THE ROYAL SOCIETY

The labile brain. II. Transients, complexity and selection

Karl J. Friston

Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, UK
(*k.friston@fil.ion.ucl.ac.uk*)

The successive expression of neuronal transients is related to dynamic correlations and, as shown in this paper, to dynamic instability. Dynamic instability is a form of complexity, typical of neuronal systems, which may be crucial for adaptive brain function from two perspectives. The first is from the point of view of neuronal selection and self-organizing systems: if selective mechanisms underpin the emergence of adaptive neuronal responses then dynamic instability is, itself, necessarily adaptive. This is because dynamic instability is the source of diversity on which selection acts and is therefore subject to selective pressure. In short, the emergence of order, through selection, depends almost paradoxically on the instabilities that characterize the diversity of brain dynamics. The second perspective is provided by information theory.

Keywords: neuronal transients; complexity; functional integration; neural codes; selection; self-organization

1. INTRODUCTION

This paper reviews the notion of complexity and how it relates to transients and dynamic instabilities in neuronal systems. In § 2, it relates neuronal transients to nonlinear dynamical concepts such as intermittency, itinerancy and dynamic instability. This section introduces the distinction between different sorts of complexity (type I and II), which is useful when considering complexity and diversity in relation to selective mechanisms that may operate in the brain. After considering the genesis of complexity, § 3 addresses the role of asynchronous or nonlinear coupling. The strengths of connections, among simulated populations, are manipulated to induce changes in (i) the nature of the coupling and (ii) the complexity of the ensuing dynamics. This allows the relative contributions of synchronous and asynchronous coupling to complexity to be characterized. In brief, we will show that complexity and nonlinear coupling go hand in hand, presiding in regimes of sparse connectivity. The importance of complexity for self-organization (Kelso 1995) and the selective consolidation of synaptic connections in terms of neuronal selection (Edelman 1993) are introduced in § 4. In this section it is suggested that selective mechanisms of a high order are sufficient to explain why the brain expresses complicated dynamics.

2. A DYNAMICAL PERSPECTIVE

(a) *Complexity*

In this section we consider transients in relation to the complexity of the dynamics that they generate. Complexity is itself a complex field with numerous definitions and perspectives (Horgan 1995). Generally, complexity refers to something in the behaviour of a system that is not ordered or predictable nor chaotic or random but something in between that reflects an underlying order that, is itself, inherently unstable or labile. There are two distinct approaches to complexity: those that derive from information theory and those that come from the field of deterministic chaos in nonlinear systems. The former approaches are based on some measure of the entropy of the system (e.g. Morgera 1985) and can be related to algorithmic complexity (framed in terms of the minimum length of an algorithm required to generate an observed time-series). More recently, entropy-based complexity measures have been proposed that try to capture the balance between integration among different neuronal systems and the preservation of information that is unique to them. Functional segregation requires the dynamics of each area to be distinct, in terms of their intrinsic activity and responses to input. Functional integration, on the other hand, requires segregated areas to influence each other in a way that facilitates coherent integration. It has been proposed that the resolution of this dialectic, between the preservation of regionally specific dynamics and global coherence, is a hallmark of complexity (Tononi *et al.* 1994; Friston *et al.* 1995). A measure of this complexity, based on the theory of stochastic processes and information theory, is found in Tononi *et al.* (1994).

In this paper we are concerned with the second approach to complexity, namely that predicated explicitly on nonlinear dynamic systems. Within this class there is another dichotomy that distinguishes between dimensional complexity in chaotic systems and dynamic instability (Kelso 1995) associated with self-organizing and pattern-forming systems. Dimensional complexity is a measure (the correlation dimension) that reflects the degree of chaos in terms of the average local behaviour of

Received 2 September 1998 *Accepted* 8 June 1999

the system's evolution (in particular the exponential divergence of trajectories). It is closely related to the Lyapunov exponents and through this to the Kolmogrov entropy (Tsonis 1992). However, this is not the sort of complexity that people are generally interested in when thinking about biological or self-organizing systems. The critical sort of complexity is that which we would intuitively appreciate as complex, namely the successive expression of different transient dynamics with stereotyped temporal patterns being continuously created and destroyed and re-emerging again. In the context of the brain, this is simply the expression of neuronal transients that arise when cell populations interact. Populations may interact in a synchronous way (e.g. phase-locking among units (e.g. Gray & Singer 1989) or populations (e.g. Sporns *et al.* 1989)), creating spatio-temporal patterns of activity that include many, if not all, of the system's components. Generally the ensuing coherent states are labile, self-limiting and short lived. Alternatively, complexity could arise from incoherent, asynchronous coupling that engenders very different transients in different populations.

In summary, complexity (of a dynamic sort) can be divided into dimensional complexity and dynamic instability. The former is a measure of chaos and the latter directly characterizes what makes a system complicated. From now on we will use complexity, dynamic instability and metastability (Kelso 1995) synonymously. For the purposes of discussion, we will distinguish between two different sorts of dynamic instability (type I and type II complexity) and illustrate them using the neuronal simulations described in paper 1 (Friston, this issue). To do this we need to introduce the notion of an attractor. Consider equation (1) in Friston (paper 1, this issue) and assume that we know all the relevant state variables $\boldsymbol{x}$.

$$\partial \boldsymbol{x}(t)/\partial t = f(\boldsymbol{x},\boldsymbol{C}). \quad (1)$$

As time goes on these variables will change. If we plotted these variables against each other (in a state space) they would trace a path or trajectory as the system evolved. An attractor is simply the surface or 'manifold' over which this trajectory courses. Clearly the dimension of the state space, in which the attractor manifold is embedded, is equal to the number of variables considered (the dimensionality of the manifold is, in fact, the dimensional complexity referred to above). The shape of the manifold will dictate the nature of the associated dynamics. The control parameters ($\boldsymbol{C}$ in equation (1)) will in turn dictate the shape the manifold. To see this, imagine that the trajectory, defined by the coordinates of the vector $\boldsymbol{x}$, is a point flowing through state space. The direction taken at each point is determined by a 'flow field' where the direction of flow is specified by equation (1). The flow field is determined only by the control parameters $\boldsymbol{C}$. This flow field in turn configures the shape of the attractor manifold. Dynamic instability suggests that this manifold is itself unstable. The distinction between type I and type II complexity rests on whether this instability arises from apparent changes in the manifold, as the trajectory explores different parts of it, or is mediated by changes in the control parameters.

(b) *Type I complexity*

This form of dynamic instability could be likened to 'intermittency' in simple nonlinear systems (Tsonis 1992) or 'itinerancy'. These phenomena reflect an apparent change in the attractor manifold that arises when the trajectory moves from one part of the manifold to another. The basic idea here is that the trajectory gets trapped in some local submanifold leading to distinct dynamics associated with that part of the attractor surface. Inevitably, after some time, the trajectory will escape to another submanifold to express a new transient-like behaviour. This wandering or 'itinerant' behaviour gives rise to dynamic instability despite the fact the that global manifold never actually changes. In other words, there is an apparent change in the manifold that is due to the itinerant nature of the trajectory getting stuck in local submanifolds. Because the attractor never changes, the underlying control parameters (e.g. connection strengths) are likewise invariant and this is what defines type I complexity (i.e. dynamic instability due to a complex attractor manifold that supports an itinerant trajectory). Perhaps the simplest illustration of this sort of behaviour is inherent in the Lorenz attractor describing the spatio-temporal modes of convection. This famous example (figure 1*a*) has two wings that can, for periods of time, capture the trajectory before it escapes to the other wing. This itinerant wandering between the two parts of the attractor manifold can be seen clearly if one plots one of the three state variables as a function of time (figure 1*b*). A more compelling example can be constructed using neuronal simulations.

(i) *An illustration using short-term plasticity*

Let us assume that the system under investigation is a single population as modelled in the simulations presented in Friston (paper 1, this issue). Here we know all the state variables and, because there are no transmission delays to consider, the entire set of equations governing the dynamics can be considered as an example of equation (1). Burst firing in these simulated cell populations will be used to illustrate type I complexity. Spontaneous bursting behaviour is intermittent where that intermittency is in part mediated by postsynaptic desensitization, engendered by the bursts of activity. Intracortical synapses display several forms of facilitation and depression. One of the most predominant forms of this short-term plasticity is depression that develops over a few action potentials and decays with time constants in the range of 200–600 ms (see Abbot *et al.* (1997) for a discussion of this in relation to cortical gain control). This phenomenon was incorporated into the simulations, described in Appendix B of Friston (paper 1, this issue) by modulating the discharge probability of subpopulation j by $\sigma_1\{-40-E_j\}$ where E_j is a trace of the transmembrane potential V_j obtained by making $\partial E_j/\partial t = (V_j - E_j)/\tau$. The trace can be thought of as modelling some intracellular calcium-dependent effect that depends on a slow build up of Ca^{2+} during prolonged depolarization. $\sigma_1\{-40-E_j\}$ is a sigmoid function that is one when the trace is low and starts to fall as it passes though -40 mV. The time constant for this short-term depression effect was $\tau = 300$ ms.

