INVITED PAPER

Cognitive Dynamics: From Attractors to Active Inference

This paper provides a link from statistical dynamics of ergodic systems to the inferential nature of human perception and ensuing action on the world.

By Karl Friston, Biswa Sengupta, and Gennaro Auletta

ABSTRACT | This paper combines recent formulations of self-organization and neuronal processing to provide an account of cognitive dynamics from basic principles. We start by showing that inference (and autopoiesis) are emergent features of any (weakly mixing) ergodic random dynamical system. We then apply the emergent dynamics to action and perception in a way that casts action as the fulfillment of (Bayesian) beliefs about the causes of sensations. More formally, we formulate ergodic flows on global random attractors as a generalized descent on a free energy functional of the internal states of a system. This formulation rests on a partition of states based on a Markov blanket that separates internal states from hidden states in the external milieu. This separation means that the internal states effectively represent external states probabilistically. The generalized descent is then related to classical Bayesian (e.g., Kalman–Bucy) filtering and predictive coding—of the sort that might be implemented in the brain. Finally, we present two simulations. The first simulates a primordial soup to illustrate the emergence of a Markov blanket and (active) inference about hidden states. The second uses the same emergent dynamics to simulate action and action observation.

KEYWORDS | Active inference; autopoiesis; cognitive; dynamics; free energy; random attractor; self-organization

I. INTRODUCTION

> "How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?"—Erwin Schrödinger (1943).

This paper draws on two recent developments in variational treatments of self-organization. The first is an application to Bayesian inference and embodied perception in the brain [1]. The second is an attempt to understand the nature of self-organization in random dynamical systems [2]–[6], in particular, our work on variational free-energy minimization [7]–[10]. Our premise is that biological self-organization is (almost) inevitable and manifests as a form of active Bayesian inference. We have previously suggested [11] that the events "within the spatial boundary of a living organism" [6] may arise from the very existence of a boundary or blanket, and that a Markov blanket may be inevitable under local coupling among dynamical systems. Here, we extend these arguments to provide a seamless progression from the basic behavior of random dynamical systems to formal (normative) accounts of the embodied or active inference that underlies cognition. To do this, we formulate flows in random dynamical systems in generalized coordinates of motion, relate this formulation to (filtering) procedures found in statistics and control theory, and then consider what this (neuronal) filtering would look like in the brain.

Most treatments of self-organization in theoretical biology use statistical thermodynamics or information theory to address the peculiar resistance of biological systems to the dispersive effects of external fluctuations [3], [5], [6], [8], [12]–[15]. We have tried to explain adaptive behavior in terms of minimizing an upper (free energy) bound on the surprise (negative log likelihood) of sensory samples [7], [16]. This minimization usefully connects the imperative for biological systems to maintain

Manuscript received July 3, 2013; accepted January 11, 2014. Date of publication March 14, 2014; date of current version March 25, 2014. This work was supported by Wellcome Trust.
K. Friston and **B. Sengupta** are with The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, London WC1N 3BG, U.K. (e-mail: k.friston@ucl.ac.uk; b.sengupta@ucl.ac.uk).
G. Auletta is with the Pontifical Gregorian University, Rome 4-00187, Italy (e-mail: auletta@unigre.it).
Digital Object Identifier: 10.1109/JPROC.2014.2306251

their sensory states within physiological bounds with an understanding of adaptive behavior in terms of inferring the causes of those states [17].

In brief, under ergodic assumptions, the long-term average of surprise is Shannon entropy. This means that minimizing free energy, through selectively sampling sensory input, places an upper bound on the entropy or dispersion of sensory states. This enables biological systems to resist the second law of thermodynamics, or more exactly the fluctuation theorem that applies to open systems far from equilibrium [18], [19]. However, because negative surprise is also *Bayesian model evidence*, systems that minimize free energy also maximize a lower bound on the evidence for an implicit model of how their sensory samples were generated. In statistics and machine learning, this is known as *approximate Bayesian inference* and provides a normative basis for the Bayesian brain hypothesis [20]–[24]. This suggests that biological systems act on the world to place an upper bound on the dispersion of their sensed states, while using those sensations to infer external states of the world. The resulting active inference is closely related to formulations in embodied cognition and artificial intelligence; for example, the perception action cycle [25], the use of predictive information [26]–[28], and homeokinetic formulations [29].

The ensuing (variational) free-energy principle has been applied widely in neurobiology [30]. The motivation for minimizing free energy has hitherto used the following sort of argument: systems that do not minimize free energy cannot exist, because the entropy of their sensory states would not be bounded and would increase indefinitely, by the fluctuation theorem [19]. Therefore, biological systems must minimize free energy. Here, we turn the argument around to suggest: any system that exists will appear to minimize free energy and therefore engage in active inference. What follows is an attempt to substantiate this argument using (heuristic) proofs and simulations.

This paper comprises seven sections. In addition to the Introduction, Section II considers the behavior of random dynamical systems, under the assumption that their flow is ergodic. The focus here is on flow in generalized coordinates of motion, which allows us to express any dynamics in terms of a Lagrangian or probabilistic model of flow. We try to relate this formulation to fundamental principles, such as the principle of stationary action that underpins most classical and quantum mechanics. Section III considers flows in systems with a Markov blanket that separates, in a statistical sense, internal from external states. We show that the dynamics of internal states can be expressed as a generalized gradient descent on a free-energy functional that is induced by a probability density encoding beliefs about external states. This generalized descent has two important implications. First, it endows internal dynamics with a representational interpretation. Second, the dynamics appears to place an upper bound on the Shannon entropy of the internal states and their Markov blanket, thereby preserving their structural and dynamical integrity. We consider this apparent representational autopoiesis as *active inference*. Section IV shows that the generalized descent is formally equivalent to Bayesian filtering and predictive coding schemes that have become popular for explaining neuronal dynamics. Section V illustrates the emergence of a Markov blanket and provides a proof of principle for the active inference, using simulations of coupled dynamical subsystems with chaotic flow. Section VI takes a complementary approach by using an explicit generalized descent on variational free energy. This allows one to interpret the dynamics in terms of action on the environment and, crucially, inference about the agency of that action. The final section, Section VII, includes the conclusion.

II. GENERALIZED DYNAMICS

We consider (weakly mixing) ergodic random dynamical systems described by stochastic differential equations of the following form:

$$\dot{\tilde{x}} = f(\tilde{x}) + \tilde{\omega}. \tag{1}$$

Here, the flow of generalized states $f(\tilde{x})$ is subject to random fluctuations $\tilde{\omega}$. Generalized states $\tilde{x} = (x, x', x'', \ldots)$ comprise the states *per se*, their motion, velocity, acceleration, and so on. This is a slightly unusual construction because the generalized motion $D\tilde{x} = (x', x'', x''', \ldots)$ is not necessarily the motion of generalized states $\dot{\tilde{x}} = (\dot{x}, \dot{x}', \dot{x}'', \ldots)$. In essence, these equations specify probability distributions over paths in generalized coordinates of motion and can be regarded as a probabilistic model of system dynamics that accommodates analytic (smooth) fluctuations.

Because the system is ergodic (and weakly mixing) it will, after a sufficient amount of time, converge to an invariant set of states called *a pullback* or *random global attractor* [31], [32]. The associated ergodic density $p(\tilde{x}|m)$ for any system or model m is the solution to the Fokker–Planck equation (also knows as the Kolmogorov forward equation) [33] describing the time evolution of the probability density over states

$$\dot{p}(\tilde{x}|m) = \nabla \cdot (\Gamma\nabla - f)p. \tag{2}$$

Here, the diffusion tensor Γ is the half the covariance (amplitude) of the random fluctuations. Equation (2) shows that the ergodic density depends upon flow, which can always be expressed in terms of curl-free and divergence-free components. This is the Helmholtz decomposition (also known as the fundamental theorem of vector calculus) and can be formulated in terms of an

antisymmetric matrix $Q(\tilde{x}) = -Q(\tilde{x})^T$ and a scalar potential $L(\tilde{x})$ that, as we will see below, plays a role of a Lagrangian [34]

$$f(\tilde{x}) = (Q - \Gamma)\nabla L(\tilde{x}). \tag{3}$$

Using this standard form [35], it is straightforward to show that $p(\tilde{x}|m) = \exp(-L(\tilde{x}))$ is the solution to the Fokker Planck equation [16]

$$p(\tilde{x}|m) = \exp(-L(\tilde{x})) \Rightarrow \nabla p = -p\nabla L \Rightarrow \dot{p} = 0. \tag{4}$$

This is an important result because it shows the flow can be decomposed into a component that flows toward regions with a higher ergodic density (the curl-free or irrotational component) and an orthogonal (divergence-free or solenoidal) component that circulates on isocontours of the ergodic density. Heuristically, this component can neither act as source nor sink and essentially conserves volume. The divergence-free flow plays a crucial role as the basis of generalized motion that creates a space-filling attractor that can have a low measure or volume. This can be seen clearly by substituting the above decomposition into the equations of motion to produce the following lemma.

Lemma (Generalized Descent): Any ergodic random dynamical system can be expressed in the form of a generalized gradient descent

$$\dot{\tilde{x}} = D\tilde{x} - \Gamma \cdot \nabla L(\tilde{x}) + \tilde{\omega}. \tag{5}$$

Proof: By construction, the divergence of generalized motion $\nabla \cdot \tilde{x} = \partial x'/\partial x + \partial x''/\partial x' + \ldots = 0$ is zero, which means generalized motion corresponds to divergence-free flow

$$\begin{aligned} D\tilde{x} &= Q\nabla L(\tilde{x}) \Rightarrow \\ f(\tilde{x}) &= (Q - \Gamma)\nabla L(\tilde{x}) = D\tilde{x} - \Gamma \cdot \nabla L(\tilde{x}). \end{aligned} \tag{6}$$

Substitution into (1) gives (5). □

Remarks: In short, the generalized motion $D\tilde{x}$ corresponds to the conservative divergence-free flow. Intuitively, this formulation casts motion as a stochastic gradient ascent in a frame of reference that moves with the generalized motion $\dot{\tilde{x}} - D\tilde{x} = -\Gamma \cdot \nabla L(\tilde{x}) + \tilde{\omega}$. The generalized descent lemma is quite important for our purposes. It shows that one can either specify a system probabilistically in terms of its equations of motion [see (1)] or a Lagrangian [see (5)]. Sections V and VI illustrate the emergent properties of random dynamical systems using these complementary formulations of flow.

Heuristically, one can picture the flow as a drift toward regions of high ergodic density in a direction that is orthogonal to the divergence-free flow. When the flow is conservative, one recovers classical equations of motion because $\Gamma = \tilde{\omega} = 0$, and the motion of the generalized states reduces to the (conservative) generalized motion of the states: $\dot{\tilde{x}} = D\tilde{x} = Q\nabla L(\tilde{x})$. For example, with the Lagrangian $L(\tilde{x}) = \varphi(x) + (1/2)x'^2$, we have

$$\begin{aligned} Q &= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \Rightarrow \\ \dot{\tilde{x}} &= D\tilde{x} = Q\nabla L(\tilde{x}) = \begin{bmatrix} x' \\ x'' \end{bmatrix} = \begin{bmatrix} x' \\ -\partial_x \varphi(x) \end{bmatrix}. \end{aligned} \tag{7}$$

This corresponds to Newtonian dynamics, where the gradient of a potential energy $\varphi(x)$ exerts a force on a body of unit mass. Fig. 1 provides an example using a double well potential. The key thing to note here is that the (classical) Lagrangian is just the (negative log) probability of generalized states at (nonequilibrium) steady state. This provides a nice connection between the most likely path and the principle of stationary action.

