 100-ms sliding window centered at a given time point in each trial. The L2 norm was computed by taking the square root of the sum of squares of the vectors within the sliding window, taking the average within the window, and then averaging across trials.

Simulations. We performed simulations using MATLAB (The MathWorks, Inc.) to validate *dva* and test its behavior under different controlled conditions (Fig. 1 *C* and *D* and *SI Appendix*, Fig. S10). See *SI Appendix, Supporting Methods* for details.

Single-trial classification of subjective state. For the fixed threshold level of color contrast chosen for each subject, we attempted to classify each trial as either seen or unseen based only on the time course of directional variance. We used a variant of a nearest-mean classifier: the mean seen time course and mean unseen time course were computed with one test trial left out. We then computed the mean-squared distance of the test time course from each of the two means. The category (seen or unseen) with the smallest mean squared distance was taken as the predicted class (the classifier's decision).

Time-frequency analysis. Time-frequency decomposition was performed using the FieldTrip toolbox for MATLAB (87) using Morlet wavelets, with the ratio parameter set to 7. The time-frequency decompositions were performed on the output of a spatial filter derived based on the overall difference between seen and unseen trials and on the output of a spatial filter derived based on the overall difference between face and house trials (27). For statistical analysis, a pointwise two-sided signed-rank test was performed across subjects ($n = 12$) for each point in time-frequency space. The resulting statistics were then corrected for multiple comparisons using a cluster-based permutation test (87), with a pointwise threshold of $P < 0.05$.

States of Consciousness Subjects.

Patients. We analyzed 165 EEG recordings taken from 116 unique DOC patients (a single recording from 86 of the patients, and two or more recordings from each of the remaining 30 patients). Because of the variability within patient, between recording sessions, and the large number of patients, we treated each recording as an independent sample. Each patient was classified as being in a VS, MCS, or (recovered) CS, based on clinical evaluation by trained neurologists. The Ethical Committee of the Pitie-Salpetriere hospital approved this research under the French rubric of *Recherche en soins courants* (routine care research). The number of recordings in each category was as follows: 25 CS (3 female, 22 male), median age 49 y (minimum 16, maximum 83); 66 MCS (18 female, 48 male), median age 45 y (minimum 16, maximum 78); and 74 VS (23 female, 51 male), median age 45 y (minimum 18, maximum 80).

Healthy subjects. Experiments were approved by the Ethical Committee of the Pitie-Salpetriere Hospital. All 12 healthy subjects gave written informed consent. Of the 12 healthy control subjects 2 were female, 10 were male, and the median age was 21 y (minimum 20, maximum 30).

Auditory stimulation. Patients and healthy subjects were exposed to auditory stimuli following the "local–global" protocol (32). Stimuli were trains of five brief auditory tone pips (5 × [50 ms tone, 100 ms silence] with 750 ms gap in between each train) played repeatedly in blocks of 3–4 min. The pitch of the last tone in each train could be the same as or different from that of the other four. Further details of the protocol can be found in refs. 16, 18, and 32.

ACKNOWLEDGMENTS. We are grateful to the NeuroSpin infrastructure groups, in particular to Ghislaine Dehaene-Lambertz, Andreas Kleinschmidt, Caroline Huron, Lucie Hertz-Pannier, Véronique Joly-Testault, and Laurence Laurier for their support in subject recruitment and testing; Virginie van Wassenhove, Marco Buiatti, and Leila Rogeau and the magnetoencephalography (MEG) team for their help with technical matters; and to Alexandre Gramfort for suggesting the singular-value decomposition analysis. We also wish to thank the editor and two anonymous reviewers. This work was supported by INSERM, Commissariat à l'Energie Atomique (CEA), the James S. McDonnell Foundation (collaborative activity award), the Fondation pour la Recherche Médicale (FRM) (Equipe FRM 2010 Grant to L.N.), Institut pour le Cerveau et la Moëlle épinière (ICM Institute), the program ''Investissements d'avenir'' ANR-10-IAIHU-06, and a European Research Council senior grant "NeuroConsc" and a Spelberch Foundation award (to S.D.). The Neurospin MEG facility was sponsored by grants from INSERM, CEA, Fondation pour la Recherche Médicale, the Bettencourt-Schueller Foundation, and the Région île-de-France. A.S. was supported by a Marie Curie postdoctoral fellowship from the European Commission (FP7-PEOPLE-2009-IIF, Project 252665). I.S. was supported by an LLP/ERASMUS scholarship from the European Commission.