Typical results evidencing intermittent bursting are seen in figure 2*a*, over a 3 s period, in terms of the simulated

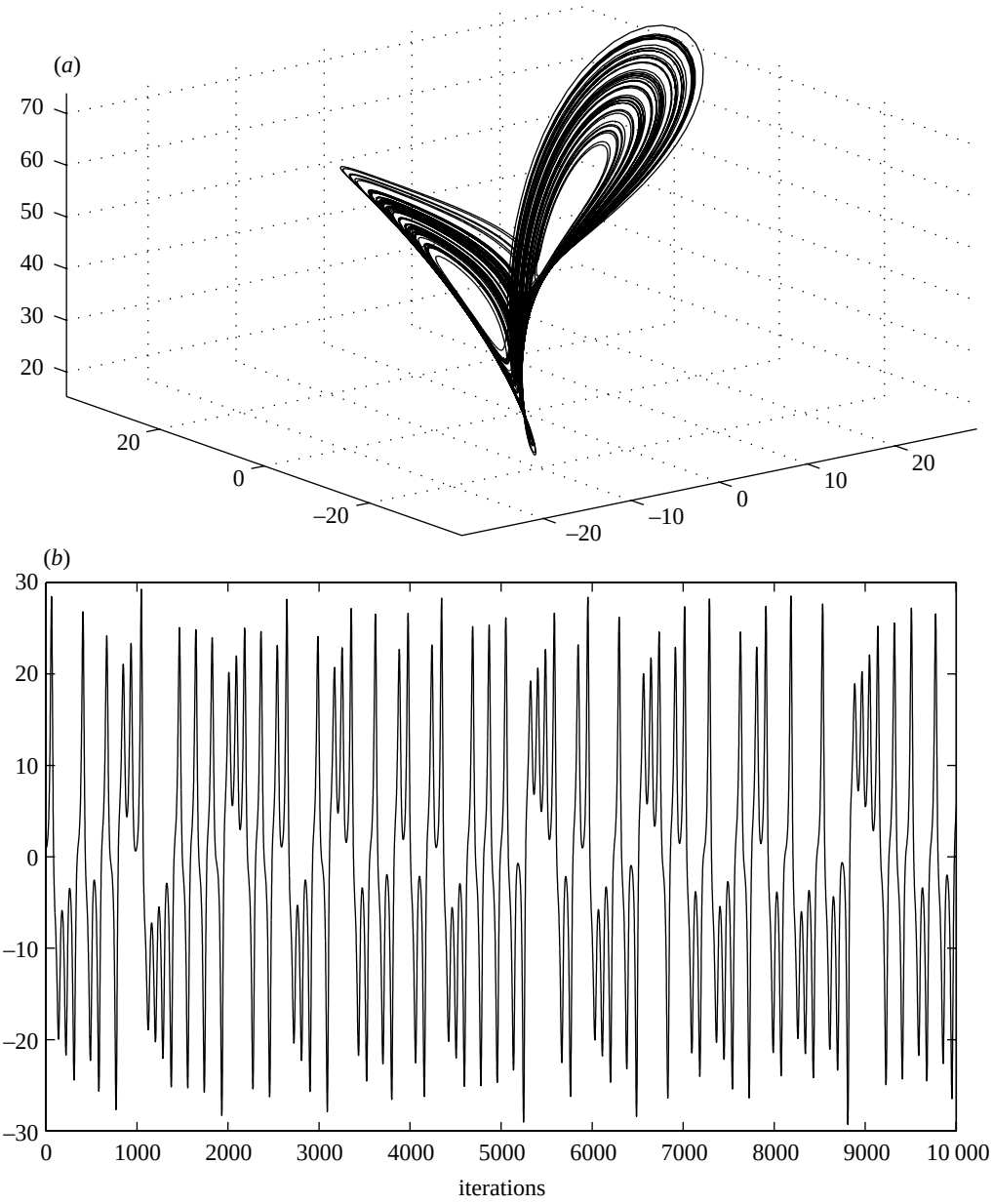

Figure 1. (*a*) The Lorenz attractor, and (*b*) the 'itinerant' dynamics of the first state variable.

local field potential (LFP). This intermittency renders the dynamics complex and, by virtue of the fact that the control parameters (time constants and connection strengths) were constant, has a type I complexity. After a period of bursting, the trajectory falls on to another part of the manifold while the trace of depolarization recovers sufficiently to allow another burst of spontaneous activity. In the next example, we will introduce a new control parameter that is time dependent. Although a similar sort of intermittency is produced the mechanism is very different.

(c) *Type II complexity*

With this sort of complexity dynamic instabilities are introduced by actually changing the attractor manifold through changes in the underlying control parameters. An obvious candidate for a control variable, in relation to a single neuronal population, is some diffuse neuromodulatory effect mediated by afferents of a neuromodulatory ascending neurotransmitter system. In compelling experiments by Munk *et al.* (1996), ascending neurotransmitter systems were activated by stimulation of the mesencephalic reticular formation. This resulted in facilitation of oscillatory activity in the gamma range and enhanced the stimulus-specific synchronization of neuronal spike-trains in the visual cortex of cats. To simulate this we introduced a new, time-dependent, control variable $M(t)$, representing neuromodulatory input, that varied between 0 and 1. This control parameter simply attenuated the closing of open fast excitatory channels. This effect entered into the simulation equations though the time constant governing the first-order kinetics of AMPA channel closure; τ_{AMPA} was replaced by $\tau_{\mathrm{AMPA}}/(1-M(t))$. In figure 2*b*, intermittent bursting engendered by this facilitatory modulatory effect is seen. In this example, $M(t)$ was varied periodically (dotted line) at about 2 Hz.

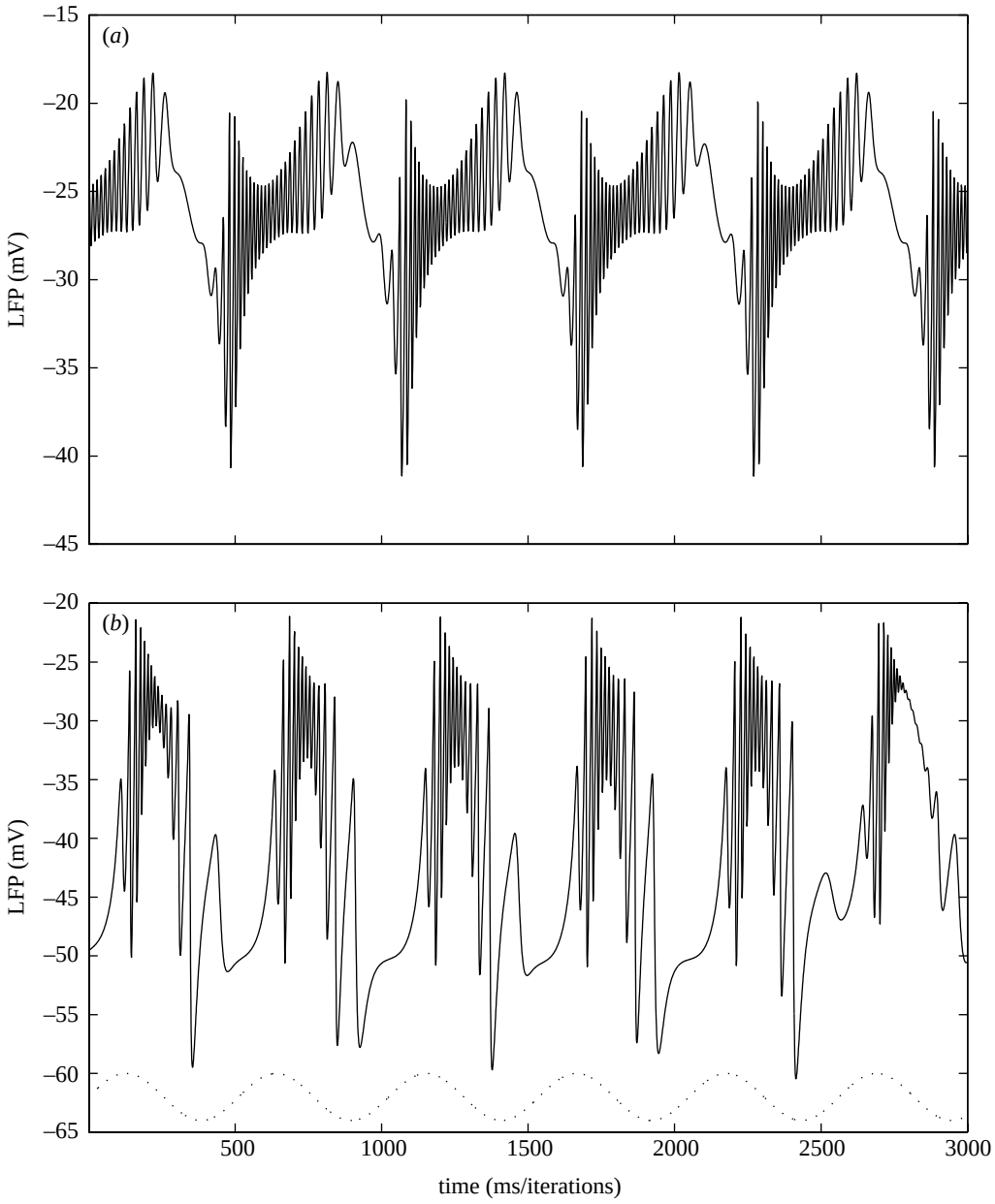

Figure 2. Different sorts of itinerant dynamics or dynamic instabilities. (*a*) The simulated LFP of a single population showing short-term depression (desensitization of postsynaptic responses), modelled as described in the main test. Characteristic here is the intermittent refractoriness and bursting activity that is attributable to the dynamics of some intrinsic state variables (i.e. type I complexity). The AMPA self-excitation connection strength was 0.44 in this simulation. (*b*) Similar intermittent dynamics but in this instance engendered by extrinsic modulatory input emulating a classical neuromodulatory effect on postsynaptic conductances (see main text for details). The sinusoidal modulation $M(t)$ had a period of 512 ms and varied between 0.6 and 1. The time-course of this input is depicted by the dotted line. The self-excitation connection strength was here 0.18 for AMPA-like synapses and 0.022 for NMDA-like synapses.

In this instance, $M(t)$ is a control variable that changes the shape of the attractor manifold and consequent dynamics. But there is an important observation to be made here. If $M(t)$ had come from another simulated population then should it be considered a control variable from the perspective of the first population or as a state variable when considering both populations together? Clearly, ignoring time-delays, in the latter context $M(t)$ would be a state variable where the collective control variables of both populations would be invariant. This is a crucial observation because it suggests that the distinction between type I and type II complexity is simply a matter of perspective. In other words, what may be a type II complexity from the point of view of one system may turn out to be a type I complexity when one 'stands back' and considers a larger system in which the first was embedded. The distinction is, however, critical, as will be shown below, particularly if one considers that it is impossible to stand sufficiently 'far back' in open systems like the brain to get a type I perspective. The point here is that neuronal activities can be construed as state variables or control variables depending on whether they are intrinsic or extrinsic to the neuronal population examined.