Corollary (Principle of Stationary Action): The most likely path $\hat{x}(t)$ of any ergodic random dynamical system satisfies the principle of stationary action $\delta_{\tilde{\mu}} S = 0$, where action is the path integral of the Lagrangian

$$S = \int_{t=0}^{t=T} dt L(\tilde{x}(t)) = -\int_{t=0}^{t=T} dt \ln p(\tilde{x}|m). \tag{8}$$

Proof: The most likely motion of any generalized state is the motion under the most likely fluctuation $\tilde{\omega} = 0$, where

$$\dot{\hat{x}} = D\hat{x} - \Gamma \cdot \partial_{\hat{x}} L(\hat{x}). \tag{9}$$

Furthermore, the most likely generalized state minimizes $L(\tilde{x}) = -\ln p(\tilde{x}|m)$, such that (by the fundamental lemma of variational calculus)

$$\partial_{\hat{x}} L(\hat{x}) = 0 \Longleftrightarrow \delta_{\hat{x}} S = 0 \Longleftrightarrow \dot{\hat{x}} = D\hat{x}.$$

This means that small variations around the most likely path do not change action. Furthermore, the most likely path is self-consistent in that its generalized motion is the motion of the generalized states. □

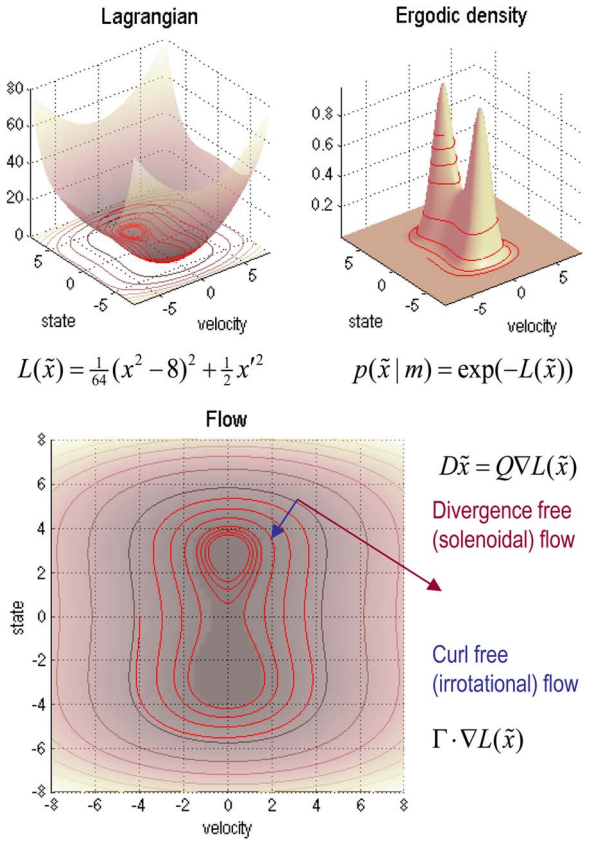

Fig. 1. ***Lagrangian dynamics and flow. This schematic illustrates the flows prescribed by a Lagrangian. The upper right panel shows the Lagrangian as a function of a single state and its velocity. This example illustrates a double well potential of the sort used in quantum mechanics. The red line corresponds to a trajectory that circulates on the isocontours of the Lagrangian, while drifting slowly downward. This trajectory was obtained using low amplitude random fluctuations: $\Gamma = (1/64)$. The upper right panel shows the same results but plotted on the ergodic density. The lower panel illustrates the decomposition of flow into orthogonal divergence-free components (that follow the isocontours) and curl-free components that drift toward regions with high ergodic density.***

This formulation can be related to the path integral formulation of quantum mechanics and, in particular, the Feynman interpretation of the probability of a path in terms of action [36]. Heuristically, one can see how the path integral of the Lagrangian corresponds to the accumulated (log) likelihood of that path [see (8)]. Interestingly the Schrödinger equation can be recovered from the path integral formulation, and both are essentially reformulations of the Fokker Planck equation.

In summary, this section has formulated dynamical random systems in generalized coordinates of motion and has established the construct validity of this formulation in terms of classical Lagrangian mechanics and path integral formulations. The key result that we will call on below is that any ergodic random dynamical system can be formulated as a generalized ascent on the log likelihood of its trajectories. In Section III, we will interpret this likelihood as a marginal likelihood and see that any random dynamical system can be interpreted as performing some form of inference on itself.

III. DYNAMICS AND ACTIVE INFERENCE

This section pursues the following lemma: any ergodic random dynamical system that possesses a Markov blanket will appear to actively maintain its structural and dynamical integrity. A Markov blanket is a set of states that separates two other sets in a statistical sense. The term Markov blanket was introduced in the setting of Bayesian networks or graphs [37] and refers to the children of a set (the set of states that are influenced), its parents (the set of states that influence it), and the parents of its children.

A Markov blanket induces a partition of states into *internal states* and *external states* that are hidden (insulated) from the internal (insular) states by the Markov blanket. For example, the surface of a cell may constitute a Markov blanket separating intracellular (internal) and extracellular (external) states [9], [11]. Statistically speaking, external states can only be seen vicariously by the internal states, through the Markov blanket. The Markov blanket can itself be partitioned into two sets that are, and are not, children of external states. We will refer to these as surface or *sensory states* and *active states*, respectively. Put simply, the existence of a Markov blanket $S \times A$ implies a partition of states into external, sensory, active, and internal states: $\tilde{x} \in X = \Psi \times S \times A \times R$. External states cause sensory states that influence, but are not influenced by, internal states, while internal states cause active states that influence, but are not influenced by, external states (see Table 1). Crucially, the dependencies induced by Markov blankets create a circular causality that is reminiscent of the action–perception cycle (see Fig. 2). The circular causality here means that external states cause changes in internal states, via sensory states, while the internal states couple back to external states through active states, such that internal and external states cause each other in a reciprocal fashion.

Equipped with this partition, we can now consider the dependencies among states implied by the Markov blanket, in terms of their equations of motion

$$\begin{aligned} f_\psi(\tilde{\psi}, \tilde{s}, \tilde{a}) &= (\Gamma - Q)\nabla_{\tilde{\psi}} \ln p(\tilde{\psi}, \tilde{s}, \tilde{a}, \tilde{r}|m) \\ f_s(\tilde{\psi}, \tilde{s}, \tilde{a}) &= (\Gamma - Q)\nabla_{\tilde{s}} \ln p(\tilde{\psi}, \tilde{s}, \tilde{a}, \tilde{r}|m) \\ f_r(\tilde{s}, \tilde{a}, \tilde{r}) &= (\Gamma - Q)\nabla_{\tilde{r}} \ln p(\tilde{\psi}, \tilde{s}, \tilde{a}, \tilde{r}|m) \\ f_a(\tilde{s}, \tilde{a}, \tilde{r}) &= (\Gamma - Q)\nabla_{\tilde{a}} \ln p(\tilde{\psi}, \tilde{s}, \tilde{a}, \tilde{r}|m). \end{aligned} \quad (10)$$

As noted in [11], there is something rather curious about this flow: it seems as if the flow of each subset of states knows the value of subsets that are hidden from it. For

Table 1 Definitions of the Tuple $(\Omega, \Psi, S, A, \Lambda, p, q)$ Underlying Active Inference

- *A sample space* Ω or non-empty set from which random fluctuations or outcomes $\omega \in \Omega$ are drawn
- *External states* $\Psi : \Psi \times A \times \Omega \to \mathbb{R}$ – hidden states of the world that cause sensory states and depend on action
- *Sensory states* $S : \Psi \times A \times \Omega \to \mathbb{R}$ – the agent's sensations that constitute a probabilistic mapping from action and external states
- *Active states* $A : S \times R \times \Omega \to \mathbb{R}$ – an agent's action that depends on its sensory and internal states
- *Internal states* $R : R \times S \times \Omega \to \mathbb{R}$ – representational states of the agent that cause action and depend on sensory states
- *Ergodic density* $p(\tilde{\psi}, \tilde{s}, \tilde{a}, \tilde{r} \mid m)$ – a probability density function over external $\tilde{\psi} \in \Psi$, sensory $\tilde{s} \in S$, active $\tilde{a} \in A$ and internal states $\tilde{r} \in R$ for a system denoted by m
- *Variational density* $q(\tilde{\psi} \mid \tilde{r})$ – an arbitrary probability density function over external states that is parameterized by internal states

example, the flow of internal states and action appear to be guided by the ergodic density over hidden external states, and yet their flow is not a function of hidden states. This apparent paradox can be finessed by noting that the flow from any state is the expected motion averaged over time. By the ergodic theorem [38], this is also the flow averaged over the external states, which does not depend on the external state at any particular time. More formally, the flow through any point $(\tilde{s}, \tilde{a}, \tilde{r})$ in the space of the internal states and their Markov blanket is [11]

$$\begin{aligned} f_r(\tilde{s}, \tilde{a}, \tilde{r}) &= \int_{\Psi} p(\tilde{\psi}|\tilde{s}, \tilde{a}, \tilde{r})(\Gamma - Q)\nabla_{\tilde{r}} \ln p(\tilde{x}|m) d\psi \\ &= (\Gamma - Q)\nabla_{\tilde{r}} \ln p(\tilde{s}, \tilde{a}, \tilde{r}|m) \\ f_a(\tilde{s}, \tilde{a}, \tilde{r}) &= \int_{\Psi} p(\tilde{\psi}|\tilde{s}, \tilde{a}, \tilde{r})(\Gamma - Q)\nabla_{\tilde{a}} \ln p(\tilde{x}|m) d\psi \\ &= (\Gamma - Q)\nabla_{\tilde{a}} \ln p(\tilde{s}, \tilde{a}, \tilde{r}|m). \end{aligned} \tag{11}$$

This shows that the flow of internal and active states performs a circuitous gradient ascent on the *marginal* ergodic density over internal states and their Markov blanket. This means that the internal states will appear to respond to sensory fluctuations based on posterior beliefs about underlying fluctuations in external states. We can formalize this notion by associating these beliefs with a probability density over external states $q(\tilde{\psi}|\tilde{r})$ that is encoded (parameterized) by internal states.

Lemma (Free Energy): for any random dynamical system with a Markov blanket and Lagrangian $L(\tilde{x}) = -\ln p(\tilde{\psi}, \tilde{s}, \tilde{a}, \tilde{r})$, there is a free energy $F(\tilde{s}, \tilde{a}, \tilde{r})$ that describes the flow of internal and active states as a generalized descent

$$\begin{aligned} f_r(\tilde{s}, \tilde{a}, \tilde{r}) &= (Q - \Gamma)\nabla_{\tilde{r}} F = D\tilde{r} - \Gamma \cdot \nabla_{\tilde{r}} F \\ f_a(\tilde{s}, \tilde{a}, \tilde{r}) &= (Q - \Gamma)\nabla_{\tilde{a}} F = D\tilde{a} - \Gamma \cdot \nabla_{\tilde{a}} F \\ F(\tilde{s}, \tilde{a}, \tilde{r}) &= E_q[L(\tilde{x})] - H[q(\psi|\tilde{r})]. \end{aligned} \tag{12}$$

This free energy is a functional of a variational density $q(\tilde{\psi}|\tilde{r})$, parameterized by internal states, that corresponds

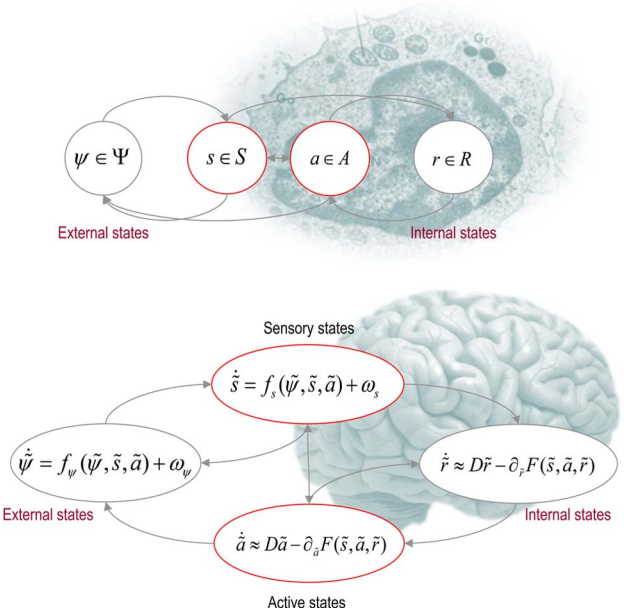

Fig. 2. ***Markov blankets and the free-energy principle. These schematics illustrate the partition of states into internal states and hidden or external states that are separated by a Markov blanket, comprising sensory and active states. The upper panel shows this partition as it might be applied to a cell: where the internal states can be associated with the intracellular states of a cell, while sensory states become the surface states or cell membrane overlying active states (e.g., the actin filaments of the cytoskeleton). The lower panel shows the same dependencies but rearranged so that they can be related to action and perception in the brain: where active and internal states minimize a free energy functional of sensory states. The ensuing self-organization of internal states then corresponds to perception, while action couples brain states back to external states. See Table 1 for a definition of variables.***

to the expected Lagrangian minus the entropy of the variational density.

