ARTICLE **Communicated by Karl Friston**

Recognition Dynamics in the Brain under the Free Energy Principle

Chang Sub Kim
cskim@jnu.ac.kr
Department of Physics, Chonnam National University, Gwangju 61186, Republic of Korea

We formulate the computational processes of perception in the framework of the principle of least action by postulating the theoretical action as a time integral of the variational free energy in the neurosciences. The free energy principle is accordingly rephrased, on autopoetic grounds, as follows: all viable organisms attempt to minimize their sensory uncertainty about an unpredictable environment over a temporal horizon. By taking the variation of informational action, we derive neural recognition dynamics (RD), which by construction reduces to the Bayesian filtering of external states from noisy sensory inputs. Consequently, we effectively cast the gradient-descent scheme of minimizing the free energy into Hamiltonian mechanics by addressing only the positions and momenta of the organisms' representations of the causal environment. To demonstrate the utility of our theory, we show how the RD may be implemented in a neuronally based biophysical model at a single-cell level and subsequently in a coarse-grained, hierarchical architecture of the brain. We also present numerical solutions to the RD for a model brain and analyze the perceptual trajectories around attractors in neural state space.

1 Introduction

The quest for a universal principle that may explain the cognitive and behavioral operation of the brain is of great scientific interest. The apparent difficulty in this quest is the gap between information processing and the biophysics that governs neurophysiology in the brain. However, it is evident that the base material for brain functions comprises neurons obeying the laws of physics. Thus, any biological principles that attempt to explain the brain's large-scale functioning must be consistent with our accepted physical reality (Schrödinger, 1967). It appears that among the current approaches, the one that prevails is the classical, effective epistemology of regarding perceptions as the construction of hypotheses that may represent the truth by producing symbolic structures matching physical reality (von Helmholtz, 1962; Gregory, 1980; Dayan, Hinton, Neal, & Zemel, 1995).

Neural Computation **30, 2616–2659** (2018)
doi:10.1162/neco_a_01115

One influential candidate at present for such a rubric in neuroscience is the free energy principle (FEP; Friston, 2009, 2010, 2013). For a technical appraisal of the FEP, we refer to (Buckley, Kim, McGregor, & Seth, 2017) where the theoretical assumptions and mathematical structure involved in the FEP are reviewed in great detail. A recent study (Ramstead, Badcock, & Friston, 2017) suggested variational neuroethology, which integrates the FEP with evolutionary systems theory to explain how living systems persist as bounded, self-organizing systems over time. To state compactly, the FEP suggests that all viable organisms perceive and act on the external world by instantiating a probabilistic causal model embodied in their brain in a manner that ensures their adaptive fitness or autopoiesis (Maturana & Varela, 1980). The biological mechanism that endows the organism's brain with the operation is theoretically framed into an information-theoretic measure, which is termed *variational* or *informational free energy* (IFE). According to the FEP, a living system attempts to minimize sensory surprisal (i.e., self-information) when exposed to environmental perturbations by calling on active inference. However, the brain does not preside over in-streaming sensory distribution; accordingly, the brain cannot directly minimize the sensory surprisal; instead, it minimizes its upper bound, which is the IFE. This is the same quantity used in machine learning, namely, the evidence lower bound, when using a negative IFE. The probabilistic rationale of the FEP argues that the brain's representations of the uncertain environment are the sufficient statistics of a probability density encoded in the brain—for example, means and variances for gaussian densities. The variational parameters are supposed to be encoded as physical variables in the brain. The brain statistically infers the external causes of sensory input by Bayesian filtering through its internal top-down model for predicting, or generating, sensory data. Filtering is a probabilistic approach to determining external states from noisy measurements of sensory data (Jazwinski, 1970). There is growing experimental support for the brain's maintenance of internal models of the environment to predict sensory inputs and to prepare actions (see Berkes, Orban, Lengyel, & Fiser, 2011, for instance). The computational operation of the abductive (Bayesian) inference is subserved by the brain variables, and the resulting perceptual mechanics is termed *recognition dynamics* (RD).

Although the FEP is promising in terms of accounting for inference in the brain (and active inference), several technical issues arise when it is applied to continuous state-space models of the world. First, the FEP minimizes the IFE at each point in time for successive sensory inputs (Friston, Stephan, Li, & Daunizeau, 2010). However, the objective function to be minimized is precisely the IFE continuously accumulated over a finite time.[1] The minimization must be performed considering trajectories over a temporal horizon

[1] According to the FEP, the updating or learning of the generative model occurs in the brain on a longer timescale than that associated with perceptual inference. To derive the

across which an organism encounters atypical events in its natural habitat and biology.

Second, the FEP employs the gradient-descent method for practically executing the minimization of the IFE (Friston et al., 2010), which is widely used in data analysis (e.g., dynamic causal modeling) and offers a solution to engineering optimization problems. The scheme enables one to find optimal solutions on the FE landscape, but the underlying variational principles (of least action) are not explicit.

