Radical embodiment: neural dynamics and consciousness

Evan Thompson and Francisco J. Varela

We propose a new approach to the neuroscience of consciousness, growing out of the 'enactive' viewpoint in cognitive science. This approach aims to map the neural substrates of consciousness at the level of large-scale, emergent and transient dynamical patterns of brain activity (rather than at the level of particular circuits or classes of neurons), and it suggests that the processes crucial for consciousness cut across the brain–body–world divisions, rather than being brain-bound neural events. Whereas standard approaches to the neural correlates of consciousness have assumed a one-way causal-explanatory relationship between internal neural representational systems and the contents of consciousness, our approach allows for theories and hypotheses about the two-way or reciprocal relationship between embodied conscious states and local neuronal activity.

Embodiment and situatedness are now common themes in cognitive science[1]. Recently, Andy Clark drew attention to three controversial 'radical embodiment' propositions often found in current research[2]:

(1) understanding the complex interplay of brain, body and world requires the tools and methods of nonlinear dynamical systems theory;

(2) traditional notions of representation and computation are inadequate;

(3) traditional decompositions of the cognitive system into inner functional subsystems or modules ('boxology') are misleading, and blind us to arguably better decompositions into dynamical systems that cut across the brain–body–world divisions.

These claims taken together express the viewpoint known as 'enactive' cognitive science[3]. Here we explore the implications of this viewpoint for the neuroscience of consciousness.

NCCs: radical embodiment and the standard view

It has become commonplace to think that one of the first steps to a scientific theory of consciousness is to discover the 'neural correlates of consciousness' (NCCs). The term 'correlate' is potentially misleading, however, for the goal is to discover not mere correlates, but rather neural events that are necessary and/or sufficient for the contents of

consciousness[4], as can be seen from this useful definition of the NCC notion:

> An NCC (for content) is a minimal neural representational system N such that representation of a content in N is sufficient, under conditions C, for representation of that content in consciousness[5].

In this definition the causal-explanatory relationship is one-way, from internal neural events to conscious experience. We propose a more expansive, two-way account, based on the three radical-embodiment propositions. Using the tools of dynamical systems theory, we suggest: (1) that as a result of the generic feature of 'emergence' in complex systems, one can expect there to be two-way or reciprocal relationships between neural events and conscious activity; and (2) that the processes crucial for consciousness cut across brain–body–world divisions, rather than being brain-bound neural events.

Resonant neural assemblies and large-scale brain integration

Cognitive neuroscience now leaves little doubt that specific cognitive acts require the transient integration of numerous, widely distributed, constantly interacting areas of the brain[6–11]. Therefore, any hypothesis about the neural correlates of consciousness must account for the integrated or coherent operation of large-scale brain activity. Recently, Varela and colleagues extensively reviewed this issue of large-scale integration[11]. Here we briefly summarize only the main points.

The most plausible mechanism for large-scale integration is the formation of dynamic links mediated by synchrony over multiple frequency bands. Neuronal groups exhibit a wide range of oscillations (in the theta to gamma ranges, 6–80 Hz), and can enter into precise synchrony over a limited period of time (a fraction of a second). Synchrony in this context means precise phase-locking as directly quantified by novel statistical methods[12] (rather than indirect measures of synchrony in terms of spectral coherence that do not separate phase and amplitude components). The role played by such synchronization of neuronal discharges has been greatly highlighted by recent results from microelectrode physiology in animals[13]. Two scales of phase synchrony can be distinguished: short-range and long-range. Most electrophysiological studies in animals have dealt with short-range synchronies[14] or synchronies between adjacent areas corresponding to a single sensory modality[15]. These local synchronies have usually been interpreted as subserving 'perceptual binding'. Detailed evidence for long-range synchronizations between widely separated brain regions during cognitive tasks has also been found[16–20]. Figure 1 illustrates this in terms of gamma activity recorded

Evan Thompson
Dept of Philosophy and Centre for Vision Research, York University, 4700 Keele Street, North York, Ontario, Canada M3J 1P3.
e-mail: evant@yorku.ca

Francisco J. Varela*
Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale (LENA), CNRS UPR 640, Hôpital de la Salpètrîere, 47 Blvd de l'Hôpital, 75651 Paris Cedex 13, France.