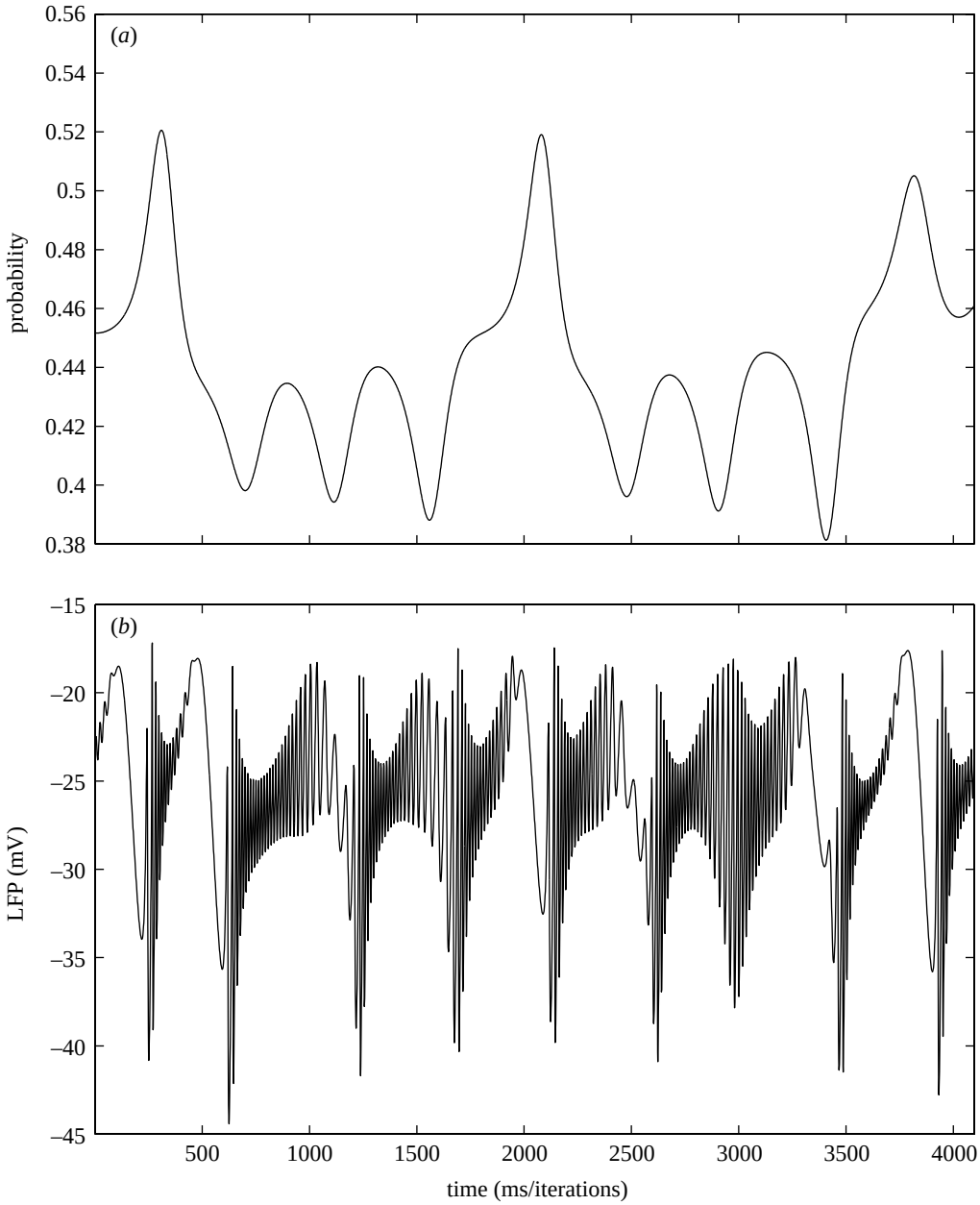

Figure 3. A further example of dynamic instability in a simulated population. In this example the self-excitatory connection strength was used as a dynamic control variable to elicit type II complexity in the ensuing LFP. (*a*) The time-dependent self-excitation obtained by adding 0.44 to the (normalized) state variable of the Lorenz attractor in figure 1. (*b*) The resulting dynamics. It can be seen that as self-excitation becomes too great there is something akin to a depolarization block and the dynamics move out of an oscillatory regime, only to re-emerge when the dynamic connection strength falls sufficiently.

(d) *Attractors within attractors: type I or type II?*

As a final example of how dynamicism in control variables might arise, consider what would happen if we took some control variables (e.g. the self-excitatory connection strengths) from the simulation above and made them dynamic, not by borrowing the outputs or state variables of another population (e.g. $M(t)$ above), but by giving each its own little attractor. This model, of attractors with control variables, that are themselves state variables of smaller attractors, leads naturally to the notion of multiple and recurrent embedding of attractors, within attractors at different spatio-temporal scales. This model seems relevant to the brain and furthermore is capable of generating great complexity. Figure 3 illustrates an example of this by making the AMPA self-excitatory connections of a single simulated population a state variable of the Lorenz attractor (figure 1). The dynamics of the latter can be thought of as modelling some reasonably low-dimensional dynamical system at the level of transcription, or translational mechanisms, that is associated with maintaining the pre- and postsynaptic infrastructure of intrinsic excitatory connectivity. Depending on whether one construes the ensuing plastic connection strengths (figure 3*a*) as dynamic control variables or new state variables, one can regard this resulting dynamics (figure 3*b*) as type II or type I complex. Bear this example in mind because we shall return to it below when discussing neuronal selection and the importance of complexity.

(e) *Dynamic control variables and complexity*

(i) *Complexity at a microscopic level*

The distinction between type I and type II complexity depends on whether or not the underlying control variables are invariant. Here we will refer to control variables that change with time as dynamic control variables. Type II complexity is generated by dynamic control variables. However, a dynamic control variable can be thought of as a state variable in a larger system (even though the dynamics of that superordinate system may not be known). The distinction between type I and type II complexity is essentially a question of perspective. In other words, it depends on what we consider to be the boundaries of the system that determine whether a variable is an extrinsic control variable ar whether it is intrinsic to the system and enters as a state variable. From the point of view of self-organizing systems, and their empirical characterization, perturbations due to stochastic noise play a critical role in dynamic instability (Kelso 1995). In the current framework, these stochastic perturbations can be regarded as extrinsic influences rendering the ensuing transitions in the dynamics an example of type II complexity. The reason for introducing the notion of dynamic control variables will become apparent later when we (i) consider mesoscopic levels of description, and (ii) consider selection in the light of type II complexity.

The complexity that arises from interactions among populations can be seen as type II complexity if we regard the extrinsic inputs to any single population as dynamic control variables. Is this sensible? Consider in more detail how the voltage-dependent NMDA synaptic interactions in the simulations of § 2 were modelled. They are based on the probability that a voltage-dependent channel will open in response to a presynaptic input from some other population, times a sigmoid function of the postsynaptic transmembrane potential (see Appendix B, Friston (paper 1, this issue)). This term has two interpretations. First, we can construe it as a voltage-dependent effect that is mediated by extrinsic inputs (the conventional view) or equivalently it is a self-excitatory connection that is modulated or gated by extrinsic input. The latter interpretation follows from the fact that for any fixed afferent input the number of excitatory channels opening is an increasing monotonic function of depolarization. The reason for focusing on this interpretation is that it is a concrete example of the more general notion that inputs from distal populations can always be considered as dynamic control parameters, that express their effect vicariously through the dynamics intrinsic to the population in question. This notion can be generalized to all inputs, both driving and modulatory. At a simple level of analysis this perspective is clearly correct, in the sense that extrinsic inputs, even of a AMPA-like or driving sort, will lead to an increased sensitivity to intrinsic presynaptic inputs through a tendency to subthreshold postsynaptic depolarization.

(ii) *Complexity at a mesoscopic level*

Hitherto we have looked at complexity directly in terms of the state and control variables at a microscopic level. How does the distinction between type I and type II complexity translate at a mesoscopic level in terms of the input–output relationships among neurons? At the mesoscopic level there is an enforced partitioning of the brain into an ensemble of coupled input–state–output dynamical systems that calls naturally for a type II perspective. Here, extrinsic inputs enter as dynamic control variables. In brief, we will suggest that type II complexity can be represented as dynamic changes in Volterra kernels that characterize neuronal interactions or equivalently, and more intuitively, dynamic changes in effective connectivity within and among populations. Consider again equation (2) of Friston (paper 1, this issue) where we select just those functions that pertain to the activities of a small number of units or populations $\boldsymbol{x}_i$:

$$\boldsymbol{x}_i(t) = f(\boldsymbol{x}(t-u), \boldsymbol{C}). \tag{2}$$

We can divide the activities $\boldsymbol{x}$ on the right-hand side into those that come from the units or populations in question $\boldsymbol{x}_i$ and those that are extrinsic $\boldsymbol{x}_e$:

$$\boldsymbol{x}_i(t) = f(\boldsymbol{x}_i(t-u), \boldsymbol{x}_e(t-u), \boldsymbol{C}). \tag{3}$$

This has an alternative form were the extrinsic inputs are grouped with the control parameters $\boldsymbol{C}$ that govern the intrinsic dynamics.

$$\boldsymbol{x}_i(t) = f(\boldsymbol{x}_i(t-u), \boldsymbol{C}^*),$$

where

$$\boldsymbol{C}^* = f_C(\boldsymbol{x}_e(t-u), \boldsymbol{C}), \tag{4}$$

where f_C is a vector function (i.e. returns a vector). If we pretend for a moment that the extrinsic inputs $\boldsymbol{x}_e(t-u)$ are constant, then we have a simple (Volterra series) characterization in terms of state variables and control parameters $f_C(\boldsymbol{x}_e, \boldsymbol{C})$ that determine the Volterra kernels. This is a smaller version of the entire ensemble of units or populations that constitute the brain. In other words, equation (4) has exactly the same form as equation (2).

Now allowing for time-dependent changes in extrinsic input we can regard the subset of units or populations x_i as a collection of input–state–output systems with time-varying Volterra kernels. In other words, $\boldsymbol{C}^*$ specifies activity-dependent Volterra kernels. The implications of rewriting equation (2) as equation (4) are quite subtle but important. In equation (2), the control parameters $\boldsymbol{C}$ are fixed and represent the Volterra kernels that are applied to the activity throughout the brain to give the response in the populations one is characterizing. These control parameters 'stand in for' the fixed casual structure of the dynamics intrinsic to these populations, including, for example, all the intrinsic connectivity, short-term plasticity, facilitation, etc. Equation (4) suggests that this subsystem can also be regarded as an isolated system if one allows for time-varying control parameters $\boldsymbol{C}^*$. These dynamic control parameters are a function of extrinsic inputs and will cause type II complexity. In short, all neuronal activity in extrinsic afferents can be construed as dynamic control parameters that influence the recipient population by changing its attractor manifold in some way.