Proof: Using Bayes rule, we can rearrange the expression for free energy in terms of a Kullback–Leibler divergence [39]

$$F(\tilde{s},\tilde{a},\tilde{r}) = -\ln p(\tilde{s},\tilde{a},\tilde{r}|m) + D_{KL}[q(\psi|\tilde{r})\|p(\psi|\tilde{s},\tilde{a},\tilde{r})]. \quad (13)$$

When $q(\tilde{\psi}|\tilde{r}) = p(\tilde{\psi}|\tilde{s},\tilde{a},\tilde{r})$, the divergence term disappears from (12) and we recover the ergodic flow in (11)

$$f_r(\tilde{s},\tilde{a},\tilde{r}) = (\Gamma - Q)\nabla_{\tilde{r}} \ln p(\tilde{s},\tilde{a},\tilde{r}|m)$$
$$f_a(\tilde{s},\tilde{a},\tilde{r}) = (\Gamma - Q)\nabla_{\tilde{a}} \ln p(\tilde{s},\tilde{a},\tilde{r}|m). \quad (14)$$

In other words, the ergodic flow ensures that the variational density is the posterior density, such that the variational density represents the hidden states in a Bayes-optimal sense. □

Remarks: All this proof says is that if one interprets internal states as parameterizing Bayesian beliefs about external states, then the dynamics of internal and active states can be described as a gradient descent on a variational free-energy function of internal states and their Markov blanket. Variational free energy was introduced by Feynman to solve difficult integration problems in path integral formulations of quantum physics [40]. This is also the free-energy bound that is used extensively in *approximate Bayesian inference* (e.g., variational Bayes) [39], [41], [42]. The expression for free energy in (13) discloses its Bayesian interpretation: the first term is the negative log evidence or *marginal likelihood* of the internal states and their Markov blanket. The second term is a *relative entropy* or Kullback–Leibler divergence [43] between the variational density and the posterior density over external states. Because (by Gibbs equality) this divergence cannot be less than zero, the internal flow will appear to have minimized the divergence between the variational and posterior density. In other words, the internal states will appear to have solved the problem of Bayesian inference by encoding posterior beliefs about hidden (external) states, under a generative model provided by the Lagrangian. This is known as *exact* Bayesian inference because the variational and posterior densities are identical. In Section IV, we will consider approximate forms (under the Laplace assumption) leading to *approximate* Bayesian inference. In short, the internal states will appear to engage in Bayesian inference: but what about action?

Because the divergence in (13) can never be less than zero, free energy is an upper bound on the negative log evidence. Now, because the system is ergodic, we have

$$F(\tilde{s},\tilde{a},\tilde{r}) \geq -\ln p(\tilde{s},\tilde{a},\tilde{r}|m) \Rightarrow$$
$$E_t[F(\tilde{s},\tilde{a},\tilde{r})] \geq E_t[-\ln p(\tilde{s},\tilde{a},\tilde{r}|m)] = H[p(\tilde{s},\tilde{a},\tilde{r}|m)]. \quad (15)$$

This means that action will (on average) appear to place an upper bound on the entropy of the internal states and their Markov blanket. Together with the Bayesian modeling perspective, this is exactly consistent with the good regulator theorem (every good regulator is a model of its environment) and related treatments of self-organization [2], [5], [16], [44], [45]. Furthermore, we have shown elsewhere [7], [30] that free-energy minimization is consistent with information-theoretic formulations of sensory processing and behavior [27], [46], [47]. Equation (12) also shows that minimizing free energy entails maximizing the entropy of the variational density, in accord with the maximum entropy principle [48]. Finally, because we have cast this treatment in terms of random dynamical systems, there is an easy connection to dynamical formulations that predominate in the neurosciences; e.g., [45] and [49]–[51]. In summary, for any ergodic random dynamical system, we have the following.

- The existence of a Markov blanket necessarily implies a partition of states into internal states, their Markov blanket (sensory and active states) and external or hidden states.
- This partition endows internal states with the apparent capacity to represent hidden states probabilistically, so that they appear to infer the hidden causes of sensory states (by minimizing a free-energy bound on log Bayesian evidence). By the circular causality induced by the Markov blanket, sensory states depend on active states, rendering inference active or embodied.
- Because active states change, but are not changed by, hidden states (see Fig. 2), they will appear to place an upper (free-energy) bound on the dispersion (entropy) of internal states and their Markov blanket. This means active states will appear to maintain the structural and functional integrity of the Markov blanket. See also [52].

In Section IV, we consider this dynamics from the point of view of Bayesian (e.g., Kalman) filtering and its implementation in the brain in terms of predictive coding.

IV. BAYESIAN FILTERING AND PREDICTIVE CODING

In Section III, we saw how the ergodic flow decomposes into divergence and curl-free components, where the divergence-free component corresponds to generalized flow. This component cannot change the (free-energy

bound on) log evidence or marginal likelihood, whereas the curl-free component increases the marginal likelihood. Readers familiar with time-series analysis will recognize the formal similarity between these two components and the prediction and update terms in Bayesian (e.g., Kalman) filtering. In this section, we pursue this formal equivalence and see how it can illuminate cognitive (neuronal) dynamics, in terms of message passing in the brain.

The free-energy lemma means that we can express the dynamics of internal and active states in terms of variational free energy where (assuming random fluctuations are small in relation to free-energy gradients)

$$\begin{aligned}\dot{\tilde{r}} &= D\tilde{r} - \Gamma \cdot \partial_{\tilde{r}} F(\tilde{s}, \tilde{a}, \tilde{r}) \\ \dot{\tilde{a}} &= D\tilde{a} - \Gamma \cdot \partial_{\tilde{a}} F(\tilde{s}, \tilde{a}, \tilde{r}).\end{aligned} \tag{16}$$

This form reveals its close connection to inference schemes used in the Bayesian inversion of state–space models [53]. To illustrate this clearly, we will look at a special case of generalized descent: under the (Laplace) assumption that the variational density $q(\tilde{\psi}|\tilde{r}) = N(\tilde{\mu}, C)$ is Gaussian with sufficient statistics $\tilde{r} = \{\tilde{\mu}, C\}$. Under this assumption, the solution of (16) for the variational precision is $C^{-1} = \Pi = \partial_{\tilde{\mu}\tilde{\mu}} L(\tilde{\mu}) \Rightarrow \partial_C F = 0$ [54]. This means that free energy can be reduced to a function of the variational mean (omitting constants)

$$F = L(\mu) + \frac{1}{2}\ln\left|\partial_{\tilde{\mu}\tilde{\mu}} L(\tilde{\mu})\right|. \tag{17}$$

This provides an important simplification that enables us to associate internal states $\tilde{\mu} = \tilde{r}$ with the variational or (approximate) posterior expectation of the hidden states, where

$$\begin{aligned}\dot{\tilde{\mu}} &= D\tilde{\mu} - \Gamma \cdot \partial_{\tilde{\mu}} F(\tilde{s}, \tilde{a}, \tilde{\mu}) \\ \dot{\tilde{a}} &= D\tilde{a} - \Gamma \cdot \partial_{\tilde{a}} F(\tilde{s}, \tilde{a}, \tilde{\mu}).\end{aligned} \tag{18}$$

The first equality is known as a generalized filter [53], which has classical filtering as a special case. Classical filtering under Markovian or Weiner assumptions is equivalent to assuming the precision of the motion of random fluctuations is zero. In this limiting case, one only has to consider the states and their first derivative. This means generalized filtering takes the form of (extended) Kalman–Bucy filtering, with the usual prediction and correction terms [55]

$$\begin{aligned}\dot{\mu} &= \mu' - \Gamma \cdot \partial_{\mu} F(s, s', \mu, \mu') \\ \dot{\mu}' &= -\Gamma \cdot \partial_{\mu'} F(s, s', \mu, \mu').\end{aligned} \tag{19}$$

Having established the formal connection to classical Bayesian filtering, we now turn to hierarchical Bayesian filtering and predictive coding of the sort that may underlie cognitive dynamics in the brain.

A. Generalized Filtering and Predictive Coding

Generalized filtering is usually used to invert hierarchical models or systems of the following form:

$$\begin{aligned}s &= g^{(1)}\left(\psi_u^{(1)}, \psi_v^{(1)}\right) + \omega_u^{(1)} \\ \dot{\psi}_v^{(1)} &= f^{(1)}\left(\psi_u^{(1)}, \psi_v^{(1)}\right) + \omega_v^{(1)} \\ &\vdots \\ \dot{\psi}_u^{(i-1)} &= g^{(i)}\left(\psi_u^{(i)}, \psi_v^{(i)}\right) + \omega_u^{(i)} \\ \dot{\psi}_v^{(i)} &= f^{(i)}\left(\psi_u^{(i)}, \psi_v^{(i)}\right) + \omega_v^{(i)} \\ &\vdots\end{aligned} \tag{20}$$

Gaussian assumptions about the random fluctuations $\tilde{\omega}$ prescribe the likelihood and empirical priors on the generalized motion of hidden states that define the Lagrangian or generative model, where (suppressing action)

$$\begin{aligned}L(\tilde{x}) &= -\ln p(\tilde{\psi}, \tilde{s}|m) \\ p(\tilde{\psi}, \tilde{s}|m) &= \prod_i p\left(\tilde{\psi}_u^{(i-1)}|\tilde{\psi}_u^{(i)}, \tilde{\psi}_v^{(i)}\right) p \\ &\quad \times \left(D\tilde{\psi}_v^{(i)}|\tilde{\psi}_u^{(i)}, \tilde{\psi}_v^{(i)}\right) p\left(\tilde{\psi}_v^{(1)}|m\right) \\ p\left(\tilde{\psi}_u^{(i-1)}|\tilde{\psi}_u^{(i)}, \tilde{\psi}_v^{(i)}\right) &= \mathcal{N}\left(\tilde{g}^{(i)}, \Sigma_u^{(i)}\right) \\ p\left(D\tilde{\psi}_u^{(i)}|\tilde{\psi}_u^{(i)}, \tilde{\psi}_v^{(i)}\right) &= \mathcal{N}\left(\tilde{f}^{(i)}, \Sigma_v^{(i)}\right).\end{aligned} \tag{21}$$

In hierarchical form, hidden states $\psi_u^{(i)} \in \psi$ couple hierarchical levels (where $\tilde{s} \triangleq \tilde{\psi}_u^{(0)}$) and the dynamics of hidden states $\psi_v^{(i)} \in \psi$ within each level confer memory on the system. The ensuing generalized gradient descent on free energy can then be expressed compactly in terms of prediction errors, where (suppressing high-order terms)

$$\begin{aligned}\dot{\tilde{\mu}}_u^{(i)} &= D\tilde{\mu}_u^{(i)} - \frac{\partial \tilde{\varepsilon}^{(i)}}{\partial \tilde{\mu}_u^{(i)}} \cdot \Pi^{(i)} \tilde{\varepsilon}^{(i)} - \Pi_u^{(i+1)} \tilde{\varepsilon}_u^{(i+1)} \\ \dot{\tilde{\mu}}_v^{(i)} &= D\tilde{\mu}_v^{(i)} - \frac{\partial \tilde{\varepsilon}^{(i)}}{\partial \tilde{\mu}_v^{(i)}} \cdot \Pi^{(i)} \tilde{\varepsilon}^{(i)} \\ \tilde{\varepsilon}_u^{(i)} &= \tilde{\mu}_u^{(i-1)} - \tilde{g}^{(i)}\left(\tilde{\mu}_u^{(i)}, \tilde{\mu}_v^{(i)}\right) \\ \tilde{\varepsilon}_v^{(i)} &= D\tilde{\mu}_v^{(i-1)} - \tilde{f}^{(i)}\left(\tilde{\mu}_u^{(i)}, \tilde{\mu}_v^{(i)}\right).\end{aligned} \tag{22}$$

Here, $\Pi^{(i)}$ is the precision (inverse covariance $\Sigma^{(i)}$) of random fluctuations at the ith level. This is known as generalized predictive coding [56], with linear predictive coding [57] as a special case. Predictive coding has become a popular metaphor for message passing in the brain: In neural network terms, (22) says that error units receive predictions from the same hierarchical level $\tilde{\mu}_u^{(i-1)}$ and the level above $\tilde{\mu}_u^{(i)}$. Conversely, posterior expectations are driven by prediction errors from the same level $\tilde{\varepsilon}^{(i+1)}$ and the level below $\tilde{\varepsilon}_u^{(i)}$. These constitute bottom-up and lateral messages that drive expectations toward a better prediction to reduce the prediction error in the level below. This is the essence of recurrent message passing between hierarchical levels to suppress free energy or prediction error; see [56] for a detailed discussion. This form of generalized filtering is used routinely in the analysis of time-series data [53] and is much more computationally efficient than equivalent sampling schemes such as particle filtering. This efficiency is largely due to the Laplace assumption that is inherent in schemes like Kalman filters.

In neurobiological implementations of predictive coding [58], the sources of bottom-up prediction errors are thought to be superficial pyramidal cells that send forward connections to higher cortical areas. Conversely, predictions are conveyed from deep pyramidal cells by backward connections, to target (polysynaptically) the superficial pyramidal cells encoding prediction error [59]. Equation (22) shows how precision $\Pi^{(i)}$ plays an important role in weighting the influence of prediction errors at any particular level of the hierarchy. In other words, by changing the precision on the prediction errors, one can bias inference toward sensory information or top–down predictions. In the current context, precision corresponds to the gain of (superficial pyramidal) populations encoding prediction error and has been discussed as mediating attention and action selection [60].

B. Active Inference and Action

In neurobiological implementations of active inference, posterior expectations elicit action by sending predictions down the hierarchy to be unpacked into proprioceptive predictions at the level of (pontine) cranial nerve nuclei and spinal cord. These engage classical reflex arcs to suppress proprioceptive prediction errors and produce the predicted motor trajectory, where (suppressing generalized motion)

$$\dot{a} = -\partial_a F = -\partial_a \tilde{s} \cdot \Pi_u^{(1)} \tilde{\varepsilon}_u^{(1)}. \tag{23}$$

The reduction of action to classical reflexes follows because the only way that action can minimize free energy is to change sensory (proprioceptive) prediction errors by changing sensory signals; cf., the equilibrium point formulation of motor control [61]. In short, active inference can be regarded as equipping a generalized predictive coding scheme with classical reflex arcs. See Fig. 3.