*Francisco Varela passed away May 28, 2001. An obituary prepared by Evan Thompson appeared in the August 2001 issue of this journal. A longer version is available on the internet at http://psyche.csse.monash.edu.au/v7/psyche-7-12-thompson.html.
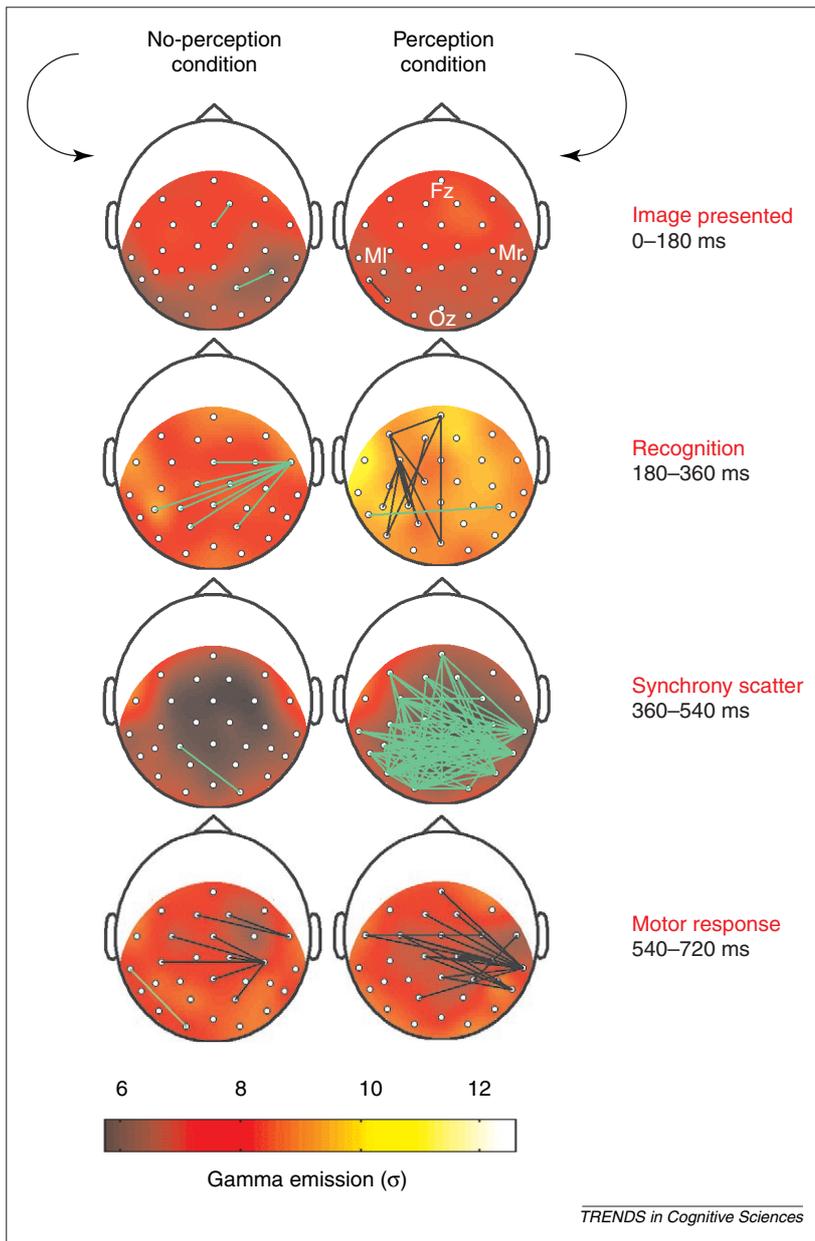
Fig. 1. The shadow of a perception. Average scalp distribution of gamma activity and phase synchrony. EEG was recorded from electrodes at the scalp surface. Subjects were shown upright and upside-down Mooney figures (high contrast faces), which are easily perceived as faces when presented upright, but usually perceived as meaningless black-and-white forms when upside-down. The subjects' task was a rapid two-choice button response of whether or not they perceived a face at first glance. Color coding indicates gamma power (averaged in a 34–40 Hz frequency range) over a given electrode and during a 180 ms time window, from stimulation onset (0 ms) to motor response (720 ms). Gamma activity is spatially homogenous and similar between conditions over time. By contrast, phase synchrony is markedly regional and differs between conditions. Synchrony between electrode pairs is indicated by black and green lines lines, corresponding to a significant increase or decrease in synchrony, respectively. These are shown only if the synchrony value is beyond the distribution of shuffled data sets (*P* < 0.01; see methods, Ref. 18). Modified from Ref. 18.

when subjects viewed upright and upside-down Mooney faces (high contrast face stimuli). Marked quantitative and qualitative differences are observed between the 'perception' (upright face) and 'no perception' (upside-down face) conditions. Between 200 and 260 ms, a first period of significant synchronization was observed in the perception condition, involving left parieto-occipital and fronto-temporal regions. It was followed by a massive period of loss of synchronization well below the reference level, peaking around 500 ms ('phase scattering'), and involving parietal and occipto-temporal regions bilaterally. (Synchrony scatter may reflect a process of active desynchronization that is necessary for the transition between different synchronous neural assemblies; see Refs 7,11,18). A second period of synchrony increase appears around 700 ms. This coincides with the subject's button-pressing, and so might be related to the motor response in both conditions. This is the only period of time in which some similarity between the phase synchrony patterns under the two conditions was noted.

These results tally with the more general notion that phase synchrony should subserve not simply the binding of sensory attributes, but the overall integration of all the dimensions of a cognitive act, including associative memory, emotional tone and motor planning[6,7,11]. This evidence suggests that the best opportunity for understanding the neural basis of consciousness is likely to be at the level of 'dynamical brain signatures'[21] (large-scale dynamical patterns of activity over multiple frequency bands) rather than the structural level of specific circuits or classes of neurons[22]. Nevertheless, the available evidence so far regarding synchronization in the vertebrate brain is only correlative, not causal; there is still no direct proof that changes in synchrony lead to changes in either behavior or consciousness. Future studies in this direction are a priority.

Emergence and reciprocal causation

The neural processes just discussed fall into the broad category of networks of nonlinear oscillators, a paradigmatic example of self-organization[23]. A key concept in this context is that of 'emergent processes' (Box 1). Emergent processes correspond to the collective behaviors of large ensembles, in which positive and negative feedback interactions give rise to non-proportional (or nonlinear) consequences.

Emergence through self-organization has two directions. First, there is local-to-global determination or 'upward causation', as a result of which novel processes emerge that have their own features, lifetimes and domains of interaction. Second, there is global-to-local determination, often called 'downward causation'[24], whereby global characteristics of a system govern or constrain local interactions. This aspect of emergence is less frequently discussed, but has long been noted by researchers in the field of complex dynamical systems[25–27]. It is central to some views about consciousness and the brain[28,29], and plays a key role in an important recent analysis that uses dynamical-systems theory to understand intentional action[30].