Recall that Volterra kernels are synonymous with effective connectivity, implying that the effective connectivity intrinsic to the units, or populations, considered is time dependent or, more specifically, activity dependent. The

type II perspective allows for activity-dependent changes in the kernels or connectivity that is due to afferents from other populations. Perhaps the simplest example is the modulation of effective connectivity between two populations by activity in a third. This sort of effect has been posited as a mechanism for the attentional modulation of the sensitivity of higher-order sensory areas to inputs from lower areas (see Büchel & Friston (1997) for an empirical example using neuroimaging). Put simply, if we took a small part of the brain and tried to characterize all the interactions among its components in terms of Volterra kernels (i.e. effective connectivity) we might be surprised to see that, for some reason, these kernels were themselves inherently unstable. This would reflect the fact that we had ignored the effect of extrinsic input from other parts of the brain (e.g. modulation of connection strengths between V2 and V5 by parietal afferents). The value of this perspective is that it (i) highlights the central role of activity-dependent changes in effective connectivity in the genesis of complexity; and (ii) motivates a way of characterizing the effects of extrinsic input when this extrinsic input is not known, or rather is known but the particular subset of afferents responsible are not (G. Green, personal communication). This will be pursued elsewhere.

(f) *Complexity and transients*

The above suggests that brain dynamics can be thought of in two equivalent ways. First, they arise from a single, immensely complicated, dynamical system with a global attractor manifold that is stable over time but supports an itinerant trajectory or, second, it is an ensemble of small, loosely coupled systems, wherein each smaller system has a dynamically changing attractor manifold due to extrinsic influences from other systems. The only distinction is whether one considers all the populations collectively or focuses on a single population (at any scale) embedded among the others. The existence of dynamic control variables must engender some degree of type II complexity in each population. Given that a type II perspective can be adopted for any complex dynamics, the genesis of complexity reduces to the source of dynamicism in the extrinsic control variables. Clearly there are three possibilities: (i) dynamics in subordinate systems (e.g. slow dynamics of protein synthesis, axonal transport and translational mechanisms); (ii) the effect of extrinsic input from superordinate systems (e.g. neuromodulatory input from another population); or (iii) coupling to the dynamics on different time-scales within the system (e.g. activity-dependent plasticity such as long-term depression). At a mesoscopic scale one can regard inputs from other neuronal populations as dynamic control parameters that enter as variables determining the form of the Volterra kernels expressed at that time.

A key notion here is that the cortical sheet can be considered as an ensemble of separable neuronal populations. Each population enacts its dynamics by tracing out a trajectory on its own attractor manifold. The influence of other populations changes the attractor manifold and the associated dynamics. This speaks to an important question, that has been addressed by a number of investigators: 'Are brain dynamics best described by a single dynamical system (i.e. a global attractor) or an ensemble of separable systems (i.e. a collection of smaller attractors)?' In Friston (1997, p. 164), we addressed this issue and concluded that both viewpoints can be reconciled by representing brain dynamics as a global attractor, where 'this single attractor has a special complexity that emulates a collection of constituent attractors whose manifolds appear to change as a function of time'. This is simply type I complexity. From the point of view of any one neuronal population, the constantly changing attractor manifold will give rise to a succession of transient dynamics, each with its own distinct and recurring spatio-temporal organization. Generally, a 'transient' describes the behaviour of a system that occurs in the initial period as a dynamical system settles down and approaches an attractor. Here neuronal systems are considered to be perpetually in an 'initial period' by virtue of continuous changes in the underlying attractor. These changes may be construed in terms of the effect of dynamic control parameters (type II) from other populations (e.g. changes in connection strengths caused by modulatory interactions) or attributed to itinerant dynamics where the trajectory wanders from one part of the global attractor manifold to another (type I). This behaviour is the essence of neuronal transients and engenders the special complexity typical of the brain. This complexity is not mandatory and only emerges under certain conditions. The aim of the analysis presented in Friston *et al.* (1997) was to show it arises when and only when the anatomical connectivity among populations is sparse. In what follows we revisit this issue but now ask whether the nature of the extrinsic coupling (linear or nonlinear) is an important factor in the genesis complexity.

3. COUPLING AND COMPLEXITY

(a) *The principle of sparse connectivity*

There are clearly many aspects of neuronal interactions that can render neuronal dynamics complex. In this section, we consider what happens when the connections between simulated neuronal populations are gradually increased. On the basis of simulations and electrophysiology one can make some predictions about the effects of increasing extrinsic connectivity: at very low levels of connectivity each neuronal population will express its own dynamics, unaffected by those of its neighbours. As connectivity increases, the dynamics should come to resemble the complicated, intermittent dynamics seen in the real brain. As extrinsic connectivity is increased further, dynamic instability disappears, with every population locked into a single, coherent pattern of activity (see Friston *et al.* 1995). In Friston (1997), we addressed directly the effects of increasing extrinsic connectivity in simulations to confirm these predictions. Initially the dynamics move from stable incoherence, where each population preserves its own unique and relatively stable behaviour, through a regime of dynamic instability (transients and periods of stable coherence that are themselves inherently unstable) to, finally, a regime of stable coherence with phase-locking and complete entrainment. The intermediate regime of dynamic instability is the subject of interest because it is in this regime that transients flourish with a diversity that has important implications for adaptive neuronal responses (see § 4).

In this section, we will revisit the relationship between complexity and sparse connectivity using a more comprehensive analysis, based on the model described in Friston (paper 1, this issue) and looking at driving (AMPA-like) connections, modulatory (NMDA-like) connections and the interactions between these two types. Furthermore, we will look at the relationship between complexity and the relative contributions of synchronous and asynchronous coupling. The idea here is that asynchronous, nonlinear interactions will supervene in the regimes of high complexity that best characterize real brain dynamics. To pursue this, complexity, and in particular dynamic instability, has to be defined and measured.

(b) *A measure of dynamic instability*

Here we describe a simple measure (Friston 1997) of dynamic instability framed in terms of the instability or entropy of the spectral density $g(\omega,t)$ of a measured neuronal process. If metastability is characterized by transient periods of stability, or the recurrent expression of different transients, then the frequency composition, or spectral density of the time-series should change with time. However, if the dynamics are stable, then the corresponding spectral densities will not change; irrespective of whether that stability results from the expression of independent intrinsic dynamics (i.e. no connectivity) or from complete entrainment and coherence (i.e. dense connectivity). Consequently the changeability or stability of the spectral densities can be used to measure metastability.

The entropy of a process reflects its predicability or the amount of information one would derive from observing it. A metric that has a high entropy is unpredictable and therefore provides a lot of information when known. For example tossing a coin provides less information than throwing a die. The former has only two outcomes (one bit of information) whereas the latter has six outcomes (i.e. $\log_2(6)$ bits). Dynamic instability is simply a characteristic of a system whose attractor manifold is unpredictable. Therefore a measure of dynamic instability is provided by the entropy of the spectral densities associated with that manifold. Here the spectral densities are being used as a 'signature' for a particular manifold that is relatively insensitive to the actual trajectories themselves. The measure used in this paper is a simple measure predicated on a time-frequency analysis of the underlying process (that was first described in Friston (1997)). It is, however, suboptimal in that it only approximates what one is really trying to measure. Ideally a measure of dynamic instability should reflect the causes of the complexity, which reduces to variability in the dynamic control variables that underpin it. A measure of this variability would be a direct measure of changes in the attractor manifold whilst being completely insensitive to the trajectories extant at any time. Practically, this sort of measure obtains from the entropy of the estimated Volterra kernels describing the dynamics. This is the subject of current work.

$$H = H\{g(\omega,t)\}, \tag{5}$$

where $H\{.\}$ returns the entropy. See Friston (paper 1, this issue) for a description of how $g(\omega,t)$ is computed. Under Gaussian assumptions (Jones 1979),

$$H = \log\left((2\pi e)^m \det\{\mathrm{Cov}\{g(\omega,t)\}\}\right)/2. \tag{6}$$

$\det\{.\}$ means the determinant of a matrix and m is the number of frequencies considered. This simple expression provides the measure used below to assess metastability as a function of extrinsic connectivity.

(c) *Nonlinear coupling and complexity*

To look at the relationship between dynamic instability and the relative role of synchronous and asynchronous coupling, we used the neuronal model described in Appendix B of Friston (paper 1, this issue). In these simulations we varied both NMDA- and AMPA-like extrinsic connections between several simulated populations. In this way, we were able to elicit dynamic instability in any one population and examine the nature of the coupling between it and connected populations. The simulations comprised four populations, where the intrinsic self-excitatory connection strengths were fixed at random values sampled from a uniform distribution in the range 0.12 to 0.16. The extrinsic excitatory connections were varied using a parameter ε, such that the connection strengths of all NMDA or AMPA connections were $1/(1+\exp(-\varepsilon))$. This allowed us to explore the effects of increasing extrinsic connectivity at very low or sparse levels (from 0 to 0.2). The architecture of the model was simple and comprised four serially linked populations. In the forwards direction the connections were all driving or AMPA-like. Backwards connections were all modulatory or NMDA-like. This biologically plausible organization (Crick & Koch 1998) allowed us to manipulate both driving and modulatory connections in the context of the same architecture. For each pair of AMPA- and NMDA-like connectivity strengths we simulated the dynamics over a 4 s period. The activities of each population were subject to time-frequency analyses as described in Friston (paper 1, this issue) to give $g_i(\omega,t)$ $(i=1,\ldots 4)$. The metastability H was computed for the third population according to equation (6) and the proportion of variance in $g_3(\omega,t)$ predicted by synchronous and asynchronous contributions from connected populations (2 and 4) was computed, using the regression analysis described in Friston (paper 1, this issue) extended to model multiple inputs. The results of this analysis are shown in figures 4 and 5 in terms of the connection strength parameter ε.

Figure 4 shows how complexity varies with extrinsic connectivity. Consistent with previous results (Friston 1997), dynamic instability is evident only in a limited regime of connection strengths, namely when they are sparse. Figure 4*a* shows complexity as a function of AMPA- and NMDA-like extrinsic connectivity in image format (white areas correspond to high levels of dynamic instability). The interaction between driving and modulatory connections is not simple, i.e. the effect of increasing the strength of AMPA connections depends on the level of NMDA connectivity and vice versa. In fact, there seems to be an optimum balance to attain maximal complexity in these simulations, where driving connections are slightly sparser than modulatory connections. The dynamics, averaged over all four populations under conditions of maximal complexity, are shown in figure 4*b*, and are remarkably complicated considering we only used four loosely coupled oscillators.