In summary, the dynamics of internal states and their Markov blanket of any ergodic random dynamical system can be interpreted as performing generalized Bayesian filtering of its sensory states. This provides an interesting perspective on Bayesian filtering as finding the most likely path of internal states that are coupled to hidden states through sensory states. The key trick here is to render the internal states a representation of hidden states through assuming a particular parameterization of the variational density; for example, the Laplace assumption. In neurobiological terms, this can be implemented fairly simply with predictive coding that, when equipped with classical reflexes, provides a principled metaphor for action and perception in the brain. We now turn to simulations of generalized dynamics to illustrate the emergence of inferential self-organization at the level of simple cell-like structures and at the level of cognitive dynamics in the brain.

V. ACTIVE INFERENCE AND SELF-ORGANIZATION

Section II suggested that dynamical systems can either be specified in terms of their equations of motion or in terms of a generalized descent on a Lagrangian. In this section, we take some arbitrary equations of motion and examine the ensuing dynamics for evidence of active inference, using the criteria established at the end of Section III. In Section VI, we will take the complementary approach and examine the behavior of systems specified by a Lagrangian or probabilistic model of dynamics.

Here, we consider simulations of a primordial soup reported in [11] to illustrate the emergence of active inference of a simple and prebiotic sort. This soup comprises an ensemble of dynamical subsystems, each with its own structural and functional states, that are coupled through short-range interactions. These simulations are similar to (hundreds of) simulations used to characterize pattern formation in dissipative systems; for example, Turing instabilities [62] and reaction–diffusion systems such as the Belousov–Zhabotinsky reaction [63]. The simulations considered here are solutions to stochastic differential equations for coupled structural and functional states. In other words, we consider states from classical mechanics that determine physical motion, and functional states that could describe electrochemical states. The agenda here is not to explore the repertoire of patterns and self-organization these ensembles exhibit, but rather take an arbitrary example and show that, buried within it, there is a clear and discernible dynamical structure that satisfies the criteria for active inference.

A. The Primordial Soup

To simulate a primordial soup we used an ensemble of elemental subsystems with Newtonian and electrochemical

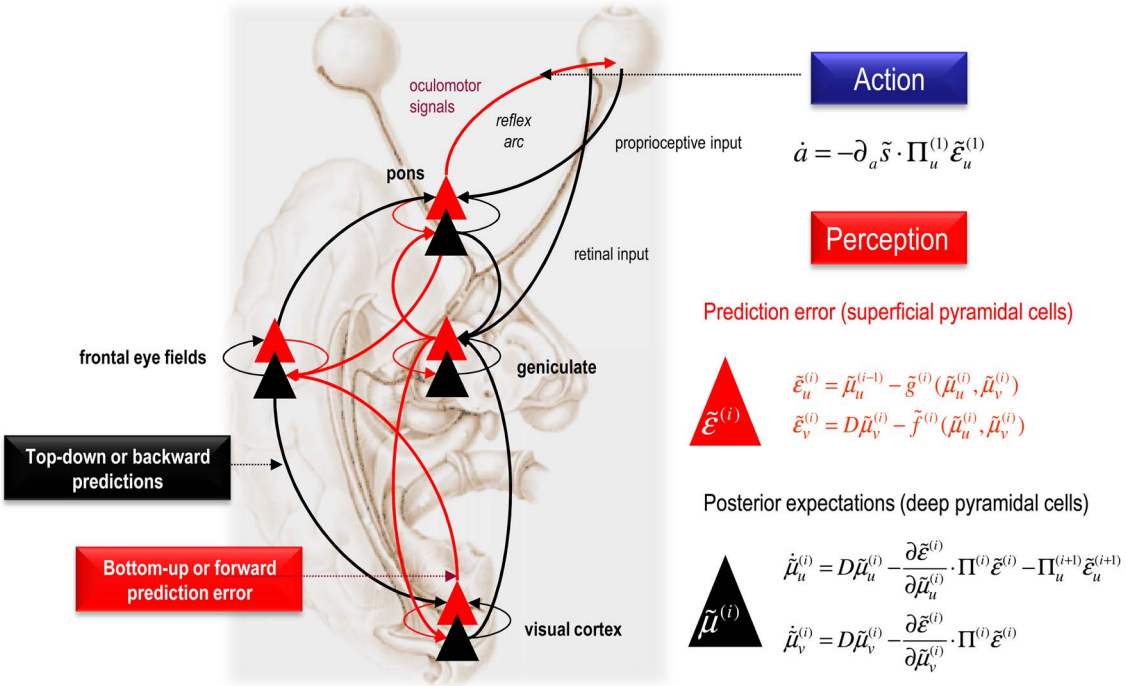

Fig. 3. ***Hierarchical message passing in the oculomotor system. Schematic detailing a neuronal message passing scheme (generalized Bayesian filtering or predictive coding) that optimizes posterior expectations about hidden states of the world, given sensory (visual) data and the active (oculomotor) sampling of those data. This schematic shows the cells of origin of forward driving connections (in red) that convey prediction error from a lower area to a higher area and the backward connections (in black) that construct predictions [58]. These predictions try to explain away prediction error in lower levels. In this scheme, the sources of forward and backward connections are superficial (red) and deep (black) pyramidal cells respectively [59]. The equations on the right-hand side represent a generalized descent on free energy under the hierarchical model described in the main text. In this example, visual input is passed to the lateral geniculate nuclei (LGN) and to higher visual (e.g., V1) and prefrontal (e.g., frontal eye fields) areas in the form of prediction errors. Crucially, proprioceptive sensations are also predicted, creating prediction errors at the level of the cranial nerve nuclei (pons). The special aspect of these proprioceptive prediction errors is that they can be resolved through classical reflex arcs; in other words, they can elicit action to change the direction of gaze and close the visual oculomotor loop.***

dynamics $\tilde{x} = (\tilde{p}, \tilde{q})$

$$\begin{aligned} \dot{\tilde{p}} &= f_p(\tilde{p}, \tilde{q}) + \tilde{\omega} \\ \dot{\tilde{q}} &= f_q(\tilde{p}, \tilde{q}) + \tilde{\omega}. \end{aligned} \tag{24}$$

Here, $\tilde{p}(t)$ describes position and motion, while $\tilde{q}(t)$ corresponds to electrochemical states. One can think of these generalized states as describing the physical and electrochemical state of large macromolecules. Crucially, these states are coupled within and between subsystems comprising an ensemble. The electrochemical dynamics were chosen to have a Lorenz attractor, as indicated in Fig. 4. Changes in electrochemical states are coupled through the local average of the states of subsystems $\overline{q}^{(i)}$ within a Euclidean distance of one. This spatial dependency is mediated by an (unweighted) adjacency matrix A that encodes the dependencies among the functional (electrochemical) states of the ensemble. The local average enters the equations of motion both linearly and nonlinearly to provide an opportunity for generalized synchronization [64].

The Lorenz form for these dynamics was a somewhat arbitrary choice but provides a ubiquitous model of electrodynamics, lasers, and chemical reactions [65]. Each subsystems rate parameter $\kappa^{(i)} = (1/32)(1 - \exp(-4 \cdot I))$ was selected randomly, where $U \in (0, 1)$ was selected from a uniform distribution. This introduces heterogeneity in the rate of electrochemical dynamics, with a large number of fast subsystems, with a rate constant of nearly one, and a small number of slower subsystems. To augment this heterogeneity, we randomly selected a third of the subsystems and prevented them from (electrochemically) influencing others, by setting the appropriate column of the adjacency matrix to zero. We refer to these as functionally *closed* systems.

The Newtonian motion of each subsystem also depends upon the functional status of its neighbors. This motion rests on forces $\varphi^{(i)}$ exerted by other subsystems that comprise a strong repulsive force (with an inverse square law) and a weaker attractive force that depends on their electrochemical states. This force was chosen so that

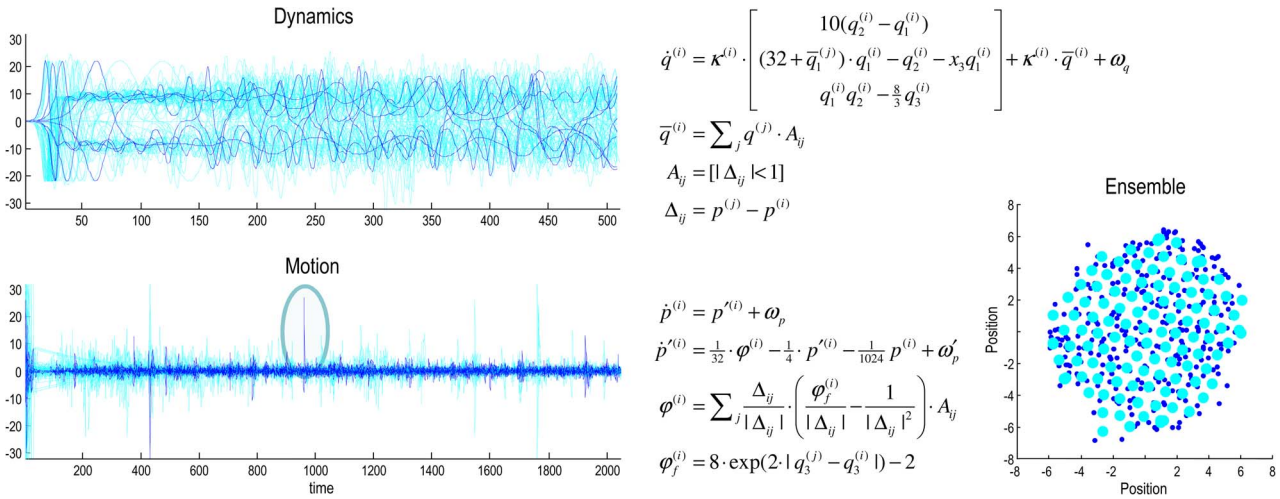

Fig. 4. ***Ensemble dynamics. The lower right panel shows the position of (128) subsystems comprising an ensemble after 2048 s, in terms of the dynamical status (three blue dots per subsystem) of each subsystem centered on its location (larger cyan dots). The left panels show the evolution of functional (upper panel) and structural (lower panel) states as a function of time. The (electrochemical) dynamics of the internal (blue) and external (cyan) states are shown for the 512 s. One can see initial (chaotic) transients that resolve fairly quickly, with itinerant behavior as they approach their attracting set. The lower panel shows the position of internal (blue) and external (cyan) subsystems over the entire simulation period and illustrates critical events (circled) that occur every few hundred seconds, especially at the beginning of the simulation. These events generally reflect a pair of particles (subsystems) being expelled from the ensemble to the periphery, when they become sufficiently close to engage short-range repulsive forces. These simulations integrated the stochastic differential equations using a forward Euler method with 1/512-s time steps and random fluctuations of unit variance; see [11] for details.***

systems with coherent (third) states were attracted to each other but repelled otherwise. The remaining two terms in the expression for acceleration (see Fig. 4) model viscosity that depends upon velocity and an exogenous force that attracts all locations to the origin, as if they were moving in a simple (quadratic) potential energy well. This ensures that the synthetic soup falls to the bottom of the well and enables local interactions. Note that the ensemble is dissipative at two levels: first, the classical motion includes dissipative friction or viscosity; and second, the functional dynamics are dissipative in the sense that they are not divergence free.

In the examples presented here, 128 subsystems were integrated using Euler's (forward) method with step sizes of 1/512 s and initial conditions sampled from the normal distribution. Random fluctuations were sampled from the unit normal distribution. By changing the parameters in the equations of motion, one can produce a repertoire of interesting behaviors. For most values of the parameters, ergodic behavior emerges as the ensemble approaches its random global attractor (usually after about 1000 s): generally, subsystems repel each other initially and then fall back toward the center, finding each other as they coalesce. Local interactions then mediate a reorganization, in which subsystems are passed around (sometimes to the periphery) until neighbors jostle gently with each other. In terms of the dynamics, transient synchronization can be seen as waves of dynamical bursting (due to the nonlinear coupling). In brief, the motion and electrochemical dynamics look very much like a restless soup (not unlike solar flares on the surface of the sun)—but does it have any self-organization beyond this?