Global-to-local effects do not take the same form as local-to-global ones: they are typically manifest

Box 1. Emergence and downward causation

Although the notion of emergence is prominent in the sciences of nonlinear complexity[a,b], there is still no philosophical consensus on this notion. We offer the following provisional definition of emergence in the case of complex dynamical systems:

Defn: A network, *N*, of interrelated components exhibits an emergent process, *E*, with emergent properties, *P*, if and only if:

(1) *E* is a global process that instantiates *P* and arises from the nonlinear dynamics, *D*, of the local interactions of *N*'s components.

(2) *E* and *P* have a global-to-local ('downward') determinative influence on the dynamics *D* of the components of *N*.

And (possibly):

(3) *E* and *P* are not exhaustively determined by the intrinsic properties of the components of *N*, that is, they exhibit 'relational holism'.

Although the term 'emergent property' is widespread, we prefer 'emergent process'. Strictly speaking, a property *per se* does not emerge, but rather comes to be instantiated in a process or entity that emerges in time[c]. (For example, the property of being alive as such does not emerge, but is instantiated in the emergent process of autopoiesis that constitutes living cells[d]). The emergent processes of concern to us arise in the case of large ensembles that form 'non-decomposable' systems[e,f], in which the nonlinear interactions of the components result in whole-system processes that have a global-to-local ('downward') determinative influence on those components. Nonlinearity is the result of positive and negative feedback relationships. These relationships account for the determination of patterns of behavior, seen as constrained alternatives in the space of all possible global states of the system (i.e. as attractors in phase space). The emergent whole-system processes are 'morphodynamical'[g], in the sense that they determine the system's identity through time, as a unity having its own proper domain of interactions and whose properties and behaviors can be physically realized in multiple ways. The paradigm of such an emergent identity[h] is the autopoiesis of the living cell[d], but includes also neural assemblies[i,j], and immune networks[k].

Some philosophers reject the notion of downward causation because they find it irreconcilable with the causal closure of the physical domain[l,m]. But the notion of emergent processes as collective variables or order parameters does not seem to succumb to this problem. One might object that although we need to appeal to global order parameters for explanatory purposes, they are not genuinely *causal*, because all causation is local, efficient causation. But this objection begs the question of whether 'cause' can also legitimately mean 'standing' or 'structuring' causes in the form of order parameters operating as 'context-sensitive constraints' that modify a system's phase space or the probability distribution of events in that space[n].

An open question about emergence in complex systems is whether they involve 'relational holism', that is, relationships that are not reducible to the intrinsic features of the components[o]. The paradigm case of relational holism is 'quantum entanglement'[p], but it has been suggested that the phenomena studied in nonlinear dynamical systems theory might also qualify[o]. If this were true, then emergence and downward causation in complex systems would violate not simply the methodological doctrine of microphysical reduction, but the ontological doctrine that a whole supervenes entirely on the intrinsic properties of its parts (known as 'mereological supervenience').

References

a Kelso, J.A.S. (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior*, MIT Press
b Holland, J. (1999) *Emergence: From Chaos to Order*, Perseus
c Hattiangadi, J. The emergence of minds in space and time. In *The Mind as a Scientific Object* (Johnson, D. and Erneling, C., eds), Oxford University Press (in press)
d Maturana, H.R. and Varela, F.J. (1980) *Autopoiesis and Cognition: The Realization of the Living*, Boston Studies in the Philosophy of Science (Vol. 42.), D. Reidel
e Wimsatt, W. (1986) Forms of aggregativity. In *Human Nature and Natural Knowledge* (Donagan, A. *et al*., eds), pp. 259–291, D. Reidel
f Bechtel, W. and Richardson, R.C. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton University Press
g Petitot, J. (1995) Morphodynamics and attractor syntax: constituency in visual perception and cognitive grammar. In *Mind as Motion: Explorations in the Dynamics of Cognition* (Port, R.F. and Van Gelder, T., eds), pp. 227–281, MIT Press
h Varela, F.J. (1997) Patterns of life: intertwining identity and cognition. *Brain Cognit.* 34, 72–87
i Bressler, S. and Kelso, J.A.S. (2001) Cortical coordination dynamics and cognition. *Trends Cognit. Sci.* 5, 26–36
j Varela, F.J. *et al.* (2001) The brain web: phase-synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239
k Varela, F.J. and Coutinho, A. (1991) Second generation immune networks. *Immunol. Today* 12, 159–166
l Kim, J. (1993) *Supervenience and Mind*, Cambridge University Press
m Kim, J. (1998) *Mind in a Physical World: An Essay on the Mind–Body Problem and Mental Causation*, MIT Press
n Juarero, A. (1999) *Dynamics in Action: Intentional Behavior as a Complex System*, MIT Press
o Silberstein, M. and McGeever, J. (1999) The search for ontological emergence. *Philos. Q.* 49, 182–200
p Teller, P. (1986) Relational holism and quantum mechanics. *Br. J. Philos. Sci.* 37, 71–81

through changes in control parameters and boundary conditions, rather than through the interacting dynamical variables. Coherent collective behaviors, called 'collective variables' or 'order parameters'[10,25,26], constrain or prescribe the behavior of the individual components, 'enslaving' them, as it were, so that they no longer have the same behavioral alternatives open to them as would be the case if they were not interdependently linked in the system. On the other hand, the behavior of the components generates and sustains the order parameters. Although usually called 'circular causality', this reciprocal (but not symmetrical) relationship between local and global levels seems better described as 'reciprocal causality'.