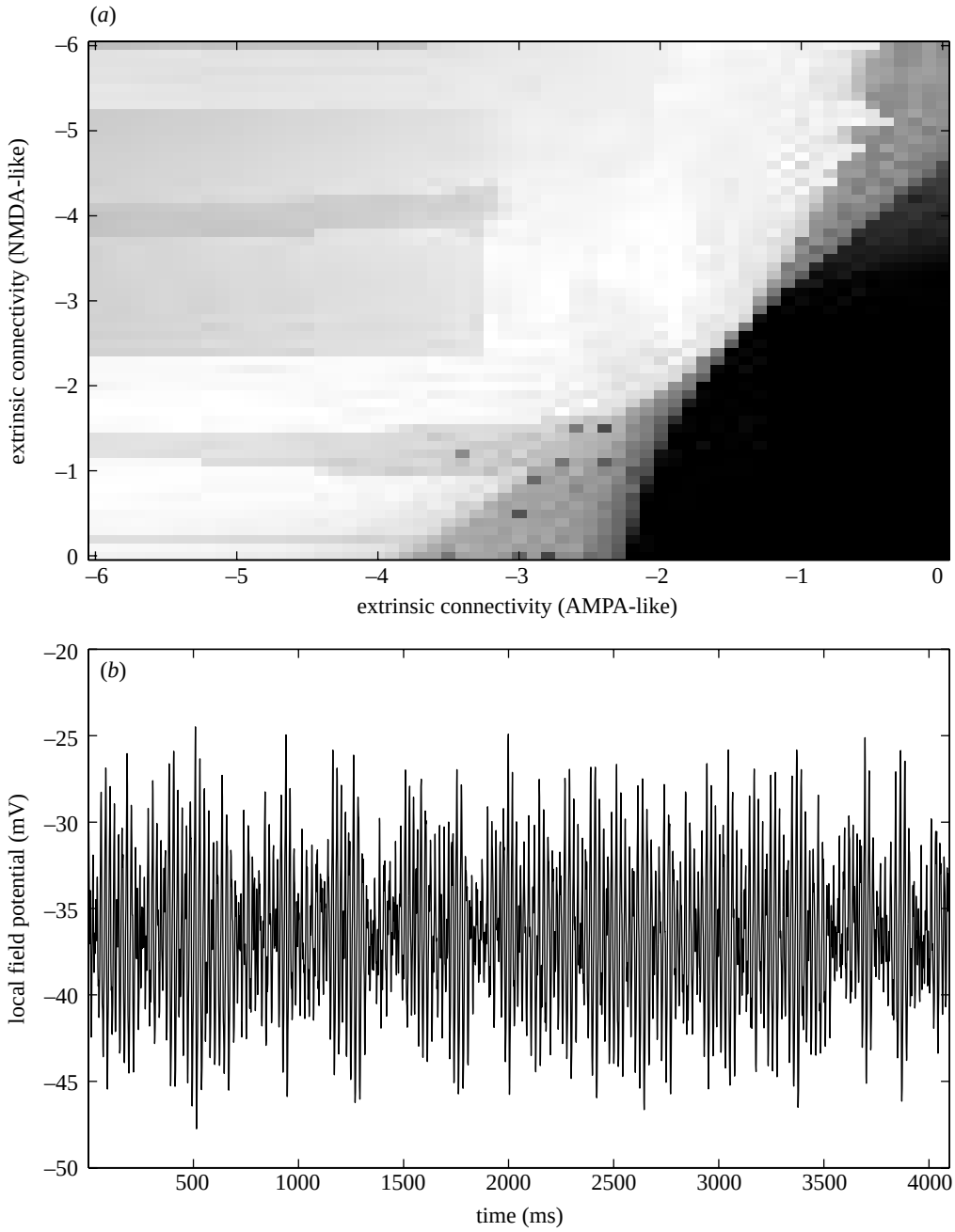

Figure 4. Dynamic instability in coupled neuronal populations. (*a*) Dynamic instability, indexed by the entropy of the spectral density changes (from 8 to 96 Hz), as a function of extrinsic AMPA- and NMDA-like connection strengths. The architecture of this system is described in the main text. The key thing to observe is that metastable dynamics are limited to a regime of sparse connectivity, when both AMPA- and NMDA-like connections are very low. The mean LFP of the four simulated populations is shown in (*b*) for the most complex pairing of both connection types. Extrinsic connectivity is expressed in terms of ε, where connection strength $= 1/(1 + \exp(-\varepsilon))$.

Figure 5 shows the relationship between complexity and the degree of synchronous and asynchronous coupling at a fixed level of AMPA-like connections (0.1). Figure 5*a* shows the complexity and corresponds to a row from figure 4*a*. Figure 5*b* shows the nature of the coupling that underlies this complexity, expressed as the proportion of the variance in $g_3(\omega,t)$, over all frequencies, explained by synchronous and asynchronous effects from connected populations. These results reveal something quite significant; namely that asynchronous coupling prevails when the dynamics are complex or metastable. As complexity starts to fall, with increasing connectivity, the degree of asynchronous coupling also falls. As complexity becomes very low, synchronous interactions increase substantially and overtake asynchronous coupling. This phenomenon is important because the real brain shows dynamic instabilities, speaking again to the importance of nonlinear interactions. In other words, functional integration is not simply a question of engendering synchronized interactions throughout the brain but a delicate balance between preservation of dynamics intrinsic to each population and the mutual influence among these populations. This balance facilitates a dynamic instability that is induced and maintained by

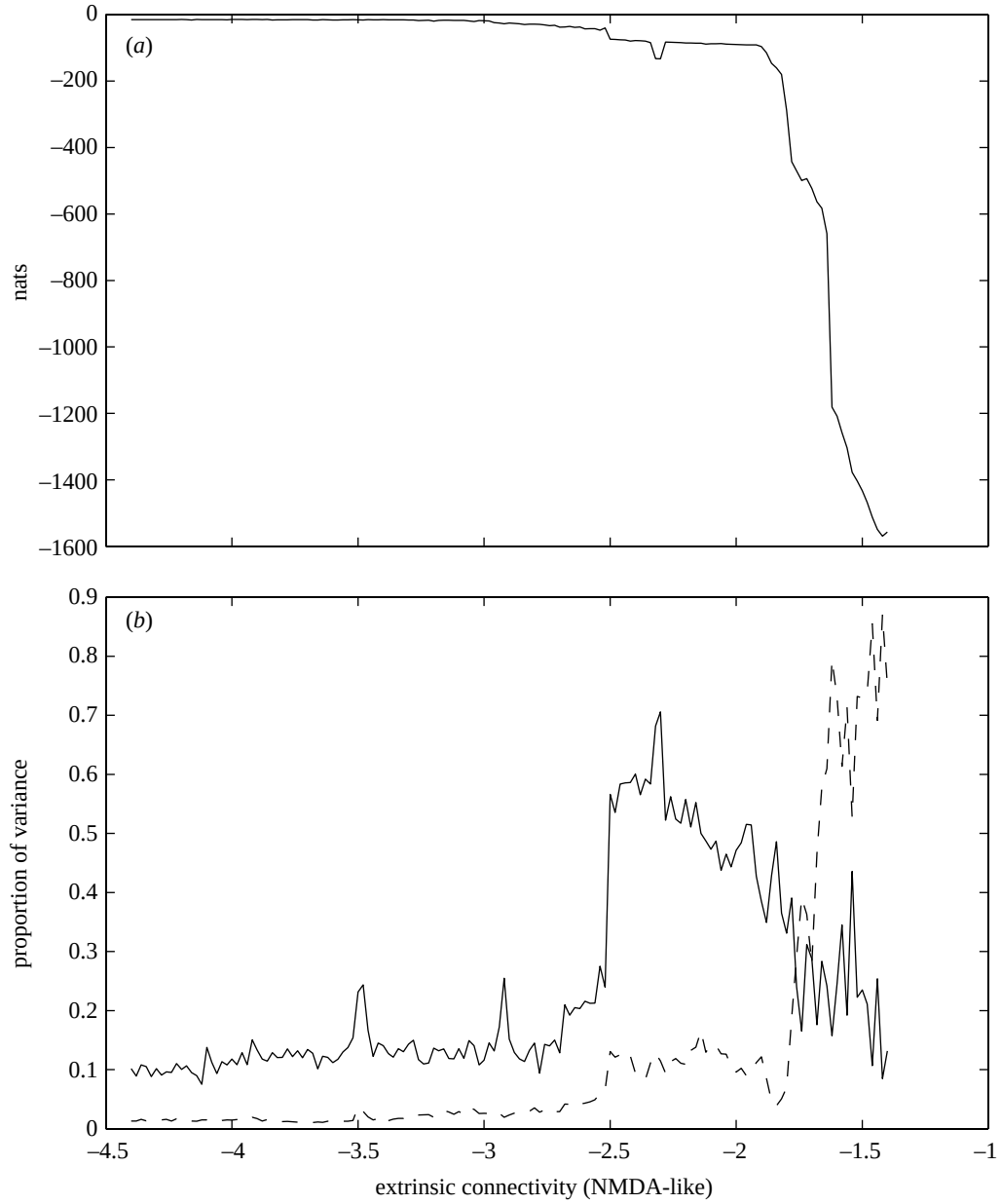

Figure 5. The relationship between asynchronous coupling and (*a*) dynamic instability. For a fixed AMPA-like extrinsic connectivity of 0.1 the NMDA-like connection strengths were varied from $1/(1+\exp(-4.4))$ to $1/(1+\exp(-1.4))$. (*b*) The proportion of variance in frequency modulation, in the third of the four simulated populations, explained by synchronous (broken line) and asynchronous (solid line) coupling with the second and fourth populations are shown in the lower panel. The point being made here is that asynchronous coupling supervenes in domains of high metastability whereas synchronous coupling emerges when the dynamics become more coherent and less complex.

asynchronous interactions (see Freeman & Barrie (1994) and Erb & Aertsen (1992) for convergent discussions).