B. The Markov Blanket

Because the structural and functional dependencies share the same adjacency matrix, it can be used to identify the principal Markov blanket by appealing to spectral graph theory: the Markov blanket of any subset of states encoded by a binary vector with elements $\chi_i \in \{0,1\}$ is given by Iverson bracket $[B \cdot \chi] \in \{0,1\}$, where the Markov blanket matrix $B = A + A^T + A^T A$ encodes children, parents, and parents of children. The principal eigenvector of the (symmetric) Markov blanket matrix will, by the Perron–Frobenius theorem, contain positive values. These values reflect the degree to which each state belongs to the cluster that is most interconnected (cf., spectral clustering). In what follows, the internal states were defined as belonging to subsystems with the $k = 8$ largest values. Having defined the internal states, the Markov blanket can be recovered from the Markov blanket matrix using $[B \cdot \chi]$ and divided into sensory and active states, depending upon whether they are influenced by the hidden states.

Given the internal states and their Markov blanket, we can now follow their assembly and visualize any structural or functional characteristics. Fig. 5 shows the adjacency matrix used to identify the Markov blanket. This adjacency matrix had nonzero entries if two subsystems were coupled over the penultimate 256 s of a 2048-s simulation. In other

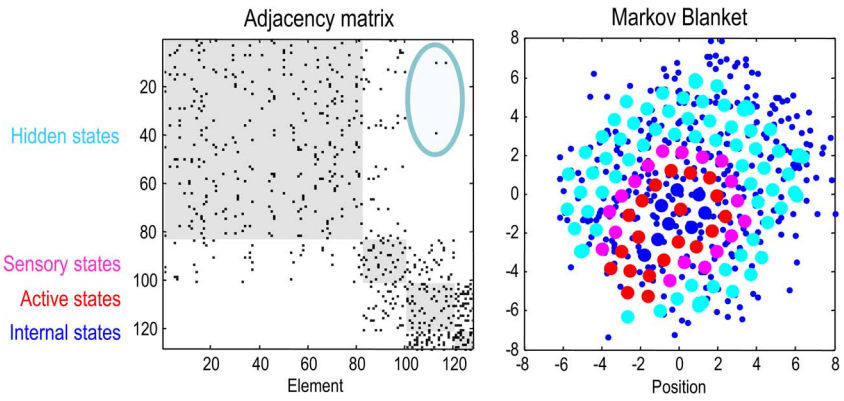

Fig. 5. ***Emergence of the Markov blanket. The left panel shows the adjacency matrix that indicates a conditional dependency (spatial proximity) on at least one occasion over the last 256 seconds of the simulation. The adjacency matrix has been reordered to show the partition of hidden (cyan), sensory (magenta), active (red), and internal (blue) subsystems, whose positions are shown in the right-hand-side panel, using the same format as in the previous figure. Note the absence of direct connections (edges) between external or hidden and internal subsystem states. The circled area illustrates coupling between active and hidden states that are not reciprocated. The spatial self-organization in the upper left panel is self evident; internal states have arranged themselves in a small loop structure with a little cilium, protected by the active states that support the surface or sensory states. When viewed as a movie, the entire ensemble pulsates in a chaotic but structured fashion, with the most marked motion in the periphery.***

words, it accommodates the fact that the adjacency matrix is itself an ergodic process, because it is defined by the ergodic flow of states. The right-hand-side panel shows the location of subsystems with internal states (blue) and their Markov blanket, in terms of sensory (magenta) and active (red) locations. A clear structure can be seen here, where the internal subsystems are (unsurprisingly) close together and enshrouded by the Markov blanket. Interestingly, the active subsystems support the sensory subsystems that are exposed to hidden states. This is reminiscent of a biological cell with a cytoskeleton that supports some sensory epithelia or receptors within its membrane.

C. Active Inference

If the internal states encode a probability density over the hidden or external states, then it should be possible to predict external states from internal states. In other words, if internal events represent external events, they should exhibit a significant statistical dependency. To establish this dependency, we examined the functional (electrochemical) status of internal subsystems to see if they could predict structural events (movement) in the external milieu. This is not unlike the approach taken in brain mapping that searches for statistical dependencies between, say, motion in the visual field and neuronal activity [66].

To test for statistical dependencies, the principal patterns of activity among the internal (functional) states were summarized using singular value decomposition and temporal embedding (see Fig. 6). A classical canonical variate analysis was then used to assess the significance of a simple linear mapping between expression of these patterns and the movement of each external subsystem. The upper left panel of Fig. 6 illustrates these internal dynamics, while the lower left panel shows the Newtonian motion of the external subsystem that was best predicted. The agreement between the actual (dotted line) and predicted (solid line) motion is self-evident, particularly around the negative excursion at 300 s. The internal dynamics that predict this event appear to emerge in their fluctuations before the event itself (see Fig. 6), as would be anticipated if internal events were modeling external events. Interestingly, the subsystem best predicted was the furthest away from the internal states (magenta circle in the lower right panel). This probably reflects the fact that peripheral subsystems have the greatest latitude for movement and show the largest excursions.

This example illustrates how internal states infer or register distant events in a way that is not dissimilar to the perception of auditory events through sound waves. The lower right panel suggests that motion predictions are the most significant at the periphery of the ensemble, where the ensemble has the greatest latitude for movement. These movements are coupled to the internal states, via the Markov blanket, through generalized synchrony. Generalized synchrony refers to the synchronization of chaotic dynamics, usually in skew–product (master–slave) systems [67], [68]. However, in our setup, there is no master–slave relationship but a circular causality induced by the Markov blanket. Generalized synchrony was famously observed by Huygens in his studies of pendulum clocks that synchronized themselves through the imperceptible motion of beams from which they were suspended [69]. This nicely illustrates the "action at a distance" caused by chaotically synchronized waves of motion. Circular causality begs the question of whether internal states predict external causes of their sensory states or actively cause them through action. Exactly

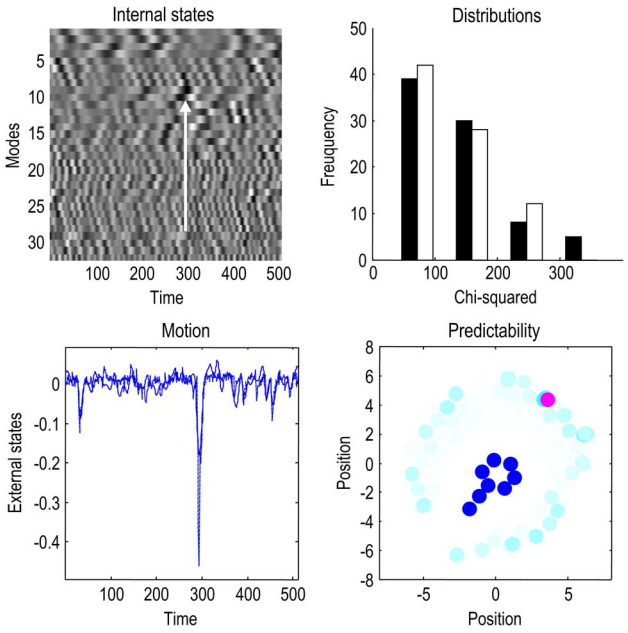

Fig. 6. *Self-organized perception. This figure illustrates the Bayesian perspective on self-organized dynamics. The upper left panel shows the first (principal) 32 eigenvariates of the internal (functional) states as a function of time over the last 512 s of the simulations reported in Figs. 4 and 5. These eigenvariates were obtained by a singular value decomposition of the time-series over all internal functional states (lagged between plus and minus 16 s) and were used to predict the (2-D) motion of each external subsystem. The significance of this prediction was assessed using the Wilks lambda (following a standard transformation to the chi-squared statistic). The actual (dotted line) and predicted (solid line) position for the most significant external subsystem is shown on the lower left. The fluctuations in internal states are visible in the upper panel and provide a linear mixture that correlates with the external fluctuation (highlighted with a white arrow). The location of the external subsystem that was best predicted is shown by the magenta circle on the lower right. The lower right panel also shows the significance with which the motion of the remaining external states could be predicted (with the intensity of the cyan being proportional to the chi-squared statistic above). Interestingly, the motion that is predicted with the greatest significance is restricted to the periphery of the ensemble, where the external subsystems have the greatest latitude for movement. To ensure that this inferential coupling was not a chance phenomenon, we repeated the analysis after flipping the external states in time. This destroys any statistical coupling between the internal and external states but preserves the correlation structure of fluctuations within either subset. The distribution of the ensuing chi-squared statistics (over 82 external elements) is shown in the upper right panel for the true (black) and null (white) analyses. Crucially, five of the subsystems in the true analysis exceeded the largest statistic in the null analysis. The largest value of the null distribution provides protection against false positives at a level of 1/82. The probability of obtaining five chi-squared values above this threshold by chance is vanishingly small* $p = 0.00052$.

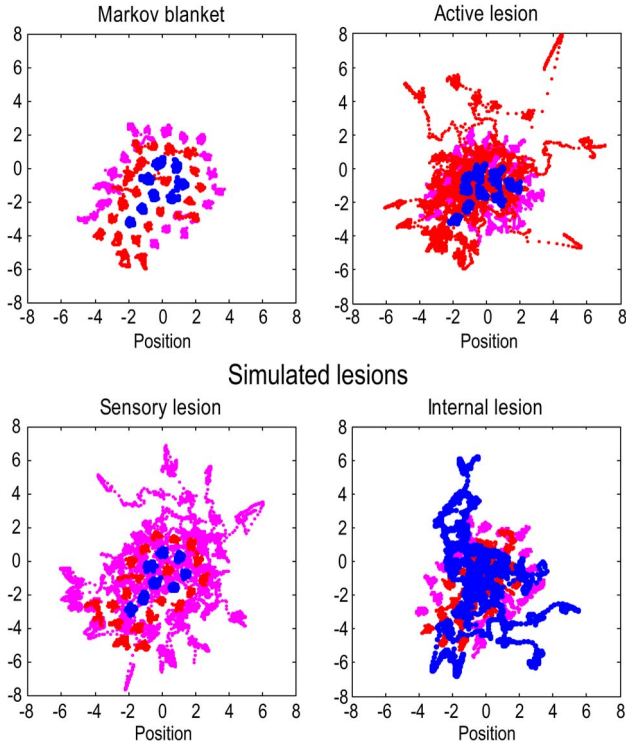

Fig. 7. *Autopoiesis and oscillator death. These results show the trajectory of the subsystems for 512 s after the last time point characterized in Fig. 4. The upper left panel shows the trajectories under the normal state of affairs; showing a preserved and quasi crystalline arrangement of the internal states (blue) and the Markov blanket (active states in red and sensory states in magenta). Contrast this self-organized maintenance of form with the decay and dispersion that ensues when the internal states and Markov blankets are synthetically lesioned (remaining three panels). In all simulations, a subset of states was lesioned by precluding influences on the functional states of neighboring subsystems. The upper right panel shows the effect of this relatively subtle lesion on active states that are rapidly expelled from the interior of the ensemble, allowing sensory states to invade and disrupt the internal states. A similar phenomenon is seen when the sensory states were lesioned, as they drift out into the external system (lower left panel). There is a catastrophic loss of structural integrity when the internal states themselves cannot affect each other, with a rapid migration of internal states through and beyond their Markov blanket (lower right panel).*

the same sorts of questions apply to perception [70], [71]: for example, are visually evoked neuronal responses caused by external events or by our movements? We will return to this question in Section VI.

D. Autopoiesis and Oscillator Death

To test for autopoietic maintenance of structural and functional integrity, the sensory, active, and internal subsystems were selectively lesioned by rendering them functionally closed; in other words, by preventing them from influencing their neighbors. Again, this recapitulates a common approach in neuroscience; namely, neuropsychology, where the functional consequences of a lesion are examined. Our lesion was relatively mild, in the sense that lesioned subsystems retained their dynamics and could respond to neighboring elements.

Fig. 7 illustrates the effects of these interventions by following the evolution of the internal states and their

Markov blanket over 512 s. The upper left panel shows the conservation of structural (and implicitly functional) integrity in terms of spatial configuration over this time period in the absence of a lesion. Contrast this with the remaining three panels that show structural disintegration as the integrity of the Markov blanket is lost and internal elements are extruded into the environment.

In summary, this section provides proof of principle that an arbitrary random dynamical system can, when suitably inspected, show evidence of active inference in the sense that there are statistical dependencies between internal and hidden states. Crucially, the emergent dynamics are consistent with active inference, in which internal states couple back to hidden external states to (apparently) preserve the attracting set. In Section VI, we consider the other extreme of self-organization and look at simulations of action and action observation in the brain.