Downward causation: consciousness and neural dynamics

Given that the coupled dynamics of brain, body and environment exhibit self-organization and emergent processes at multiple levels[26,27], and that emergence involves both upward and downward causation, it seems legitimate to conjecture that downward causation occurs at multiple levels in these systems, including that of conscious cognitive acts in relation to local neural activity. Indeed, this point has been noted by authors concerned with the dynamical-systems approach to cognition[25–30]. For instance, Freeman describes consciousness as an order parameter and 'state variable-operator' in the brain that 'mediates relations among neurons… Far from being an epiphenomenon… it must play a crucial role in intentional behavior. It is the task of the neurodynamicist to define and measure what that role is'(Ref. 29, p. 132).

Clearly, the issue of downward causation is not only philosophical (see Box 1) but empirical and experimental. If conscious cognitive acts are emergent phenomena, then accordingly we can hypothesize that they have causal effects on local neuronal activity, and therefore that it might be possible to observe the effects of a moment of consciousness and its substrate large-scale neural assemblies at the level of local properties of neuronal activity. Let us mention two cases that can be used to address this point.

Human epileptic activity
It is clear that epileptic activity modifies the subject's mental competencies. The converse, although less documented, seems also to be the case: the subject can voluntarily affect his or her electrical condition that normally would lead to a seizure. As long ago as 1954, Penfield and Jasper described the blocking of a parietal seizure by the initiation of a complex mathematical calculation[31], and recently more extensive observations have confirmed such cognitive influences[32]. One can assume that such intervention is possible because the epileptogenic zones are embedded in a complex network of other brain regions that actively participate in mental life. Such network interactions are multiple and distributed over a large scale[11]. It follows that the global pattern of integration (the result of upward causation) can produce downward effects, acting eventually at the local level of epileptogenic zones, whose activity can thus be taken as an indicator of the downward influence.

More precisely, the question is whether a particular cognitive task or state can manifest as a specific effect in the local activity given by an epileptic discharge, when seen at a sufficient level of detail. Studies by Varela and colleagues have shown that there are deterministic temporal patterns within the apparently random fluctuations of human epileptic activity, and that these patterns can be modulated during cognitive tasks[33,34]. Analysis of the sequence of intervals between spike discharges from the epileptic focus in a particular patient showed that the spikes display a distinct periodic activity for a short time before they progress to a different temporal pattern. This periodic activity was differentially modulated during perceptual tasks engaged in by the patient; and the modulation was carried by frequencies in the gamma range (30–70 Hz)(see Box 2). These findings strongly suggest that the act of perception on the part of the patient contributes in a highly specific manner, via the phase synchrony of its associated neural assembly (the order parameter), to 'pulling' the epileptic activities towards particular unstable periodic orbits. Whether such cognitive 'control' can be harnessed in a therapeutic direction[32] is an open question, but these studies might point to a possible physiological basis of such interventions.

Voluntary perceptual reversals
Another case worth exploring, especially in the light of recent studies of binocular rivalry[35,36], is the visual experience of multistable or ambiguous figures. Models of multistable perception have suggested that it is based on generic properties of coupled nonlinear oscillators and their phase relationships[37], and aspects of this dynamics have been tracked to various neural levels[11,35,36]. We conjecture, on the basis of these studies, that different 'cognitive' interpretations of ambiguous figures initiated by the subject might shift the neuronal bias that defines the perceptual reversal. This conjecture could eventually be tested by studying human clinical subjects with implanted electrodes (like the patient described in Box 2) as they perform *voluntary* perceptual reversals of ambiguous figures.

From emergence to embodiment
Having suggested that there are reciprocal causal-explanatory relationships between neural events and conscious events when the latter are conceived of as

---

Box 2. Cognitive modulation of epileptic activity: a case of downward causation?

Varela and colleagues studied recordings from an implanted subdural electrode grid, in a patient ('Mr T') with an unusually focal and stable occipito-temporal epileptic discharge[a,b]. Here we review some of the main findings as they bear on the issue of downward causation in cognitive neurodynamics.

Mr T suffered from a partial drug-resistant temporal epilepsy. A subdural grid of 6×6 electrodes was apposed over the convexity of the temporo-occipital junction (Fig. Ia). Surprisingly, only one electrode (no. 9) presented a permanent discharge around 2 Hz in a very restricted region. Mr T showed no evidence of cognitive impairment, and was willing to participate in simple cognitive tests of visual and auditory discrimination following an odd-ball protocol. For the visual task, he was asked to press a button when the target stimulus (a non-triangle made up of three equidistant everted 'pac-men') appeared, but not when two other stimuli were shown (a Kanizsa triangle formed by the edge completion of three 'pac-men' and a normal triangle drawn with lines).

Short-term regularities of interspike intervals

The first study[a] began by considering the sequence of intervals between spike discharges from the epileptic focus given by the time elapsed between two successive peaks of the activity (Fig. Ia). This time series exhibited considerable variability around a mean, as clearly seen when plotted in a cumulative histogram (Fig. Ib). At first glance, the epileptic focus appeared to be a regular 2 Hz oscillation contaminated with Gaussian random noise. But closer inspection in the form of a 'first-return map' (or scatter plot) revealed a deterministic pattern.