In summary, complicated metastable dynamics can occur when the connectivity among simulated neuronal populations is sparse. Complexity of this sort is characterized by a succession of transients that belie a continuously changing attractor manifold for each neuronal population and consequent instability of spectral density of the underlying dynamics. Characteristic of this complexity is the prevalence of asynchronous or nonlinear coupling between the populations. In relation to the distinction between type I and type II complexity, one can regard the dynamics of the simulations above as a reflection of type I complexity when considering all four populations together, or as an example of type II complexity where the dynamic control variables correspond to extrinsic inputs to the third population. When these inputs result in asynchronous coupling among populations the complexity is likely to be high.

On the basis of this, and in the light of the simulations presented above, one can infer that neuronal dynamics are modelled by neither an ensemble of separate attractors nor a simple low dimensional attractor, but are consistent with the attractor surface that ensues when many separate attractors are loosely coupled together or buried within each other. This manifold has a special complexity, where the trajectories upon it show complicated dynamics, with the recurrent appearance and destruction of transients. In keeping with much of the current thinking on self-organizing systems, this rich

form of dynamical instability is found in regimes of parameter space near critical points or phase transitions (e.g. Kauffman 1992; Kelso 1995). This work suggests that, for the brain, these critical regions involve sparse, extrinsic connectivity and asynchronous interactions. Before considering why complexity of this sort might be important, dynamic instability is related to some empirical studies.

(d) *An empirical perspective on dynamic instability*

Dynamic instability has been characterized above in terms of the modulation of different frequencies expressed in neuronal time-series that accompanies the expression of transients. It has often been noted that 'gamma rhythms can occur in brief bursts with a considerable jitter in the frequency' (Jefferys *et al.* 1996) and this has been used as a rationale for explaining the difficulties in detecting synchronization. The alternative viewpoint is that, of course, it is this very transience and jitter that is the essence of what should be measured. Dynamic changes in synchronized interactions do provide the focus for research, most prominently in event-related studies. These pertain either to a single time-series (e.g. event-related desynchronization (Pfurtscheller & Aranibar 1979)) or those relating two or more time-series (e.g. dynamic coherence). See Mayer-Kress *et al.* (1991) and Fuchs *et al.* (1992) for compelling examples, and Pfurtscheller & Aranibar (1979) for spectral density changes in relation to self-paced movement.

(i) *Event-related examples*

Most of the empirical evidence for dynamic modulation of frequencies in neuronal time-series comes from event-related work. For example, Tallon-Baudry *et al.* (1996, 1997) have characterized transient episodes of synchronization in the gamma band following presentation of visual stimuli using electroencephalograph (EEG) recordings. They find early stimulus-locked synchronization unrelated to the stimulus type in contradistinction to a later component at 40 Hz that appears around 280 ms and is not locked to the stimulus onset. The latter component is stronger for coherent (illusory or real) stimuli relative to non-coherent stimuli. See Steriade *et al.* (1996) for a description of frequency and temporal coherence changes in relation to stimulation of ascending activating systems and naturally during different stages of sleep. Similar examples can be found at lower frequencies than the gamma range. For example, Vanni *et al.* (1997) demonstrate modulation of the parieto-occipital alpha rhythm during object detection, where the presentation of non-objects evoked systematically higher levels of transient alpha activity than objects. This difference emerging about 400 ms after presentation.

(ii) *Dynamic correlations and nonlinear coupling*

In Friston (paper 1, this issue) we introduced dynamic correlations (e.g. Vaadia *et al.* 1995) as an equivalent perspective on neuronal transients. Because of the intimate relationship between the successive expression of transients and complexity, there is an equally straightforward relationship between complexity (i.e. dynamic instability) and dynamic correlations. Consider two normalized neuronal time-series $x_i(t)$ and $x_j(t)$. The cross-correlation function $\rho_{ij}(u,t)$ over lag u, at time t, is given by the inverse Fourier transform IFT{.} of their cross-spectral density

$$\rho_{ij}(u,t) = \mathrm{IFT}\{\sqrt{(g_i(\omega,t)g_j(\omega,t))} \times \exp(j[\phi(\omega,t) - \phi_j(\omega,t)])\}, \quad (7)$$

where $f_i(\omega,t)$ denotes the phase at a particular frequency and time. This expression says that the dynamic changes in the cross-correlation can be induced by either changes in the spectral density (i.e. dynamic instability in one or both time-series) or transient changes in the phase relationship between them. Desmedt & Tomberg (1994) have demonstrated transient phase-locking between prefrontal and parietal EEG recordings during selective attention. A fuller discussion of the relationship between dynamic correlations and transient phase-locking is found in Friston *et al.* (1997).

This equivalence (equation (7)) means that dynamic instability necessarily results in dynamic correlations. A beautiful example of dynamic correlations is provided in MacLeod & Laurent (1996). In this instance, the dynamic correlations were expressed, following olfactory stimulation, as transient synchronization of neuronal assemblies in the olfactory system (in particular between the dynamics of projection neurons and the local field potential). Recent work on nonlinear coupling (Schiff *et al.* 1996) has been exploited in the investigation of intracranial EEG recordings during complex partial seizures. Le Van Quyen *et al.* (1998) have shown that during the inter-ictal period, nonlinear coupling was low or absent, whereas transient nonlinear interdependencies developed during the onset, offset and other critical periods of seizure development. Although this example comes from pathology, it is among the rarer empirical examples of asynchronous coupling in real brain systems.

4. WHY IS COMPLEXITY IMPORTANT?

In § 2 and 3, the importance of asynchronous coupling and neuronal transients when characterizing brain dynamics is promoted. They are motivated by the assumption that the brain should be complicated or evidence dynamic instability. Why should it? There are two main themes that emerge when thinking about this question. The first relates to selectionism and self-organization in the brain and the second to information-theoretical approaches to brain function.

(a) *Dynamic instability and second-order selection*

One of the most compelling reasons for the complexity of brain dynamics is based on population dynamics and selectionism, either at an evolutionary or somatic time-scale (Edelman 1993). The point of contact between selectionism and complexity (i.e. diversity) may be one of the most promising areas for understanding the mechanistic basis of the adaptive and self-organizing capacities of the brain.

The idea is simple. Given that one can take a type II perspective on the complexity of any neuronal system, it follows that there must be dynamic control variables that underpin it. If these dynamic control variables are subject to consolidation, that is contingent on the dynamics

expressed, then there exists a mechanism for selection and self-organization. Furthermore, irrespective of the contingencies mediating this consolidation, when, and only when, the dynamics of the control variables generate sufficient diversity to support selection will these dynamics prevail under selective pressure. In short, it is only necessary to posit dynamic control variables that are subject to consolidation, contingent on the dynamics expressed, to explain (i) how particular spatio-temporal patterns of activity can be selected, and (ii) how dynamic instability is itself selected as an emergent but necessary precondition for this selection. The analogy here with evolutionary selection is obvious. The transient dynamics correspond to the phenotype and the dynamic control variables to the underlying genotype. Dynamic instability over time corresponds to phenotypic diversity over a population and consolidation of the dynamic control variables corresponds to selection *per se*. The emergence of dynamic instability corresponds to the second-order selection of selectability traits; a theme that was part of the evolutionary synthesis when 'adaptational aspects of diversity were analyzed as due to selection forces' (Mayr 1982). We will now review this in a less abstract fashion.

Clearly for the brain to mediate adaptive sensorimotor integration the repertoire of neuronal responses to any environmental situation should be sufficiently diverse to allow for the selection of the most adaptive response. This selection is probably most easily seen in terms of consolidation of synaptic efficacies when something of value occurs. Value-dependent selection (Friston *et al.* 1994) is no different from any other form of selection and necessarily depends on diversity upon which selection can act: in this instance, spatio-temporal patterns of activity or neuronal transients and implicitly the connections that subtend them. In other words, a diversity or dynamic instability is a necessary condition for the selection of adaptive functional integration. To the extent that this selection is mediated by consolidation of activity-dependent control parameters (the type II perspective), then it depends on diversity of the activity patterns themselves. This obtains as a natural consequence of dynamic instability. There is then a simple casual explanation for why the brain should come to express complicated dynamics that depends on selection for selectability (Kauffman 1992). If it is necessary to have metastability to facilitate neuronal selection then that metastability has, by definition, adaptive value. It will therefore be selected for at both an evolutionary and neuronal level. This mechanism is a second-order aspect of selection that does not pertain to 'what' is selected but to the constraints under which selection can occur. It is easy to see a phylogenetic trend to increasing complexity on an evolutionary time-scale, but perhaps it is more interesting to look at similar phenomena on a neuronal time-scale. In the absence of a pre-selected neuronal response to some completely novel situation, the only adaptive strategy is to facilitate the (selective) mechanisms that will lead to an adaptive response, namely, to explore the diversity of alternative responses (i.e. one can do nothing, keep doing the same thing or try different things until something works. Only the latter is adaptive). This leads naturally to the induction of dynamic instability. A fanciful view of desynchronization associated with arousal (Munk *et al.* 1996) is that it may, in part, be the electrophysiological correlate of augmented selectability.

(b) *An illustrative simulation*

These points can be made clearer using a simple example that employs one of the simulations above. Consider the example of type II complexity presented in figure 3. In this case, a single population was simulated where the excitatory–excitatory AMPA-like connectivity was assigned its own dynamics using the Lorenz attractor. Say that the regular bursting of this population every 600 ms has some adaptive value and that this was reflected in a reduction of 'arousal', indexed by the afferent input from ascending modulatory neurotransmitter systems. The effect of this input could mediate (i) selection of the appropriate dynamics, and (ii) selection for selectability in the following way. First let us distinguish between the underlying connection strengths (more generally any control variables) $\boldsymbol{C}$ and the actual dynamic control parameter $\boldsymbol{C}^*$, which include some dynamic terms $\boldsymbol{D}(t)$ (e.g. those borrowed from the Lorenz attractor in previous simulation) where

$$\boldsymbol{C}^*(t) = \boldsymbol{C}(t) + \boldsymbol{D})(t) \times A(t). \tag{8}$$

It is clear that under states of high arousal the dynamic connection strengths will vary substantially, generating a fair degree of complexity in the resulting activity. Conversely when arousal is low, $\boldsymbol{C}^*(t)$ will not change much and the type II complexity, attributable to this dynamic control variable, will be minimal. Under conditions of high arousal the dynamics expressed will be diverse, exploring a large range of patterns and increasing the probability that an adaptive pattern will be expressed and selected. Equation (8) models selection for selectability (second-order selection). Arousal enters here as a control parameter for the dynamics of the dynamic control parameters (hence the second-order nature of this effect). The dynamicism modelled by $\boldsymbol{D}(t)$ can arise from three possible sources: (i) modulatory inputs from other neuronal systems (e.g. classical neuromodulation); (ii) from subordinate dynamical systems associated with the control variable in question (e.g. protein synthesis and slow changes in cellular infrastructure); or (iii) it can be 'borrowed' from the dynamics of the system itself (e.g. activity-dependent plasticity). Each of these will generate a type II complexity or diversity at the level of the activity patterns that are selected.