1. Churchland MM, et al. (2010) Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat Neurosci* 13(3):369–378.
2. He BJ (2013) Spontaneous and task-evoked brain activity negatively interact. *J Neurosci* 33(11):4672–4682.
3. Schurger A, Pereira F, Treisman A, Cohen JD (2010) Reproducibility distinguishes conscious from nonconscious neural representations. *Science* 327(5961):97–99.
4. Xue G, et al. (2010) Greater neural pattern similarity across repetitions is associated with better memory. *Science* 330(6000):97–101.
5. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79(8):2554–2558.
6. Hopfield JJ, Tank DW (1986) Computing with neural circuits: A model. *Science* 233(4764): 625–633.
7. Tononi G, Edelman GM (1998) Consciousness and complexity. *Science* 282(5395): 1846–1851.
8. Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 23(11):571–579.
9. Dehaene S, Sergent C, Changeux JP (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci USA* 100(14):8520–8525.
10. Crick F, Koch C (1998) Consciousness and neuroscience. *Cereb Cortex* 8(2):97–107.
11. Kinsbourne M (1997) What qualifies a representation for a role in consciousness? *Scientific Approaches to Consciousness.* Carnegie Mellon Symposia on Cognition, eds Cohen JD, Schooler JW (Lawrence Erlbaum, Hillsdale, NJ), pp 335–355.
12. Dehaene S, Changeux JP (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70(2):200–227.
13. Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: Phase synchronization and large-scale integration. *Nat Rev Neurosci* 2(4):229–239.
14. Melloni L, et al. (2007) Synchronization of neural activity across cortical areas correlates with conscious perception. *J Neurosci* 27(11):2858–2865.
15. Dehaene S, Changeux JP (2005) Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biol* 3(5):e141.
16. Sitt JD, et al. (2014) Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 137(Pt 8):2258–2270.
17. King J-R, et al. (2013) Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr Biol* 23(19):1914–1919.
18. Faugeras F, et al. (2012) Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. *Neuropsychologia* 50(3):403–418.
19. Nili H, et al. (2014) A toolbox for representational similarity analysis. *PLOS Comput Biol* 10(4):e1003553.
20. Fisher N (1993) *Statistical Analysis of Circular Data* (Cambridge Univ Press, Cambridge, UK).
21. Del Cul A, Baillet S, Dehaene S (2007) Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol* 5(10):e260.
22. Gaillard R, et al. (2009) Converging intracranial markers of conscious access. *PLoS Biol* 7(3):e61.
23. Wyart V, Tallon-Baudry C (2009) How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. *J Neurosci* 29(27):8715–8725.
24. Wyart V, Tallon-Baudry C (2008) Neural dissociation between visual awareness and spatial attention. *J Neurosci* 28(10):2667–2679.
25. Schurger A, Cowey A, Cohen JD, Treisman A, Tallon-Baudry C (2008) Distinct and independent correlates of attention and awareness in a hemianopic patient. *Neuropsychologia* 46(8):2189–2197.
26. Schurger A, Cowey A, Tallon-Baudry C (2006) Induced gamma-band oscillations correlate with awareness in hemianopic patient GY. *Neuropsychologia* 44(10):1796–1803.
27. Schurger A, Marti S, Dehaene S (2013) Reducing multi-sensor data to a single time course that reveals experimental effects. *BMC Neurosci* 14:122.
28. Hesselmann G, Kell CA, Eger E, Kleinschmidt A (2008) Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proc Natl Acad Sci USA* 105(31): 10984–10989.
29. Quiroga RQ, Mukamel R, Isham EA, Malach R, Fried I (2008) Human single-neuron responses at the threshold of conscious recognition. *Proc Natl Acad Sci USA* 105(9): 3599–3604.
30. Picton TW (1992) The P300 wave of the human event-related potential. *J Clin Neurophysiol* 9(4):456–479.
31. Giacino JT, Kalmar K (2005) Diagnostic and prognostic guidelines for the vegetative and minimally conscious states. *Neuropsychol Rehabil* 15(3-4):166–174.