A first-return map plots the values of two successive intervals, T_{n+1} against T_n (Fig. Id). The diagonal line from bottom left to top right is the locus where two intervals are identical ($T_n = T_{n+1}$); an accumulation of points here would indicate periodic activity. By contrast, if the density appears as a dispersed cloud of points, the probability of interval T_{n+1} is independent of the value of the previous interval T_n, and thus there is no trace of a regular or correlated activity. Visual examination of the first-return plots for the various conditions reveals regions of high densities, in which points accumulate into groups clustered around discrete positions along the diagonal, as well as clusters of points at outlying positions away from the diagonal.

Next was estimated the density of points located within a small circle (radius of 8 ms) centered at a particular value T of the diagonal. This density tells us the probability that pairs of successive intervals will remain around a chosen value of T. As shown in Fig. Ic, not all values of T have similar probabilities, signifying that successive spikes are correlated, because recurrence times fall into three evenly spaced distinct peaks, T_1, T_2, T_3. In other words, the spikes display a distinct periodic activity for a short time before they move to a different temporal pattern.

Modulations reflected in the gamma-range

The clusters of points representing short-term periodicities can also be separated according to the specific mental state the subject is undergoing. Analysis revealed that the various clusters are differentially modulated during perceptual conditions by periodicity shifts of around 15–30 ms. Such periodicities correspond in the frequency domain to the gamma-range of 30–70 Hz. Correspondingly, an increased gamma-band coherence was found between the epileptic focus and surrounding recording sites. These synchronies were shown to be precisely linked to the perceptual appearance of the stimuli in the 200–400 ms range, and were not present except during target presentation.

Taken together, these results confirm that, from the periodic shifts seen in the first return map, distinct cognitive processes can be related to gamma-band activity. Gamma-band activity has been widely reported to be involved in cognitive processes in general[c,d]. Although a causal relationship between the interspike intervals and cortical coherence in the gamma band was not directly demonstrated, the findings strongly suggest that cognitive tasks modulate epileptic activity through increased coherences in this frequency range.

Tracing the downward effect to its dynamical skeleton

These results from the first study of Mr T were analyzed further in another study[b]. This addressed whether the observed shifts along the diagonal result from a precise geometry of neural activity in the form of 'unstable periodic orbits' (UPOs) (Ref. e). To answer this question we need to focus in on a first-return map to find points that contain sequences of spikes approaching the diagonal periodicity, only to remain there briefly before moving off to another periodicity.

A typical example of recurrent observed trajectories is shown for point T_2 in Fig. Ic,d, displaying the sequence of points numbered 717–721. From points 717–719, the state of the system is drawn towards the thick diagonal along a specific linear slope, but after point 719 it diverges away to a diagonal with a different slope. These direct observations revealed that the trajectories around the fixed points look very much like motion near a fixed point of the saddle-focus type, with stable and unstable manifolds that intersect at the fixed point (a generic property independent of the periodicity). In particular, the approaches of trajectories towards the unstable periodic patterns are observed in the sequences immediately following the perceptual tasks.

These findings suggest that the act of perception contributes in a highly specific manner to 'pulling' the epileptic activities towards particular unstable periodic orbits, a clear candidate for downward causation. Mr T provided the first evidence for the existence and cognitive modulation of such UPOs in the human brain.

The fundamental nature of such a dynamical description is confirmed by the fact that UPOs are found at many biological levels[e]. Relevant to the issue of downward causation is evidence that acting

order parameters of large-scale brain dynamics, we now suggest that the processes crucial for consciousness cut across the brain–body–world divisions, rather than being limited to neural events in the head.

Although it is often tacitly assumed that consciousness must 'supervene' entirely on internal neural states, it is far from clear how one is supposed to distinguish between 'internal' and 'external' states. Despite the philosophical fiction of a 'brain-in-a-vat',
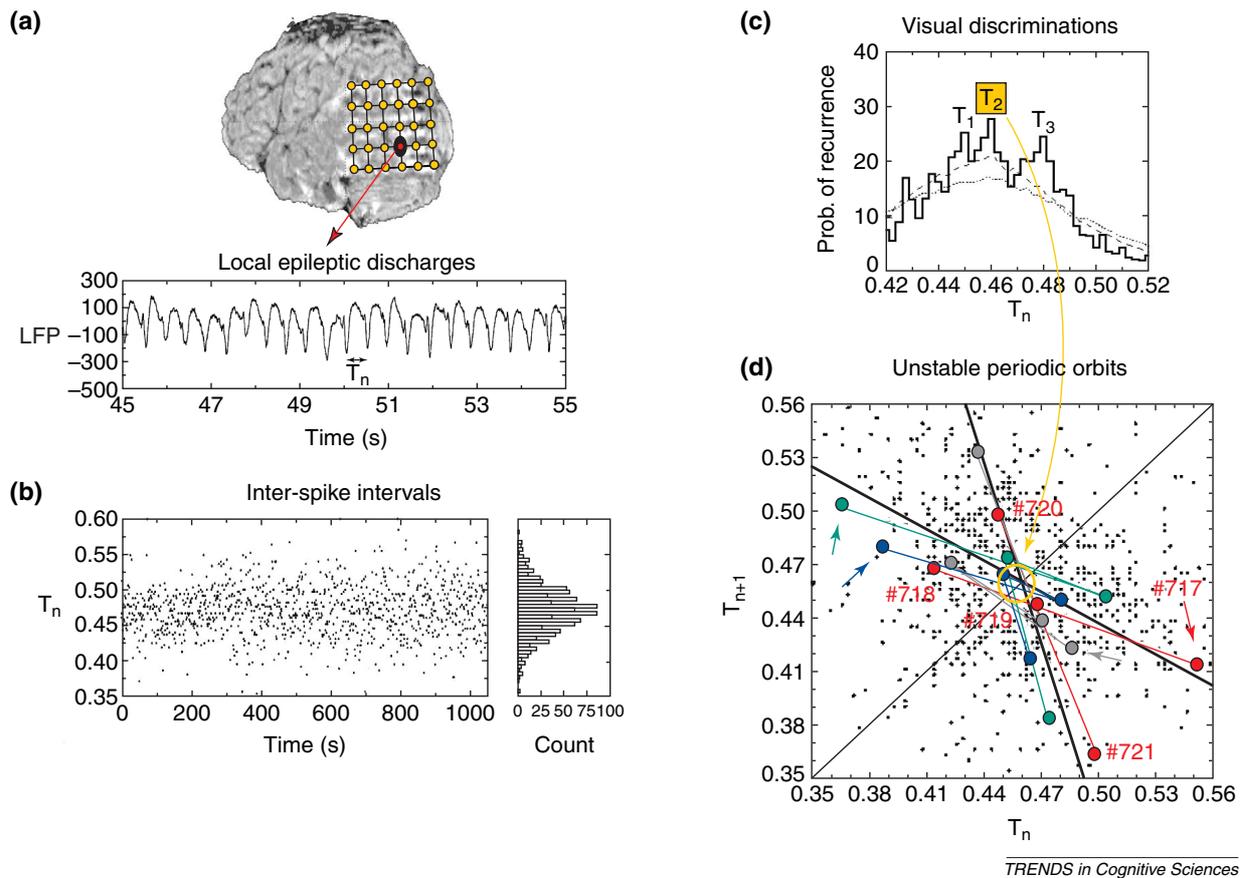
Opinion