The selective consolidation of $\boldsymbol{C}^*(t)$ itself can be modelled in many ways, perhaps the simplest is

$$\partial \boldsymbol{C}(t)/\partial t = -\partial A(t)/\partial t \times (\boldsymbol{C}^*(t) - \boldsymbol{C}(t)), \tag{9}$$

where the underlying (latent) connection strengths 'catch up' with the actual (effective) values in proportion to the ongoing reduction in arousal. In other words, when something valuable happens that reduces arousal, the short-term potentiation or depression of synaptic efficacy, extant at that time, is consolidated. The form of equation (9) is exactly the same as that used in Friston *et al.* (1994) to model value-dependent consolidation of synaptic efficacy and, through this, is closely related to temporal difference models of reinforcement learning (Sutton & Barto 1990).

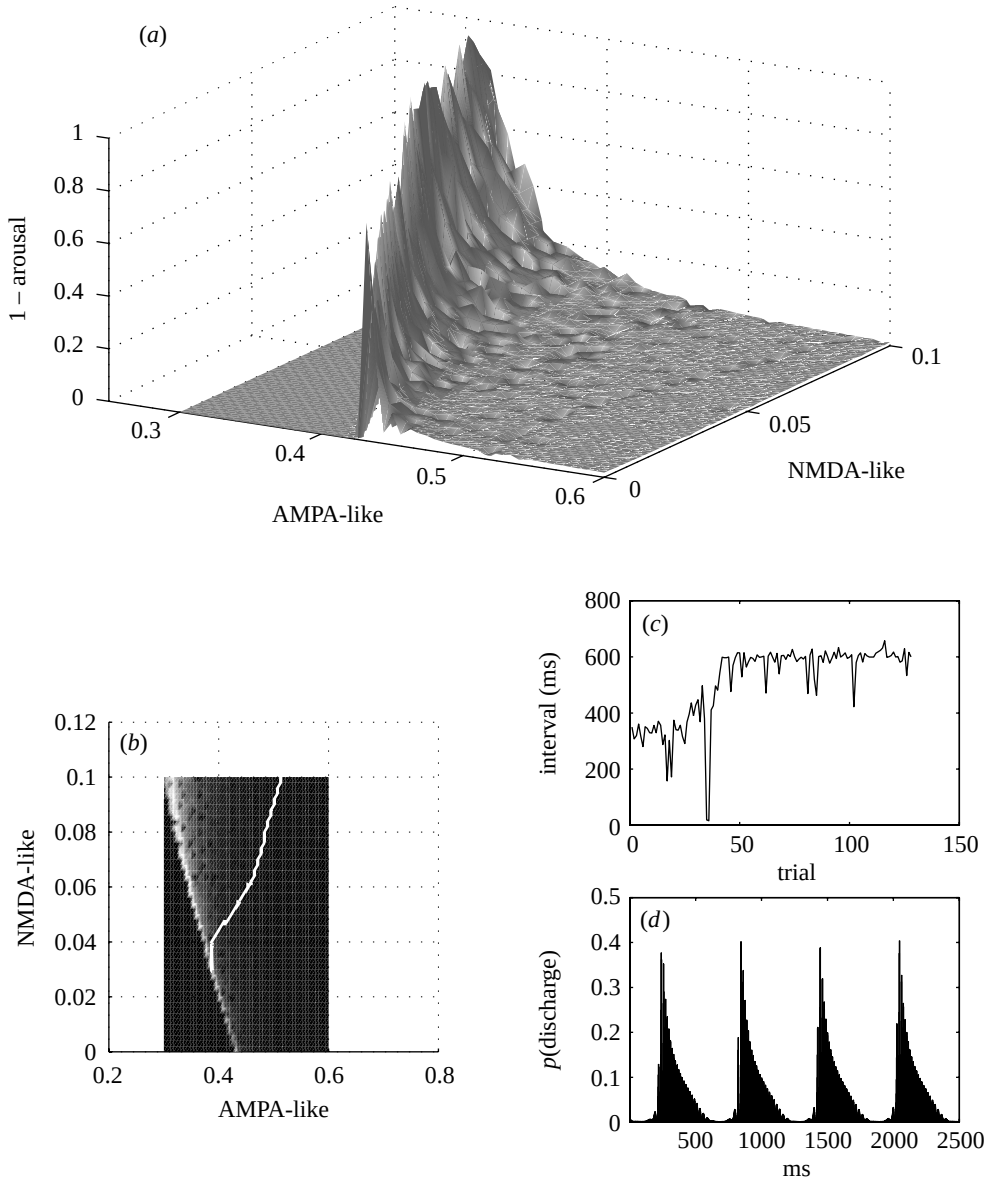

Figure 6. Selection and fitness landscapes. (*a*) The 'fitness landscape' computed on the bases of 4096 iteration trials as a function of self-excitatory AMPA- and NMDA-like connection strengths in a single neuronal population. Fitness here is defined as $1-A(t)$, where $A(t)$ is 'arousal'. $1-A(t)=\exp(-\mathrm{abs}(I-600)/64)$. I is the interval between bursting in milliseconds. Fitness is 1 (or arousal is 0) when the inter-burst interval is 600 ms. (*b*) The same landscape as above but now viewed from the top. Superimposed is the trajectory taken during a simulation using the selective consolidation scheme described in the main text. The evolution of inter-burst interval is shown in (*c*) and the dynamics eventually selected are shown in (*d*), in terms of discharge probability.

Here we will illustrate a rather more robust model of consolidation that depends on a dynamic threshold η that arousal or its complement 'value' (in this case, value $=1-A(t)$) must exceed for consolidation of short-term changes in synaptic efficacy to occur. Let $\pi(t)$ be a permissive effect mediated by ascending neurotransmitter systems (i.e. $A(t)$), such that

$$\partial \boldsymbol{C}(t)/\partial t = \pi(t) \times (\boldsymbol{C}^{*}(t) - \boldsymbol{C}(t)),$$

and

$$\partial \eta(t)/\partial t = \pi(t) \times (A(t) - \eta(t)),$$

where

$$\pi(t) = \dagger\{\eta(t) - A(t)\}. \qquad (10)$$

$\dagger\{.\}$ is some threshold function; in the example below $\dagger\{x\}=\tanh(x\times 16)$ for $x>0$ and $\dagger\{x\}=0$ otherwise. Working through these equations will reveal a very simple mechanism. Whenever arousal falls below a threshold, the underlying connection strengths change to approximate the extant dynamic values. Coincidently the threshold approaches the new level of low arousal. This system of equations ensures a progressive minimization of arousal. A more intuitive understanding of how this selective consolidation works can be gained in relation to the notion of gradient ascent on 'fitness' landscapes. Consider again the problem of selecting a dynamic of bursting every 600 ms or so. For every combination of AMPA- and NMDA-like excitatory–excitatory connection strengths in a given population

there will be a mean inter-burst interval that may range from several milliseconds to long intervals (possibly as long as 600 ms). Each pair will be associated with an arousal that is the complement of value or 'fitness'. Plotting $1-A(t)$ as a function of the two dynamic control variables creates a landscape in which the peaks represent the most adaptive values of the control parameters and associated bursting activity. The upper panel of figure 6*a* shows this landscape for a single population that includes short-term depression used in figure 3. In this instance arousal $A(t)$ was computed as an exponential function of the absolute difference between the actual and desired inter-burst interval. Imagine that you and I had to climb this mountain in the dark. First of all we elect to simply follow the steepest course in our immediate vicinity (cf. conventional gradient ascent). We soon realize that this is not going to work because the mountain is covered with little peaks, one of which we arrive at very quickly and get stuck. We then decide that I should wander around a bit and shout whenever I start going uphill. At this point you join me and we repeat the procedure (cf. equation (9) and temporal difference models). This seems to work better for a while until I find myself walking up from a local gully and calling you down into it! Finally we agree that I should explore the local terrain and whenever I am higher than you, you will then join me. This works. It is slow but infallible and is exactly what equation (10) implements. I am the dynamic control variable exploring the local fitness landscape and you are the underlying control variable. Clearly the dynamics of the exploration have to be slow in order to build the landscape, that is a function of the faster dynamics of the activity engendered. An example of a path or trajectory taken in a simulation of these selective dynamics is given in figure 6*b*. The arousal in this example was evaluated every 4096 iteration 'trial' and the consolidation of connection strengths and threshold effected at these times. The short-term plasticity $\boldsymbol{D}(t)$ in this example was assumed to be driven by a high-dimensional deterministic system and emulated by sampling from a random Gaussian distribution (s.d. $=1/32$). The trial by trial evolution of inter-burst interval is shown in figure 6*c* and the final dynamics selected in figure 6*d*, in terms of the discharge probability. In this simulation $\eta(\mathrm{t})$ was substituted for $A(t)$ in equation (8) to give a more graceful convergence.

This example is a simple one, used to illustrate the idea, and ignores all sorts of issues relating to dynamic and contextual changes in the fitness landscape, how this landscape is constructed or, from the point of view reinforcement learning, temporal and spatial credit assignment. However, it does possess some features that make it unique in relation to other adaptive neural networks. First, it explicitly invokes a second-order selection (in terms of arousal in the current example) of adaptive exploration of the fitness landscape, and second, there is no dependence on associative plasticity or Hebbian learning of any sort. It could be considered as a form of genetic algorithm that has been cast in terms of dynamical systems.

Despite its simplicity, this example shows how easy it is for any particular dynamic to be selected using mechanisms that may well have neurobiological correlates. The particular model of consolidation here deals with the typically difficult terrains associated with fitness landscapes of dynamic systems that show great sensitivity to small changes in the control parameters (in the language of self-organizing systems, near the edge of chaos or in regimes dense in phase transitions). It should be noted that this sort of nonlinear gradient ascent depends on only one scalar ('arousal' or its complement 'value') and as such is locally computable in the brain, provided the 'value' (height on the fitness landscape) is available.