VI. ACTIVE INFERENCE AND COGNITIVE DYNAMICS

In Section V, we started with equations of motion for generalized states and then identified the external, internal states and their Markov blanket based on the ensuing dynamics. In our final illustration of active inference, we start with a partition and solve for internal and active states using a generalized descent on free energy under the Laplace assumption [see (10) and (18)]

$$\begin{aligned}\dot{\tilde{\psi}} &= f_{\psi}(\tilde{\psi},\tilde{s},\tilde{a}) + \tilde{\omega}_{\psi}\\ \dot{\tilde{s}} &= f_{s}(\tilde{\psi},\tilde{s},\tilde{a}) + \tilde{\omega}_{s}\\ \dot{\tilde{\mu}} &= D\tilde{\mu} - \Gamma\cdot\partial_{\tilde{\mu}}F(\tilde{s},\tilde{a},\tilde{\mu})\\ \dot{\tilde{a}} &= D\tilde{a} - \Gamma\cdot\partial_{\tilde{a}}F(\tilde{s},\tilde{a},\tilde{\mu}).\end{aligned} \tag{25}$$

This solution requires us to specify the Lagrangian or generative model that defines free energy in terms of its equations of motion. However, we specified these equations with a twist: the generative model used for internal states is more complicated than the Lagrangian used for the remaining states. This has the important consequence that the internal states enslave external states, through action, to effectively create the sensorium expected under the generative model. In other words, we simulate willful action that is generated by prior beliefs about hidden causes, which do not actually exist (until they are emulated by action). In what follows, we simulate handwriting using the simulations originally reported in [72]. Having illustrated how inference can cause behavior, we then turn to perceptual inference, using the same simulations to model handwriting recognition or action observation. This illustrates an ambitious application of generalized filtering to invert a highly nonlinear state–space model with autonomous (heteroclinic) dynamics.

A. A Generative Model of Writing

Our agent was equipped with a simple generative model based on Lotka–Volterra dynamics. The particular form of this model has been discussed previously as the basis of putative speech decoding [73]. Here, it is used to model a heteroclinic cycle encoding successive locations to which the agent expects its two-jointed arm to be attracted. The resulting trajectory was contrived to simulate handwriting. This model comprises two sets of hidden states $\tilde{\psi} = (\tilde{u}, \tilde{q})$. The first $(\tilde{u}_1, \ldots, \tilde{u}_6)$ occupies an abstract state space, in which a series of (unstable) attracting points are visited in succession. More formally, the (Lotka–Volterra) equations of motion for these hidden attractor states ensure that only one has a high value at any one time, and impose a particular sequence on the underlying states. These equations of motion are provided in Fig. 8 (lower left). The second set of hidden states $(\tilde{q}_1, \tilde{q}_2)$ describes the (angular) positions and velocities of arms joints. The attractor and angular states are coupled through a prior expectation that the arm will be drawn to a particular location $v^*(\tilde{u})$ specified by the attractor states. This is implemented by placing a (fictive) elastic band between the tip of the arm and the location that exerts a force on the arm ($\phi(\tilde{q},\tilde{u})$ in Fig. 8). The hidden states draw the arm to a succession of points to produce a trajectory. We chose the locations (L in Fig. 8) so that the resulting trajectory looked like handwriting. Crucially, hidden states generate both proprioceptive and visual predictions. The proprioceptive consequences are the angular positions and velocities of the two joints $\tilde{q}$, while visual information predicts the location of the arm $v(\tilde{q})$; see Fig. 8 and [72]. Crucially, because this generative model generates two (proprioceptive and exteroceptive) sensory modalities, its inversion corresponds to Bayes-optimal multisensory integration.

However, because action is also trying to reduce prediction errors, it will move the arm to minimize proprioceptive prediction errors and reproduce the expected trajectory. In other words, the arm will trace out a trajectory prescribed by prior beliefs about its itinerant motion. This closes the loop, producing autonomous self-generated sequences of behavior of the sort described below. Note that the real world does not contain any attracting locations or elastic bands. The only causes of observed movement are the self-fulfilling expectations encoded by the itinerant dynamics of the generative model. In summary, hidden attractor states entail the intended movement trajectory, because they generate predictions that action fulfils. Operationally, this is reflected in the equations of motion used by action in Fig. 8 (lower right panels): these are functions of action but not attractor states. Conversely, the equations of motion in the generative model are functions of attractor states but not action, where the attractor states generate forces that play the role of action.

A subtle but important constraint in these simulations was that action only had access to proprioceptive prediction error. In other words, action only minimized the difference between expected and sensed generalized

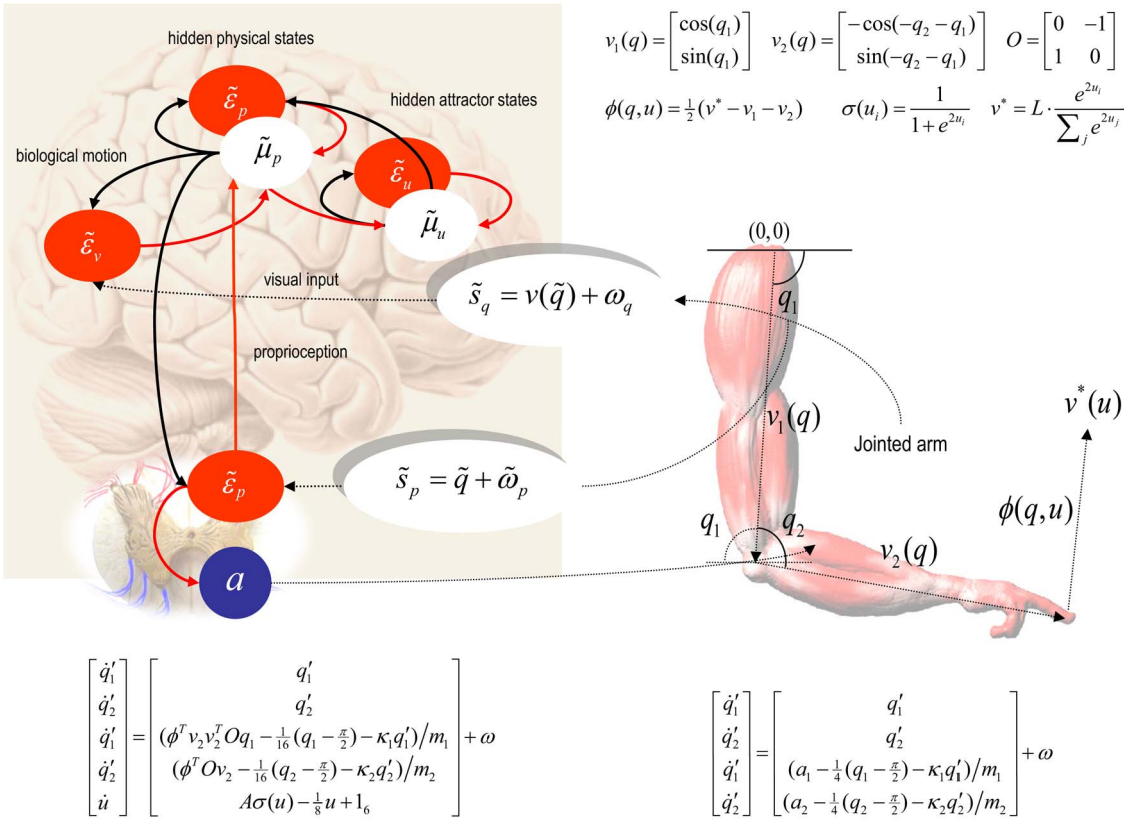

Fig. 8. ***Simulated mirror neuron system and motor plant. The right-hand-side panel depicts the functional architecture of supposed neural circuits underlying active inference. The red ellipses represent prediction error units (neurons or populations), while the white ellipses denote posterior expectations about hidden states in the world. Here, they are divided into abstract attractor states and physical states of the arm (angular positions and velocities of the two joints). Red arrows are forward connections conveying prediction errors and black arrows are backward connections mediating predictions. Motor commands are emitted by the blue units in the ventral horn of the spinal cord. Note that these just receive prediction errors about proprioceptive states. These prediction errors are the difference between sensed proprioceptive input from the two joints and descending predictions from posterior expectations in motor cortex. The two-jointed arm has a state space that is characterized by two angles, which control its position. The equations of motion on the left define the generative model or Lagrangian used for a generalized descent on free energy for the internal states (posterior expectations). The simpler equations on the right are used to define the free energy for action, and the flow of hidden states producing sensations; see [72] for details.***

motion of the joints, where the mapping between action (changing the generalized joint position) and proprioceptive input is very simple. However, this does not mean that visual information (prediction errors) cannot affect action. Visual information is crucial when optimizing posterior expectations that provide predictions in both proprioceptive and visual modalities. This means that visual input can influence action vicariously, through high level (intentional) representations that predict a (unimodal) proprioceptive component. In short, although the perception or intention of the agent integrates proprioceptive and visual information in a Bayes-optimal fashion, action is driven just by proprioceptive prediction errors. This will become important below, where we remove proprioceptive input but retain visual stimulation to simulate action observation.

Fig. 9 shows the results of integrating (25) using the generative model in Fig. 8. The top right panel shows the expected hidden states embodying Lotka–Volterra dynamics (the hidden joint states are smaller in amplitude). These generate predictions about the position of the joints (upper left panel) and consequent prediction errors that drive action. Action is shown on the lower right and displays intermittent forces to produce a motor trajectory. This trajectory is shown on the lower left and is translated with time to reproduce handwriting. Although this is a pleasingly simple way of simulating a complicated motor trajectory, it should be noted that this agent has a very limited repertoire of behaviors; it can only reproduce this sequence of graphemes, and will do so *ad infinitum*.

In summary, we have illustrated the functional architecture of a generative model whose autonomous (itinerant) expectations prescribe complicated motor sequences through active inference. This rests upon itinerant dynamics that can be regarded as a formal prior on hidden (and fictive) causes in the world. Action tries to

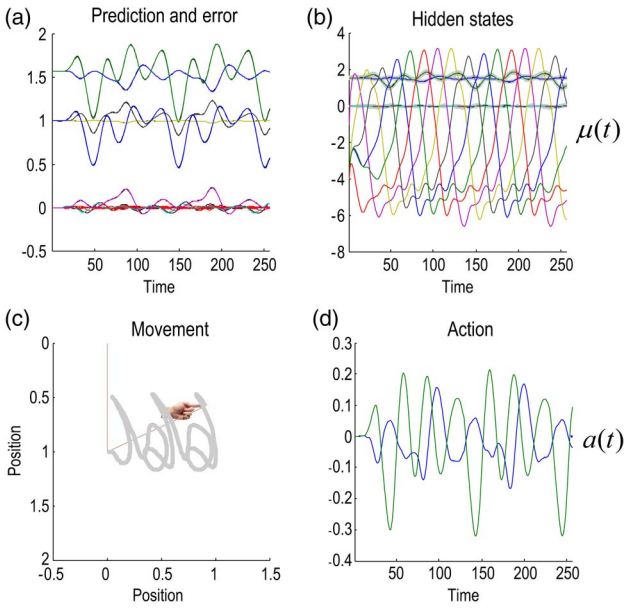

Fig. 9. *Simulating action. This figure shows the results of simulated action (writing) in terms of posterior expectations about hidden states of the world (b), consequent predictions about sensory input (a), and the ensuing behavior (c) that is caused by action (d). The autonomous dynamics that underlie this behavior rest upon the expected hidden states that follow Lotka–Volterra dynamics: these are the six (arbitrarily) colored lines in panel B. The hidden physical states have smaller amplitudes and map directly on to the predicted proprioceptive and visual signals (a). The visual locations of the two joints are shown as blue and green lines, above the predicted joint positions and angular velocities that fluctuate around zero. The dotted red lines correspond to prediction error, which shows small fluctuations about the prediction. Action tries to suppress this error by matching expected changes in angular velocity through exerting forces on the joints. These forces are shown in blue and green in panel d. The subsequent movement of the arm is shown in panel c. This trajectory has been plotted in a moving frame of reference so that it looks like synthetic handwriting. The straight lines in panel c denote the final position of the two jointed arm, and the hand icon shows the final position of its extremity.*

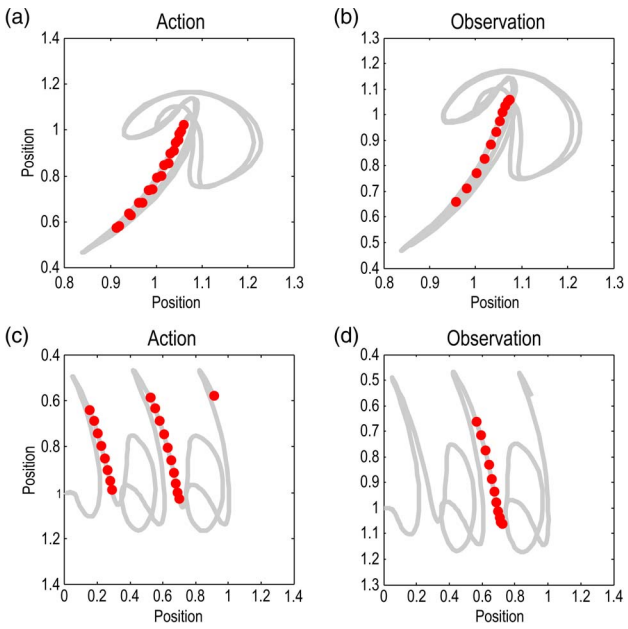

Fig. 10. *Simulating action observation. These results illustrate the perceptual correlates of units representing expected hidden states. The left-hand-side panels (a) and (c) show the activity of one (the fourth attractor) hidden state expectation under action, while the right-hand-side panels (b) and (d) show exactly the same responses under action–observation. The top rows (a) and (b) show the trajectory in visual space in terms of horizontal and vertical position (gray lines). The red dots correspond to the time bins during which the expectation exceeded two arbitrary units. The key thing to take from these results is that the simulated neuronal responses are specific to a limited part of visual space and, crucially, a particular trajectory through this space. Notice that the same selectivity is shown under action and observation. The implicit direction selectivity can be seen more clearly in the lower panels (c) and (d), in which the same data are displayed but in a moving frame of reference. The key thing to note here is that this unit responds preferentially when, and only when, the motor trajectory produces a downstroke, but not an upstroke.*

fulfill predictions about proprioceptive inputs, producing realistic behavior. These trajectories are both caused by representations of abstract (attractor) states and cause those states in the sense that they are expectations. Closing the loop in this way ensures a synchrony between internal expectations and external outcomes that recapitulates the circular causality illustrated in Section V. In Section VI-B, we will make a simple change that means that movements are no longer caused by the agent. However, we will see that the internal expectations are relatively unaffected, which means that they still anticipate observed movements.