32. Bekinschtein TA, et al. (2009) Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci USA* 106(5):1672–1677.
33. Casali AG, et al. (2013) A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Transl Med* 5(198):198ra105.
34. Haimovici A, Tagliazucchi E, Balenzuela P, Chialvo DR (2013) Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Phys Rev Lett* 110(17):178101.
35. Yang H, Shew WL, Roy R, Plenz D (2012) Maximal variability of phase synchrony in cortical networks with neuronal avalanches. *J Neurosci* 32(3):1061–1072.
36. Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi RJ (2001) Long-range temporal correlations and scaling behavior in human brain oscillations. *J Neurosci* 21(4): 1370–1377.
37. Balduzzi D, Tononi G (2008) Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLOS Comput Biol* 4(6):e1000091.
38. Rosanova M, et al. (2012) Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain* 135(Pt 4):1308–1320.
39. Massimini M, et al. (2010) Cortical reactivity and effective connectivity during REM sleep in humans. *Cogn Neurosci* 1(3):176–183.
40. Li Q, Hill Z, He BJ (2014) Spatiotemporal dissociation of brain activity underlying subjective awareness, objective performance and confidence. *J Neurosci* 34(12):4382–4395.
41. King J-R, Gramfort A, Schurger A, Naccache L, Dehaene S (2014) Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS ONE* 9(1):e85791.
42. Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J Neurosci* 12(12):4745–4765.
43. Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25(14):3661–3673.
44. Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55(1):131–141.
45. Smith MA, Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28(48):12591–12603.
46. Nauhaus I, Busse L, Carandini M, Ringach DL (2009) Stimulus contrast modulates functional connectivity in visual cortex. *Nat Neurosci* 12(1):70–76.
47. Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63(6):879–888.
48. Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12(12):1594–1600.
49. Fisch L, et al. (2009) Neural "ignition": Enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 64(4):562–574.
50. Rodriguez E, et al. (1999) Perception's shadow: Long-distance synchronization of human brain activity. *Nature* 397(6718):430–433.
51. Auksztulewicz R, Spitzer B, Blankenburg F (2012) Recurrent neural processing and somatosensory awareness. *J Neurosci* 32(3):799–805.
52. Shanahan M (2010) Metastable chimera states in community-structured oscillator networks. *Chaos* 20(1):013108.
53. O'Brien G, Opie J (1999) A connectionist theory of phenomenal experience. *Behav Brain Sci* 22(1):127–148, discussion 148–196.
54. He BJ, Raichle ME (2009) The fMRI signal, slow cortical potential and consciousness. *Trends Cogn Sci* 13(7):302–309.
55. Libet B, Alberts WW, Wright EW, Jr, Feinstein B (1967) Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science* 158(3808): 1597–1600.
56. Mealor A, Dienes Z (2012) No-loss gambling shows the speed of the unconscious. *Conscious Cogn* 21(1):228–237.
57. Jiang Y, He S (2006) Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Curr Biol* 16(20):2023–2029.
58. Sergent C, Baillet S, Dehaene S (2005) Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci* 8(10):1391–1400.
59. Fahrenfort JJ, et al. (2012) Neuronal integration in visual cortex elevates face category tuning to conscious face perception. *Proc Natl Acad Sci USA* 109(52):21504–21509.
60. Dienes Z, Seth A (2010) Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious Cogn* 19(2):674–681.
61. Persaud N, McLeod P, Cowey A (2007) Post-decision wagering objectively measures awareness. *Nat Neurosci* 10(2):257–261.