Fig. I. Perceptual modulation of human epileptic activity. (a) Surface-reconstruction showing the artifacts due to the metal induction of the electrode grid, and the projection of electrode no. 9 from which the permanent discharge was recorded. Ten seconds of the local field recording from the subdural electrode in the epileptogenic region is shown below. Highly stereotyped and nearly periodic epileptic discharges are continuously observed. The spikes were used as natural time markers to obtain a time series of interspike intervals, T_n. (b) The full data set of T_n for the resting condition (passive visual fixation). To the right, a histogram of the data set corresponds to a unimodal normal distribution of the interval. The other conditions, auditory odd-ball discrimination, visual odd-ball discrimination, and hyperventilation, gave results that were not significantly different. (c) Density of points in the return map (d) located within the small yellow circle (radius 8 ms) around a particular value of T of the diagonal. Three main probability peaks T_1, T_2, T_3 are marked. (d) First-return map showing variations of one particular fixed point, $T_2 = 0.465$ s. The starting points are indicated by arrows. Multiple trajectories (here color coded) that did not follow each other in time approach the neighborhood in a stable direction and diverge away in a different unstable direction (e.g. see T_2, points 717–720, red circles). The stable and unstable manifolds are estimated by straight line fits to all the visitations, and are indicated by heavy black lines (the more horizontal of the two referring to the unstable manifold, and the more vertical to the stable manifold). Modified from Refs 33,34.

so as to modify the dynamics at this level is causally effective, for instance in the control of cardiac arrhythmias[f] and the control of epileptic discharges *in vitro*[g]. Hence downward causation need be no metaphysical 'will-o'-the-wisp', but can be made an empirically tangible issue. Indeed, it has implications for therapeutic interventions (see Ref. b), including giving a neurophysiological rationale for exploring the strategy of 'cognitive' control of epileptic seizures, still a marginal subject of study[h].

References
a Le Van Quyen, M. *et al.* (1997) Temporal patterns in human epileptic activity are modulated by perceptual discriminations. *NeuroReport* 8, 1703–1710
b Le Van Quyen, M. *et al.* (1997) Unstable periodic orbits in human epileptic activity. *Phys. Rev. E* 56, 3401–3411
c Rodriguez, E. *et al.* (1999) Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397, 430–433
d Varela, F.J. *et al.* (2001) The brain web: phase-synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239
e So, P. *et al.* (1998) Periodic orbits: a new language for neuronal dynamics. *Biophys. J.* 74, 2776–2785
f Glass, L., ed. (1995) *Theory of Heart: Biophysics, Non-Linear Dynamics, and Clinical Applications*, Springer-Verlag
g Shiff, S.J. *et al.* (1994) Controlling chaos in the brain. *Nature* 370, 615–620
h Schmid-Schonbein, C. (1998) Improvement of seizure control by psychological methods in patients with intractable epilepsies. *Seizure* 7, 261–270

it is doubtful (even as a thought experiment) that one can 'peel away' the body and the environment as 'external' to the brain processes crucial for consciousness. The nervous system, the body and the environment are highly structured dynamical systems, coupled to each other on multiple levels[38]. Because they are so thoroughly enmeshed – biologically, ecologically and socially – a better conception of brain, body and environment would be as mutually embedded systems rather than as

Questions for future research

- Can long-range synchronization in the vertebrate brain be shown to play a causal role in behavior or consciousness?
- Can 'downward' causal relationships between large-scale dynamical correlates of conscious cognitive acts and local neuronal activity be experimentally demonstrated?
- Do nonlinear relationships in complex dynamical systems involve a form of relational holism?
- Can we show, as has been suggested here, that the contents of consciousness depend crucially on the embedding of brain dynamics in the somatic and environmental context of the animal's life?

internally and externally located with respect to one another. Neural, somatic and environmental elements are likely to interact to produce (via emergence as upward causation) global organism–environment processes, which in turn affect (via downward causation) their constituent elements. Although speculative, these points gain plausibility from considering the dimensions of embodiment.