(i) *Neurobiological mechanisms*

Selective consolidation as mediated by some consolidatory or reinforcing signal is an essential component in the emergence of adaptive dynamics. There is considerable evidence to suggest monoaminergic (dopamine DA, norepinephrine NE and serotonin 5HT) and cholinergic (ACh) neurotransmission facilitates either the consolidation or maintenance of long-term changes in synaptic strength following short-term plasticity (McGaugh 1992). The three most compelling lines of evidence are modulation of (i) experience-dependent changes in synaptic efficacy, (ii) behavioural plasticity, and (iii) experimentally induced long-term potentiation. One neurodevelopmental example is the role of ACh in facilitating experience-dependent organization of connections in striate cortex. Induction of ocular-dominance shift in cats, and its recovery following monocular deprivation depend on the integrity of NE and/or ACh neurotransmission (Bear & Singer 1986). Further evidence implicating ACh in the modulation of plasticity comes from the electrophysiological correlates of learning (Metherate & Weinberger 1989). See Friston *et al.* (1994) for a full discussion.

5. CONCLUSION

The main arguments developed in this paper can be summarized as follows.

(i) Complexity is the phenotypic variation in neuronal dynamics or transients that, from a dynamical perspective, implies a constantly changing attractor manifold.

(ii) Changes in an attractor manifold can be apparent, as itinerant trajectories explore different regions of a global attractor (type I complexity), or can arise from the extrinsic influence of dynamic control variables. The distinction depends on whether one considers the system as a whole or focuses on a component.

(iii) Considering the brain as an ensemble of loosely coupled attractors allows one to treat extrinsic inputs to any neuronal population as dynamic control variables that will engender type II complexity.

(iv) The Volterra formulation in terms of coupled input–state–output systems replaces dynamic control variables with dynamic Volterra kernels. Volterra kernels embody the intrinsic causal structure of the population dynamics and how extrinsic inputs enter in terms of effective connectivity. Dynamic Volterra kernels therefore correspond to dynamic or activity- and time-dependent effective

connectivity that is a cornerstone in the genesis of complexity.

(v) The regimes in which complexity is most prominent involve sparse extrinsic connections and asynchronous or nonlinear coupling among populations. This can be demonstrated heuristically with neuronal simulations.

(vi) The theoretical importance of complex dynamics in the brain can be motivated by appeal the second-order selection. If the selection of adaptive transients depends on some primary selective mechanism, then diversity or complexity is necessarily adaptive and will be subject to selective pressure at neuronal (e.g. the correlates of arousal) or evolutionary time-scales. The corollary of this is that activity- and time-dependent connections, sparse connectivity and nonlinear coupling, as architectural features implicated in the genesis of complexity, will themselves be subject to selective pressure.

(vii) The conjunction of dynamical and selective perspectives posits that phenotypic diversity is maintained by complexity and the diverse expression of neuronal transients over time. These 'populations' of phenotypes are caused by control parameters that stand in for the genotype. Genetic variation, over time, is mediated by dynamicism in control parameters that themselves ensure phenotypic variation through type II complexity.

(viii) This framework leads to the intuitive position that the primary selective mechanisms act at the level of dynamic control variables (i.e. changing Volterra kernels) or, equivalently, activity- and time-dependent effective connections. The consolidation of plastic changes in connection strengths is an established theme in theoretical neuroscience.

The relationships among complexity, self-organizing systems and neuronal selection is an intriguing area and will be dealt with more thoroughly, in the context of neuronal transients, elsewhere. In Friston (paper 3, this issue) we focus on neuronal transients and how information is abstracted from the sensorium in early visual processing.

K.J.F. is funded by the Wellcome Trust. I would like to thank Semir Zeki, Richard Frackowiak, Erik Lumer, Dave Chawla, Christian Büchel, Chris Frith, Cathy Price and the reviewers for their scientific input.

REFERENCES

Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. 1997 Synaptic depression and cortical gain control. *Science* **275**, 220–223.

Bear, M. F. & Singer, W. 1986 Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature* **320**, 172–176.

Büchel, C. & Friston, K. J. 1997 Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebr. Cortex* **7**, 768–778.

Crick, F. & Koch, C. 1998 Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* **391**, 245–250.

Desmedt, J. E. & Tomberg, C. 1994 Transient phase-locking of 40 Hz electrical oscillations in prefrontal and parietal human cortex reflects the process of conscious somatic perception. *Neurosci. Lett.* **168**, 126–129.

Edelman, G. M. 1993 Neural Darwinism: selection and reentrant signalling in higher brain function. *Neuron* **10**, 115–125.

Erb, M. & Aertsen, A. 1992 Dynamics of activity in biology-oriented neural network models: stability analysis at low firing rates. In *Information processing in the cortex. Experiments and theory* (ed. A. Aertsen & V. Braitenberg), pp. 201–223. Berlin: Springer.

Freeman, W. & Barrie, J. 1994 Chaotic oscillations and the genesis of meaning in cerebral cortex. In *Temporal coding in the brain* (ed. G. Buzsaki, R. Llinas, W. Singer, A. Berthoz & T. Christen), pp. 13–18. Berlin: Springer.

Friston, K. J. 1997 Transients metastability and neuronal dynamics. *NeuroImage* **5**, 164–171.

Friston, K. J., Tononi, G., Reeke, G. H., Sporns, O. & Edelman, G. E. 1994 Value-dependent selection in the brain: simulation in synthetic neural model. *Neuroscience* **39**, 229–243.

Friston, K. J., Tononi, G., Sporns, O. & Edelman, G. M. 1995 Characterising the complexity of neuronal interactions. *Hum. Brain Mapp.* **3**, 302–314.

Friston, K. J., Stephan, K.-M. & Frackowiak, R. S. J. 1997 Transient phase-locking and dynamic correlations: are they the same thing? *Hum. Brain Mapp.* **5**, 48–57.

Fuchs, A., Kelso, J. A. S. & Haken, H. 1992 Phase transitions in the human brain: spatial mode dynamics. *Int. J. Bifurcation Chaos* **2**, 917–939.

Gray, C. M. & Singer, W. 1989 Stimulus specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl Acad. Sci. USA* **86**, 1698–1702.

Horgan, J. 1995 From complexity to perplexity. *Sci. Am.* June, 74–79.

Jefferys, J. G. R., Traub, R. D. & Whittington, M. A. 1996 Neuronal networks for induced '40 Hz' rhythms. *Trends Neurosci.* **19**, 202–208.

Jones, D. S. 1979 *Elementary information theory.* Oxford, UK: Clarendon Press.

Kauffman, S. A. 1992 The sciences of complexity and 'origins of order'. In *The principles of organization in organisms* (ed. J. E. Mittenthal & A. B. Baskin), pp. 303–320. New York: Addison-Wesley.

Kelso, J. A. S. 1995 *Dynamic patterns: the self-organisation of brain and behaviour.* Cambridge, MA: MIT Press.

Le Van Quyen, M., Adam, C., Baulac, M., Martinerie, J. & Varela, F. J. 1998 Nonlinear interdependencies of EEG signals in human intracranially recorded temporal lobe seizures. *Brain Res.* **4**, 24–40.

McGaugh, J. 1992 Neuromodulatory systems and the regulation of memory storage. In *Neuropsychology of memory* (ed. L. Squire & N. Butters), pp. 386–401. New York: Guilford Press.

MacLeod, K. & Laurent, G. 1996 Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. *Science* **274**, 976–979.

Mayer-Kress, G., Barczys, C. & Freeman, W. 1991 Attractor reconstruction from event-related multi-electrode EEG-data. In *Mathematical approaches to brain functioning diagnostics* (ed. I. Dvorak & A. V. Holden), pp. 315–336. New York: Manchester University Press.

Mayr, E. 1982 *The growth of biological thought.* Cambridge, MA: Harvard University Press.

Metherate, R. & Weinberger, N. M. 1989 Acetylcholine produces stimulus-specific receptive field alterations in cat auditory system. *Brain Res.* **480**, 372–377.

Morgera, S. D. 1985 Information theoretic covariance complexity and its relation to pattern recognition. *IEEE Trans. Systems, Man & Cybern. SMC-15* **5**, 608–619.

Munk, M. H. J., Roelfsema, P. R., Konig, P., Engel, A. K. & Singer, W. 1996 Role of reticular activation in the modulation of intra-cortical synchronization. *Science* **272**, 271–277.

Pfurtscheller, G. & Aranibar, A. 1979 Evaluation of event-related desynchronisation (ERB) preceding and following voluntary self-paced movement. *Electroenceph. Clin. Neurophysiol*. **46**, 138–146.

Schiff, S. J., So, P., Chang, T., Burke, R. E. & Sauer, T. 1996 Detecting dynamical interdependence and generalized synchrony though mutual prediction in a neuronal ensemble. *Phys. Rev.* E **54**, 6708–6724.

Sporns, O., Gally, J. A., Reeke, G. N. & Edelman, G. M. 1989 Re-entrant signalling among simulated neuronal groups leads to coherence in their oscillatory activity *Proc. Natl Acad. Sci. USA* **86**, 7265–7269.

Steriade, M., Amzica, F. & Contreras, D. 1996 Synchronization of fast (30–40Hz) spontaneous cortical rhythms during brain activation. *J. Neurosci.* **16**, 392–417.

Sutton, R. S. & Barto, A. G. 1990 Time derivative models of Pavlovian reinforcement. In *Learning and computational neuroscience: foundations of adaptive networks* (ed. M. Gabriel & J. Moore), pp. 497–538, Cambridge, MA: MIT Press.

Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. 1996 Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in humans. *J. Neurosci.* **16**, 4240–4249.

Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. 1997 Oscillatory g-band (30–70Hz) activity induced by a visual search task in humans. *J. Neurosci*. **17**, 722–734.

Tononi, G., Sporns, O. & Edelman, G. M. 1994 A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl Acad. Sci. USA* **91**, 5033–5037.

Tsonis, A. A. 1992 *Chaos: from theory to applications*. New York: Plenum Press.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. & Aertsen, A. 1995 Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515–518.

Vanni, S., Revonuso, A. & Hari, R. 1997 Modulation of the parieto-occipital rhythm during object detection. *J. Neurosci.* **17**, 7141–7147.