B. Action–Observation

We now revisit the above simulations with a small but important change. Basically, we reproduced the same visual input but removed the proprioceptive consequences of action by downweighting proprioceptive precision. From the agent's perspective, this is like seeing an arm that looks like its own arm but does not generate sensations (i.e., the arm of another agent). However, the agent still expects the arm to move with a particular itinerant structure and will try to predict the trajectory with its generative model. In this instance, the hidden states still represent itinerant dynamics (intentions) that govern the motor trajectory, but these states do not produce (precise) proprioceptive prediction errors and, therefore, do not result in action [see (23)].

It is interesting to regard the ensuing dynamics of expected attractor states as representing trajectories through representational spaces; cf., the activity of place cells [74]. Fig. 10 illustrates the sensory or perceptual correlates of expected attractor states. The left-hand-side panels show the activity of one internal state (the fourth) under action, while the right-hand-side panels show exactly the same activity under action–observation. The

Table 2 Processes and Paradigms That Have Been Modeled Using the Scheme in This Paper

Domain	Process or paradigm
Perception	• Perceptual categorization (bird songs) [80] • Novelty and omission-related responses [80] • Perceptual inference (speech) [73] • Illusions [81] [82]
Sensory attenuation	• Attenuation and the force matching illusion [82]
Sensory learning	• Perceptual learning (mismatch negativity) [83]
Attention	• Attention and the Posner paradigm [60] • Attention and biased competition [60]
Motor control	• Retinal stabilization and oculomotor reflexes [84] • Saccadic eye movements and cued reaching [84] • Motor trajectories and place cells [72]
Sensorimotor integration	• Bayes-optimal sensorimotor integration [84]
Behavior	• Heuristics and dynamical systems theory [16] • Goal-directed behavior [85]
Action observation	• Action observation and mirror neurons [72]

top rows show the trajectories in visual space, in terms of horizontal and vertical displacements (gray lines). The red dots correspond to the time bins in which the posterior expectation exceeded a threshold of two arbitrary units. The key thing to take from these results is that the responses of this internal state are very specific to a limited part of space and, crucially, a particular trajectory through this space during both action and its observation.

The obvious analogy here would be mirror neuron activity [75] and its role in predicting the sensorial consequences of action [76]. However, there is also an interesting analogy between directionally selective place cells of the sort studied in hippocampal recordings, e.g., [74] and [77]. Notice that the same place and directional selectivity is seen under action and observation (Fig. 10 right and left columns). The direction selectivity can be seen more clearly in the lower panels, in which the same data are displayed but in a moving frame of reference (to simulate writing). The key thing to note is that this unit responds preferentially when, and only when, the motor trajectory produces a downstroke, but not an upstroke. These sorts of simulations beg the interesting question: would place cell activity be elicited during visual replay of movement through an environment? In principle, these questions can now be addressed using virtual navigation paradigms [78].

In summary, these simulations suggest that exactly the same neuronal representation can serve as a prescription for self-generated action, while, in another context, it encodes a perceptual representation of the intentions of another [75]. The only thing that changes is the context in which the inference is made. In these simulations, this contextual change was modeled by simply reducing the precision of proprioceptive errors. We have previously discussed this modulation of proprioceptive precision in terms of selectively enabling or disabling particular motor trajectories, which may be a potential target for the pathophysiology of Parkinson's disease. The connection with formal mechanisms of attentional gain [60] is interesting here, because it means that one could regard this contextual manipulation as an attentional bias to exteroceptive signals (caused by others) relative to interoceptive signals (caused by oneself).

VII. CONCLUSION

In conclusion, starting with some basic considerations about the ergodic behavior of random dynamical systems, we have seen how inference could be construed as an emergent property of any weakly mixing (random dynamical) system, and how it can be described in terms of a generalized descent on variational free energy. Using the same formalism, we have been able to address some fairly abstract issues in action and its observation that even touch on representation of intentions, agency, and behavior.

There are many other examples of cognitive dynamics that we could have considered using this scheme; the interested reader will find references in Table 2.

The take-home message of this work is that cognitive dynamics may conform to the same basic principles that underlie self-organization in any system with coupled dynamics. The emergence of cognitive-like phenomena rests upon the notion of a Markov blanket that separates internal states from external states. The very presence of this separation implies a generalized synchrony between external (e.g., environmental) and internal (e.g., neuronal) states that will appear to be lawful, in the sense that internal states minimize the same free-energy functional use for Bayesian inference. This lends a quintessentially inferential or predictive aspect to internal states that has many of the hallmarks of cognition. Crucially, this inference or assimilation is active, in the sense that the internal states affect the causes of sensory input vicariously, through action. The resulting circular causality between perception and action fits comfortably with many formulations in embodied cognition and artificial intelligence; for example, the perception–action cycle [25], active vision [71], the use of predictive information [26]–[28], and homeokinetic formulations [29]. Furthermore, it connects these perspectives to more general treatments of circular causality and autopoiesis in cybernetics and synergetics [3], [4].

One might argue that the formulation of cognitive dynamics offered in this paper is too general. In other words, if every dynamical system engages in some form of active inference, what is special about the cognitively adept brain? The answer may lie in the persistence of the brain's Markov blanket (i.e., the body) and the hierarchical depth of its implicit generative model. In nonlinear and dynamical generative models, there is intimate relationship between temporal scale and hierarchical level, in that higher levels generate slower dynamics that contextualize faster dynamics at lower levels. One might, therefore, imagine that (neuronal) systems that maintain the integrity of the Markov blanket for long periods of time may, necessarily, entertain deep inferences about the causal structure of their sensorium—influences that may be cognitive in nature. Perhaps the last word on active inference should go to Helmholtz [79], upon whose ideas much of this paper is based:

> "Each movement we make by which we alter the appearance of objects should be thought of as an experiment designed to test whether we have understood correctly the invariant relations of the phenomena before us, that is, their existence in definite spatial relations"—Herman von Helmholtz [79, p. 384]. ■

Acknowledgment

The authors would like to thank the anonymous reviewers for helpful guidance in communicating these ideas.

REFERENCES

[1] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *J. Physiol.*, vol. 100, no. 1–3, pp. 70–87, 2006.

[2] W. R. Ashby, "Principles of the self-organizing dynamic system," *J. Gen. Psychol.*, vol. 37, pp. 125–128, 1947.

[3] H. Haken, *Synergetics: An Introduction. Non-Equilibrium Phase Transition and Self-Organisation in Physics, Chemistry and Biology*, 3rd ed. Berlin, Germany: Springer-Verlag, 1983.

[4] H. R. Maturana and F. Varela, "Autopoiesis: The organization of the living," in *Autopoiesis and Cognition*, V. F. Maturana HR, Ed. Dordrecht, The Netherlands: Reidel, 1980.

[5] G. Nicolis and I. Prigogine, *Self-Organization in Non-Equilibrium Systems*. New York, NY, USA: Wiley, 1977.

[6] E. Schrödinger, *What is life?: The Physical Aspect of the Living Cell*. Dublin, Dublin: Trinity College, 1944.

[7] K. Friston, "A free energy principle for biological systems," *Entropy*, vol. 14, pp. 2100–2121, 2012.

[8] G. Auletta, "A paradigm shift in biology?" *Information*, vol. 1, pp. 28–59, 2010.

[9] G. Auletta, "Information and metabolism in bacterial chemotaxis," *Entropy*, vol. 15, no. 1, pp. 311–326, 2013.

[10] B. Sengupta, M. Stemmler, S. B. Laughlin, and J. E. Niven, "Action potential energy efficiency varies among neuron types in vertebrates and invertebrates," *PLoS Comput. Biol.*, vol. 6, 2010, e1000840.

[11] K. J. Friston, "Life as we know it," *Roy. Soc. Interface*, vol. 10, no. 86, 2013, DOI: 10.1098/rsif.2013.0475

[12] P. Ao, "Global view of bionetwork dynamics: Adaptive landscape," *J. Gen. Genom.*, vol. 36, no. 2, pp. 63–73, 2009.

[13] L. Demetrius, "Thermodynamics and evolution," *J. Theor. Biol.*, vol. 206, no. 1, pp. 1–16, 2000.

[14] M. J. Davis, "Low-dimensional manifolds in reaction-diffusion equations. 1. Fundamental aspects.," *J. Phys. Chem. A*, vol. 110, no. 16, pp. 5235–5256, 2006.

[15] M. I. Rabinovich, V. S. Afraimovich, V. Bick, and P. Varona, "Information flow dynamics in the brain," *Phys. Life Rev.*, vol. 9, no. 1, pp. 51–73, 2012.

[16] K. Friston and P. Ao, "Free-energy, value and attractors," *Comput. Math. Methods Med.*, vol. 2012, 2012, 937860.

[17] R. C. Conant and R. W. Ashby, "Every good regulator of a system must be a model of that system," *Int. J. Syst. Sci.*, vol. 1, no. 2, pp. 89–97, 1970.

[18] D. J. Evans, "A non-equilibrium free energy theorem for deterministic systems," *Mol. Phys.*, vol. 101, pp. 15551–15554, 2003.

[19] D. J. Evans and D. J. Searles, "Equilibrium microstates which generate second law violating steady states," *Phys. Rev. E*, vol. 50, no. 2, pp. 1645–1648, 1994.

[20] P. Dayan, G. E. Hinton, and R. M. Neal, "The Helmholtz machine," *Neural Comput.*, vol. 7, pp. 889–904, 1995.

[21] R. L. Gregory, "Perceptions as hypotheses," *Phil. Trans. Roy. Soc. Lond. B*, vol. 290, pp. 181–197, 1980.

[22] H. Helmholtz, "Concerning the perceptions in general," in *Treatise on Physiological Optics*, vol. III, 3rd ed. New York, NY, USA: Dover, 1866/1962.

[23] D. Kersten, P. Mamassian, and A. Yuille, "Object perception as Bayesian inference," *Annu. Rev. Psychol.*, vol. 55, pp. 271–304, 2004.

[24] T. S. Lee and D. Mumford, "Hierarchical Bayesian inference in the visual cortex," *J. Opt. Soc. Amer., Opt. Image Sci. Vis.*, vol. 20, pp. 1434–1448, 2003.

[25] J. M. Fuster, "Upper processing stages of the perception-action cycle," *Trends Cogn. Sci.*, vol. 8, no. 4, pp. 143–145, 2004.

[26] N. Ay, N. Bertschinger, R. Der, F. Güttler, and E. Olbrich, "Predictive information and

explorative behavior of autonomous robots," *Eur. Phys. J. B*, vol. 63, pp. 329–3239, 2008.

[27] W. Bialek, I. Nemenman, and N. Tishby, "Predictability, complexity, learning," *Neural Comput.*, vol. 13, no. 11, pp. 2409–2463, 2001.

[28] N. Tishby and D. Polani, "Information theory of decisions and actions," in *Perception-Reason-Action Cycle: Models, Algorithms and Systems*, V. Cutsuridis, A. Hussain, and J. Taylor, Eds. Berlin, Germany: Springer-Verlag, 2010.

[29] H. Soodak and A. Iberall, "Homeokinetics: A physical science for complex systems," *Science*, vol. 201, pp. 579–582, 1978.

[30] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, Feb. 2010.

[31] H. Crauel and F. Flandoli, "Attractors for random dynamical systems," *Probab. Theory Relat. Fields*, vol. 100, pp. 365–393, 1994.

[32] H. Crauel, "Global random attractors are uniquely determined by attracting deterministic compact sets," *Annali di Matematica Pura ed Applicata*, vol. 4, no. 176, pp. 57–72, 1999.

[33] T. D. Frank, *Nonlinear Fokker-Planck Equations: Fundamentals and Applications. Springer Series in Synergetics*. Berlin, Germany: Springer-Verlag, 2004.