Dimensions of embodiment
The relationship between neural dynamics and conscious situated agents can be described in terms of the participation of neural processes in the 'cycles of operation' that constitute the agent's life. Three kinds of cycles need to be distinguished for higher primates:

(1) cycles of organismic regulation of the entire body;

(2) cycles of sensorimotor coupling between organism and environment;

(3) cycles of intersubjective interaction, involving the recognition of the intentional meaning of actions and linguistic communication (in humans).

We will consider each of these in turn.

Organismic regulation
The participation of the brain in organismic regulation is diffuse and multifaceted. Its main basis is the autonomic nervous system, in which sensors and effectors to and from the body link neural processes to basic homeodynamic processes of the internal organs and viscera. Emotional states – reflecting the links between the autonomic nervous system and the limbic system via the hypothalamus – are part and parcel of homeodynamic regulation[39,40]. In the brain-stem, nuclei that regulate homeostasis are interconnected with nuclei that regulate sleep, wakefulness and arousal. Numerous mutual interactions between brain and body exist at biochemical levels, especially in the molecular components of the endocrine, immune and nervous systems. The integrity of the entire organism depends on such regulatory cycles involving brain and body at multiple levels.

Organismic regulation, because of its links with basic emotional operating systems in the mammalian brain, has a pervasive affective dimension that manifests in the range of affective behaviors and feelings typical of mammalian life[39]. This affective dimension of organismic regulation, variously called 'primal consciousness'[39] or 'core consciousness'[40], could be called sentience – the feeling of being alive – understood as the inescapable affective backdrop of every conscious state[41] (inescapable because of the need for every conscious state to be grounded on the regulatory and emotional states of the whole organism). Sentience seems not to be organized according to sensory modality, but according to the regulatory and affective processes that constitute the organism's feeling of self[40].

Sensorimotor coupling
Situated activity takes the form of cycles of sensorimotor coupling with the environment. What the organism senses is a function of how it moves, and how it moves is a function of what it senses[42–44]. The substrates of these cycles are the sensorimotor pathways of the body, which are mediated in the brain by multiple neocortical regions and subcortical structures. Transient neural assemblies mediate the coordination of sensory and motor surfaces, and sensorimotor coupling with the environment constrains and modulates this neural dynamics[38]. It is this cycle that enables the organism to be a situated agent[45].

Intersubjective interaction
The signalling of affective state and sensorimotor coupling play a huge role in social cognition, especially in apes and humans. Higher primates excel at interpreting others as psychological subjects on the basis of their bodily presence (facial expressions, posture, vocalizations, etc.)[46]. Neural structures that are known to be important in social cognition (amygdala, ventromedial frontal cortices, and right somatosensory related cortices) are also important to emotion[47]. A crucial ingredient of both seems to be 'feeling', in the sense of affective experiences linked to one's own emotional state or to perceiving the emotional states of others through empathy[40,47].

Intersubjectivity involves distinct forms of sensorimotor coupling, as seen in the so-called 'mirror neurons' discovered in area F5 of the premotor cortex in monkeys[48]. These neurons display the same pattern of activity both when the animal accomplishes certain goal-directed hand movements, and when the animal observes a conspecific (or the experimenter) performing the same actions. The recognition of the intentional meaning of actions in others apparently depends on patterns of neural activity in premotor areas that are similar to those internally generated to produce the same type of action. There is evidence for a mirror-neuron system for gesture recognition in humans[48], and it has been proposed that this system might be part of the neural basis for the development of language[49].

Consciousness embodied and embedded

Based on this overview, we conjecture that consciousness depends crucially on the manner in which brain dynamics are embedded in the somatic and environmental context of the animal's life, and therefore that there may be no such thing as a minimal internal neural correlate whose intrinsic properties are sufficient to produce conscious experience. Clearly, this conjecture requires further theoretical analysis and empirical study.

Conclusion

Our aim has been to suggest an 'enactive' or 'radical-embodiment' approach to the neuroscience of consciousness. We propose that there are two-way or reciprocal relationships between neural events and conscious activity. An attractive feature of this proposal is that it allows consciousness to be a causally efficacious participant in the cycles of operation constituting the agent's life (see also Ref. 30). Testing this proposal will require studies of neural dynamics that are not simply correlative, but causal. We also propose that the processes crucial for consciousness cut across the brain–body–world divisions rather than being located simply in the head. Evaluating this proposal could set the agenda for embodied cognitive science in coming years[50].

References

1 Clark, A. (1997) *Being There: Putting Brain, Body, and World Together Again*, MIT Press
2 Clark, A. (1999) An embodied cognitive science? *Trends Cognit. Sci.* 3, 345–351
3 Varela, F.J. *et al.* (1991) *The Embodied Mind: Cognitive Science and Human Experience,* MIT Press
4 Kanwisher, N. (2001) Neural events and perceptual awareness. *Cognition* 79, 83–113
5 Chalmers, D.J. (2000) What is a neural correlate of consciousness? In *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (Metzinger, T. ed.), p. 31, MIT Press
6 Damasio, A. (1990) Synchronous activation in multiple cortical regions: a mechanism for recall. *Semin. Neurosci.* 2, 287–297
7 Varela, F.J. (1995) Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biol. Res.* 28, 81–95
8 Friston, F.J. (1997) Another neural code? *NeuroImage* 5, 213–220
9 Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. *Science* 282, 1846–1851
10 Bressler, S.L. and Kelso, J.A.S. (2001) Cortical coordination dynamics and cognition. *Trends Cognit. Sci.* 5, 26–36
11 Varela, F.J. *et al.* (2001) The brain web: phase-synchronization and large-scale integration, *Nat. Rev. Neurosci.* 2, 229–239
12 Lachaux, J-P. *et al.* (1999) Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208
13 Roskies, A.L. (ed.) (1999) The Binding Problem (Special Issue) *Neuron* 24, 7–125
14 Gray, C.M. *et al.* (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338, 334–337
15 König, P. *et al.* (1995) Relation between oscillatory activity and long-range synchronization in cat visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 92, 290–294
16 Bressler, S.L. *et al.* (1993) Episodic multiregional cortical coherence at multiple frequencies during visual task performance. *Nature* 366, 153–156
17 Roelfsema, P.R. *et al.* (1997) Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* 385, 157–161