62. Charles L, Van Opstal F, Marti S, Dehaene S (2013) Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* 73:80–94.
63. O'Connell RG, Dockree PM, Kelly SP (2012) A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci* 15(12):1729–1735.
64. Sachidhanandam S, Sreenivasan V, Kyriakatos A, Kremer Y, Petersen CC (2013) Membrane potential correlates of sensory perception in mouse barrel cortex. *Nat Neurosci* 16(11):1671–1677.
65. Kojima H, Kawabata Y (2012) Perceived duration of chromatic and achromatic light. *Vision Res* 53(1):21–29.
66. McKeefry DJ, Parry NRA, Murray IJ (2003) Simple reaction times in color space: The influence of chromaticity, contrast, and cone opponency. *Invest Ophthalmol Vis Sci* 44(5):2267–2276.
67. Cohen JD, et al. (1997) Temporal dynamics of brain activation during a working memory task. *Nature* 386(6625):604–608.
68. Vogel EK, Machizawa MG (2004) Neural activity predicts individual differences in visual working memory capacity. *Nature* 428(6984):748–751.
69. Naghavi HR, Nyberg L (2005) Common fronto-parietal activity in attention, memory, and consciousness: Shared demands on integration? *Conscious Cogn* 14(2):390–425.
70. Pitts MA, Padwal J, Fennelly D, Martínez A, Hillyard SA (2014) Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *Neuroimage* 101: 337–350.
71. Aru J, Bachmann T, Singer W, Melloni L (2012) Distilling the neural correlates of consciousness. *Neurosci Biobehav Rev* 36(2):737–746.
72. Pitts MA, Britz J (2011) Insights from intermittent binocular rivalry and EEG. *Front Hum Neurosci* 5:107.
73. Sergent C, Naccache L (2012) Imaging neural signatures of consciousness: 'What', 'when', 'where' and 'how' does it work? *Arch Ital Biol* 150(2–3):91–106.
74. Dennett DC, Kinsbourne M (1992) Time and the observer: The where and when of consciousness in the brain. *Behav Brain Sci* 15(2):183–201.
75. Schwarzkopf DS, Rees G (2010) Neuroscience. Brain activity to rely on? *Science* 327(5961):43–44.
76. Clifford CW (2010) Consciousness: Reading the neural signature. *Curr Biol* 20(2): R61–R62.
77. Ling S, Blake R (2009) Suppression during binocular rivalry broadens orientation tuning. *Psychol Sci* 20(11):1348–1355.
78. Sackur J, Dehaene S (2009) The cognitive architecture for chaining of two mental operations. *Cognition* 111(2):187–211.
79. Wildie M, Shanahan M (2012) Metastability and chimera states in modular delay and pulse-coupled oscillator networks. *Chaos* 22(4):043131.
80. Graziano MSA, Kastner S (2011) Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cogn Neurosci* 2(2):98–113.
81. Holmes P, Shea-Brown E (2006) Stability. *Scholarpedia* 1(10):1838.
82. Moutoussis K, Zeki S (2002) The relationship between cortical activation and perception investigated with invisible stimuli. *Proc Natl Acad Sci USA* 99(14):9527–9532.
83. Schurger A (2009) A very inexpensive MRI-compatible method for dichoptic visual stimulation. *J Neurosci Methods* 177(1):199–202.
84. Parra LC, Spence CD, Gerson AD, Sajda P (2005) Recipes for the linear analysis of EEG. *Neuroimage* 28(2):326–341.
85. Parkkonen L (2010) Instrumentation and data pre-processing. *MEG: An Introduction to Methods*, eds Hansen P, Kringelbach M, Salmelin R (Oxford Univ Press, New York), Chap 2.
86. Kostianoy AG, Nihoul JCJ, Rodionov VB (2004) Frontal zones in the Bering Sea. *Physical Oceanography of the Frontal Zones in Sub-Arctic Seas* (Elsevier, Amsterdam), Vol 71, pp 135–189.
87. Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869.
88. Koivisto M, et al. (2008) The earliest electrophysiological correlate of visual awareness? *Brain Cogn* 66(1):91–103.