[34] P. Ao, "Potential in stochastic differential equations: Novel construction," *J. Phys. A, Math Gen.*, vol. 37, pp. L25–L30, 2004.

[35] R. Yuan, Y. Ma, B. Yuan, and A. Ping *Constructive proof of global Lyapunov function as potential function*, Dec. 2010. [Online]. Available: arXiv:1012.2721v1

[36] G. Auletta, M. Fortunato, and G. Parisi, *Quantum Mechanics,* 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[37] J. Pearl, *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference*. San Fransisco, CA, USA: Morgan Kaufmann, 1988.

[38] G. D. Birkhoff, "Proof of the ergodic theorem," *Proc. Nat. Acad. Sci. USA*, vol. 17, pp. 656–660, 1931.

[39] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit., Univ. College London, London, U.K., 2003.

[40] R. P. Feynman, *Statistical Mechanics*. Reading, MA, USA: Benjamin, 1972.

[41] G. E. Hinton and D. van Camp, "Keeping the neural networks simple by minimizing the description length of the weights," *Proc. 6th Annu. Conf. Comput. Learn. Theory*, pp. 5–13, 1993.

[42] R. E. Kass and D. Steffey, "Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models)," *J. Amer. Stat. Assoc.*, vol. 407, pp. 717–726, 1989.

[43] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[44] C. van Leeuwen, "Perceptual-learning systems as conservative structures: Is economy an attractor?" *Psychol. Res.*, vol. 52, no. 2–3, pp. 145–152, 1990.

[45] V. Pasquale, P. Massobrio, L. L. Bologna, M. Chiappalone, and S. Martinoia, "Self-organization and neuronal avalanches in networks of dissociated cortical neurons," *Neuroscience*, vol. 153, no. 4, pp. 1354–1369, 2008.

[46] H. Barlow, "Possible principles underlying the transformations of sensory messages," in *Sensory Communication*, W. Rosenblith, Ed. Cambridge, MA, USA: MIT Press, 1961, pp. 217–234.

[47] R. Linsker, "Perceptual neural organization: Some approaches based on network models and information theory," *Annu. Rev. Neurosci.*, vol. 13, pp. 257–281, 1990.

[48] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev. Ser. II*, vol. 106, no. 4, pp. 620–630, 1957.

[49] M. Breakspear and C. J. Stam, "Dynamics of a neural system with a multiscale architecture," *Philos. Trans. Roy. Soc. Lond. B, Biol. Sci.*, vol. 360, no. 1457, pp. 1051–1074, 2005.

[50] S. L. Bressler and E. Tognoli, "Operational principles of neurocognitive networks," *Int. J. Psychophysiol.*, vol. 60, no. 2, pp. 139–148, 2006.

[51] W. J. Freeman, "Characterization of state transitions in spatially distributed, chaotic, nonlinear, dynamical systems in cerebral cortex," *Integr. Physiol. Behav. Sci.*, vol. 29, no. 3, pp. 294–306, 1994.

[52] G. Auletta, *Cognitive Biology: Dealing With Information From Bacteria to Minds*. Oxford, U.K.: Oxford Univ. Press, 2011.

[53] K. Friston, K. Stephan, B. Li, and J. Daunizeau, "Generalised filtering," *Math. Probl. Eng.*, vol. 2010, 2010, 621670.

[54] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny, "Variational free energy and the Laplace approximation," *Neuroimage*, vol. 34, no. 1, pp. 220–234, 2007.

[55] K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau, "DEM: A variational treatment of dynamic systems," *Neuroimage*, vol. 41, no. 3, pp. 849–885, 2008.

[56] K. Friston, "Hierarchical models in the brain," *PLoS Comput. Biol.*, vol. 4, no. 11, 2008, e1000211.

[57] P. Elias, "Predictive coding-I," *IRE Trans. Inf. Theory*, vol. 1, no. 1, pp. 16–24, Mar. 1955.

[58] A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston, "Canonical microcircuits for predictive coding," *Neuron*, vol. 76, no. 4, pp. 695–711, 2012.

[59] D. Mumford, "On the computational architecture of the neocortex. II," *Biol. Cybern.*, vol. 66, pp. 241–251, 1992.

[60] H. Feldman and K. J. Friston, "Attention, uncertainty, free-energy," *Front. Human Neurosci.*, vol. 4, 2010, DOI: 10.3389/fnhum.2010.00215.

[61] A. G. Feldman and M. F. Levin, "The origin and use of positional frames of reference in motor control," *Behav. Brain Sci.*, vol. 18, pp. 723–806, 1995.

[62] A. M. Turing, "The chemical basis of morphogenesis," *Phil. Trans. Roy. Soc. Lond. B*, vol. 237, no. 641, pp. 37–72, 1952.

[63] B. P. Belousov, "'Периодически действующая реакция и ее механизм', Periodically acting reaction and its mechanism," *Сборнирефератов по радиационной медицине, Collection Abstr. Rad. Med.*, vol. 147, 1959.

[64] A. Hu, Z. Xu, and L. Guo, "The existence of generalized synchronization of chaotic systems in complex networks," *Chaos*, vol. 20, no. 1, 2010, 013112.

[65] D. Poland, "Cooperative catalysis and chemical chaos: A chemical model for the Lorenz equations," *Physica D*, vol. 65, no. 1, pp. 86–99, 1993.

[66] S. Zeki, "The Ferrier lecture 1995 behind the seen: The functional specialization of the brain in space and time," *Philos. Trans. Roy. Soc. Lond. B., Biol. Sci.*, vol. 360, no. 1458, pp. 1145–1183, 2005.

[67] B. Hunt, E. Ott, and J. Yorke, "Differentiable synchronisation of chaos," *Phys. Rev. E*, vol. 55, pp. 4029–4034, 1997.

[68] E. Barreto, K. Josic, C. J. Morales, E. Sander, and P. So, "The geometry of chaos synchronization," *Chaos*, vol. 13, pp. 151–164, 2003.

[69] C. Huygens, *Horologium Oscillatorium*. France: Paris, 1673.

[70] R. A. Adams, S. Shipp, and K. J. Friston, "Predictions not commands: Active inference in the motor system," *Brain Struct. Funct.*, vol. 218, pp. 611–643, 2013.

[71] R. H. Wurtz, K. McAlonan, J. Cavanaugh, and R. A. Berman, "Thalamic pathways for active vision," *Trends Cogn. Sci.*, vol. 5, no. 4, pp. 177–184, 2011.

[72] K. Friston, J. Mattout, and J. Kilner, "Action understanding and active inference," *Biol. Cybern.*, vol. 104, pp. 137–160, 2011.

[73] S. J. Kiebel, J. Daunizeau, and K. J. Friston, "Perception and hierarchical dynamics," *Front. Neuroinf.*, vol. 3, 2009, DOI: 10.3389/neuro.11.020.2009.

[74] S. Leutgeb, J. K. Leutgeb, M. B. Moser, and E. I. Moser, "Place cells, spatial maps and the population code for memory," *Current Opinion Neurobiol.*, vol. 15, no. 6, pp. 738–746, 2005.

[75] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annu. Rev. Neurosci.*, vol. 27, pp. 169–192, 2004.

[76] R. C. Miall, "Connecting mirror neurons and forward models," *Neuroreport*, vol. 14, no. 17, pp. 2135–2137, 2003.

[77] D. Robbe and G. Buzsáki, "Alteration of theta timescale dynamics of hippocampal place cells by a cannabinoid is associated with memory impairment," *J. Neurosci.*, vol. 29, no. 40, pp. 12597–12605, Oct. 2009.

[78] C. D. Harvey, F. Collman, D. A. Dombeck, and D. W. Tank, "Intracellular dynamics of hippocampal place cells during virtual navigation," *Nature*, vol. 461, no. 7266, pp. 941–946, 2009.

[79] H. von Helmholtz, "The facts of perception (1878)," in *The Selected Writings of Hermann von Helmholtz*, R. Karl, Ed. Middletown, CT, USA: Wesleyan Univ. Press, 1971.

[80] K. Friston and S. Kiebel, "Cortical circuits for perceptual inference," *Neural Netw.*, vol. 22, no. 8, pp. 1093–1104, 2009.

[81] H. Brown and K. J. Friston, "Free-energy and illusions: The cornsweet effect," *Front. Psychol.*, vol. 3, 2012, DOI: 10.3389/fpsyg.2012.00043.

[82] H. Brown, R. A. Adams, I. Parees, M. Edwards, and K. Friston, "Active inference, sensory attenuation and illusions," *Cogn. Process.*, vol. 14, no. 4, pp. 411–427, 2013.

[83] K. J. Friston and S. J. Kiebel, "Predictive coding under the free-energy principle," *Philos. Trans. Roy. Soc. B*, vol. 364, pp. 1211–1221, May 2009.

[84] K. J. Friston, J. Daunizeau, J. Kilner, and S. J. Kiebel, "Action and behavior: A free-energy formulation," *Biol. Cybern.*, vol. 102, no. 3, pp. 227–260, 2010.

[85] K. J. Friston, J. Daunizeau, and S. J. Kiebel, "Active inference or reinforcement learning?" *PLoS One*, vol. 4, no. 7, 2009, e6421.

ABOUT THE AUTHORS

Karl Friston received undergraduate training in Natural Sciences from Gonville and Caius College, Cambridge University, U.K. and completed his medical training at Kings College, London University, in 1983.

He is a theoretical neuroscientist and authority on brain imaging. He invented statistical parametric mapping (SPM), voxel-based morphometry (VBM), and dynamic causal modeling (DCM). These contributions were motivated by schizophrenia research and theoretical studies of value learning—formulated as the dysconnection hypothesis of schizophrenia. His mathematical contributions include variational Laplacian procedures and generalized filtering for hierarchical Bayesian model inversion. He currently works on models of functional integration in the human brain and the principles that underlie neuronal interactions. His main contribution to theoretical neurobiology is a free-energy principle for action and perception (active inference).

Dr. Friston received the first Young Investigators Award in Human Brain Mapping (1996) and was elected a Fellow of the Academy of Medical Sciences (1999). In 2000, he was President of the international Organization of Human Brain Mapping. In 2003, he was awarded the Minerva Golden Brain Award and was elected a Fellow of the Royal Society in 2006. In 2008, he received a Medal, Collège de France and an Honorary Doctorate from the University of York in 2011. He became of Fellow of the Society of Biology in 2012 and received the Weldon Memorial prize and Medal in 2013 for contributions to mathematical biology.

Biswa Sengupta received the M.Sc. degree in theoretical computer science from the University of York, York, U.K. and the Ph.D. degree from Cambridge University, Cambridge, U.K., in 2010, working on the tradeoffs between information encoding and energy consumption in single neurons.

As an electronics and computer engineering undergraduate at the University of York, he was interested in machine intelligence, consequently leading to his work on optimizing hardware–software codesign of complicated control systems such as fighter planes, space shuttles, etc. Inspired by classical work in artificial intelligence, he began to consider computations in biological systems, especially within and between neurons. As a result of this interest, his undergraduate thesis focused on short-term synaptic plasticity in pyramidal neurons in the hippocampus. During summers, he worked on optimizing linear algebra libraries for statistical parametric modeling—a widely used tool in neuroimaging. During his time at the University of York, he primarily studied fear-conditioning circuits in the auditory pathway. He then joined the Max Planck Institute, Tübingen, Germany, for a second MS level course in neuroscience to supplement my understanding of the nervous system. At Tübingen, he had the opportunity to learn theoretical and experimental techniques on both rodents and nonhuman primates working in the intersection of neurophysiology, machine learning, optimization, and statistical physics. On completion of his Ph.D., he joined the Indian Institute of Science, Bangalore, India, as a Wellcome Trust Fellow.

Gennaro Auletta was born on August 27, 1957, in Naples, Italy. He received the Ph.D. degree in philosophy (''Leibniz's Philosophy of modalities'') from the Sapienza University of Rome, Rome, Italy, in 1993.

He is an Italian Philosopher of Science actively involved in scientific research. He is an internationally acknowledged expert in quantum mechanics and in the foundation and interpretation of this discipline. His main interests in quantum information led him to focus on the way in which biological and cognitive systems deal with information. He is also active in the field of the dialog between science, philosophy and theology, and has been the Vice Director of the International Conference on Biological Evolution: Facts and Theories, held at the Pontifical Gregorian University, Rome, Italy, in March 2009. He is an Aggregate Professor at the Pontifical Gregorian University where, from 2003 to 2012, he held the position of Scientific Director of the Specialization in Science and Philosophy. He is also a Senior Researcher at the University of Cassino, Cassino, Italy. From 2003 to 2010, he was the Scientific Coordinator of the ''Science, Theology and the Ontological Quest'' Project (STOQ, a project under the patronage of the Pontifical Council for Culture involving seven Roman Pontifical Universities and supported by the John Templeton Foundation).

Dr. Auletta has been a Fellow of the Linnean Society of London and of the International Society for Science and Religion since 2009.