18 Rodriguez, E. *et al.* (1999) Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397, 430–433
19 Srinivasan, R. *et al.* (1999) Increased synchronization of neuromagnetic responses during conscious perception. *J. Neurosci.* 19, 5435–5448
20 von Stein, A. *et al.* (2000) Top-down processing mediated by interareal synchronization. *Proc. Natl. Acad. Sci. U. S. A.* 97, 14748–14753
21 Lutz, A. *et al.* Preparation strategies for visual perception: a study of endogenous neural synchronies. *Proc. Natl. Acad. Sci. U. S. A.* (in press)
22 Crick, F. and Koch, C. (1998) Consciousness and neuroscience. *Cereb. Cortex* 8, 97–107
23 Kopell, N. (2000) We got rhythm: dynamical systems of the nervous system. *Notices AMS* 47, 6–16
24 Cambell, D.T. (1974) 'Downward causation' in hierarchically organised biological systems. In *Studies in the Philosophy of Biology* (Ayala, F.J. and Dobzhansky, T., eds), pp. 179–186, University of California Press
25 Haken, H. and Stadler, M., eds (1990) *Synergetics of Cognition*, Springer-Verlag
26 Kelso, J.A.S. (1995) *Dynamic Patterns: The Self–Organization of Brain and Behavior,* MIT Press
27 Port, R.F. and Van Gelder, T., eds (1995) *Mind as Motion: Explorations in the Dynamics of Cognition*, MIT Press
28 Sperry, R.W. (1991) In defense of mentalism and emergent interaction. *J. Mind Behav.* 12, 221–245
29 Freeman, W. (1999) *How Brains Make Up Their Minds*, Wiedenfeld & Nicholson
30 Juarrero, A. (1999) *Dynamics in Action: Intentional Behavior as a Complex System*, MIT Press
31 Penfield, W. and Jasper, H. (1954) *Epilepsy and the Functional Anatomy of the Human Brain*, Little, Brown & Co.
32 Schmid-Schonbein, C. (1998) Improvement of seizure control by psychological methods in patients with intractable epilepsies. *Seizure* 7, 261–270
33 Le Van Quyen, M. *et al.* (1997) Temporal patterns in human epileptic activity are modulated by perceptual discriminations. *NeuroReport* 8, 1703–1710
34 Le Van Quyen, M. *et al.* (1997) Unstable periodic orbits in human epileptic activity. *Phys. Rev. E* 56, 3401–3411

35 Leopold, D. and Logothetis, N.K. (1999) Multistable phenomena: changing views in perception. *Trends Cognit. Sci.* 3, 254–263
36 Wilson, H.R. *et al.* (2001) Dynamics of travelling waves in visual perception. *Nature* 412, 907–910
37 Kelso, J.A.S. *et al.* (1994) Multistability and metastability in perceptual and brain dynamics. In *Ambiguity in Mind and Nature: Multistable Cognitive Phenomena* (Springer Series in Synergetics, Vol. 64) (Stadler, M. and Kruse, P., eds), pp. 159–184, Springer-Verlag
38 Chiel, H. and Beer, R. (1997) The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends Neurosci.* 20, 553–557
39 Panskepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press
40 Damasio, A. (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt Brace
41 Watt, D. (1998) Emotion and consciousness: implications of affective neuroscience for extended reticular thalamic activating system theories of consciousness. *The Association for the Scientific Study of Consciousness: Electronic Seminar on Emotion and Consciousness*, Sept. 21–Oct. 9, 1998; http://www.phil.vt.edu/assc/esem.html
42 Maturana, H.R. and Varela, F.J. (1987) *The Tree of Knowledge: The Biological Roots of Human Understanding*, Shambala/New Science Library
43 Hurley, S. (1998) *Consciousness in Action*, Harvard University Press
44 O'Regan, J.K. and Noë, A. A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* (in press)
45 Varela, F.J. (1997) Patterns of life: intertwining identity and cognition. *Brain Cognit.* 34, 72–87
46 Povinelli, D.J. and Preus, T.M. (1995) Theory of mind: evolutionary history of a cognitive specialization. *Trends Neurosci.* 21, 188–194
47 Adolphs, R. (1999) Social cognition and the human brain. *Trends Cognit. Sci.* 3, 469–479
48 Gallese, V. and Goldman, A. (1999) Mirror neurons and the simulation theory of mind-reading. *Trends Cognit. Sci.* 2, 493–500
49 Rizzolatti, G. and Arbib, M. (1998) Language within our grasp. *Trends Neurosci.* 21, 188–194
50 Thompson, E. and Varela, F.J. *Why the Mind Isn't in the Head*, Harvard University Press (forthcoming)