Third, the FEP introduces the notion of generalized coordinates of motion, which comprise an infinite number of high-order derivatives that can account for analytic (i.e., smooth) random fluctuations (Friston, 2008a). The ensuing theoretical construct is a generalization of standard Newtonian mechanics.[2] However, there is no principled approach to specify the order of generalized motion. In practice, the generalized motion is truncated at a finite embedding order by assuming that the precision of random fluctuations on higher orders of motion disappears very quickly.

Fourth, the FEP introduces the hydrodynamics-like concepts of the path of a mode (motion of expectation) and the mode of a path (expected motion) by distinguishing the dynamic update from the temporal update of a time-dependent state (Friston, 2008b). Because the distinction is essential to ensure an equilibrium solution to the RD when employing the dynamical generative models, further theoretical exploration seems worthwhile.

Fifth, the FEP considers the states of the environment as "hidden" because what the brain faces is only a probabilistic sensory mapping. Subsequently, a distinction is made between the hidden-state representations responsible for intralevel dynamics and causal-state representations responsible for interlevel dynamics in the hierarchical brain (Friston, 2006). Such a distinction is based on a hierarchical generative model with dynamics on different timescales. Accordingly, a biophysically grounded formulation that enables this separation of timescales is required.

In this article, we present a mechanical formulation of the RD in the brain in the framework of Hamilton's principle of least action (Landau

RD of the slow variables for synaptic efficacy and gain, the time integral of the IFE is taken as an objective function; however, the gradient descent method is again executed in a pointwise manner in time (Friston & Stephan, 2007).

[2] In standard Newtonian mechanics, the mechanical state of a particle is specified by position and velocity, which is the first-order time derivative of position. The velocity changes in the presence of an applied force, and the resulting rate of change is termed *acceleration*, which is the second-order derivative of position. No physical observables are assigned to the dynamical orders beyond the second order. In some literature (see Schot, 1978, for instance), the concept of "jerk" is assigned to the third-order derivative as a physical meaning. From the mathematical perspective, such a generalization is not forbidden. However, higher orders are difficult to measure (Visser, 2004). More seriously, the third order raises the question of what causes jerk, like how force causes acceleration according to Newton's second law. The same impasse occurs for all higher orders.

& Lifshitz, 1976). Motivated by the aforementioned theoretical observations, we attempt to resolve some of the technical complexities in the FEP framework. Specifically, the goal is to recast the gradient-descent strategy of minimizing the IFE, which has thus far eluded an undergirding formal description, into a mathematical framework that is consistent with the normative physics principles. We do this by hypothesizing the IFE as a Lagrangian of the brain that enters a theoretical action, being the fundamental objective function to be minimized in continuous time under the principle of least action (see Sengupta, Tozzi, Cooray, Douglas, & Friston, 2016, for a technical essay sketching a model-independent Lagrangian formalism relevant to our idea). Consequently, we reformulate the RD by considering only the canonical, physical realities to eschew the generalized coordinates of infinitely recursive time derivatives of the continuous states of an organism's environment and brain. In the ensuing description, the dynamical state of a system is specified only by positions and their first-order derivatives.

In this work, supported by recent evidence (Markov et al., 2014; Michalareas, Vezoli, van Pelt, Schoffelen, & Kennedy, 2016), we admit the bi directional facet in informational flow in the brain. The environment begets sensory data at the brain-environment interface through structures such as sensory receptors or interoceptors within an organism. The incited electro-opto-chemical interaction in sensory neurons must transduce forward in the anatomical structure of the brain. While complying with the idea of perception as the construction of hypotheses, there must be a backward pathway as well in information processing in the functional hierarchy of the brain. To understand how such a bidirectional functional architecture emerges from the electrophysiology of biophysics and anatomical organization of the brain is a primary research interest (see Markov & Kennedy, 2013, for instance). We shall consider a simple model that effectively incorporates the functional hierarchy while focusing on the brain's perceptual mechanics for inferring the external world, given sensory data. The problem of learning of the environment via updating the internal model of the world and of active inference—changing sensations via action on the external world (see Friston, Daunizeau, & Kiebel, 2009; Buckley & Toyoizumi, 2018, for instance)—is deferred for an upcoming paper. Instead, we provide a broad discussion in section 5 on how the learning may work in our formulation.

Here, we outline how in this work we cast Bayesian filtering in the FEP by using a variational principle of least action and how we articulate the minimization of the sensory uncertainty in terms of the associated Lagrangian and Hamiltonian. Furthermore, given a particular form of the differential equations, afforded by computational neuroscience, one can see relatively easily how neuronal dynamics could implement the Bayesian filtering. First, according to the FEP, the brain represents the environmental features statistically efficiently by using the sufficient statistics μ. We assume that μ represents the state of the basic computational unit of the

neural attributes of perception in the brain. Such a constituent is considered a "perceptual particle," which may be a single neuron or physically coarse-grained population of neurons forming a small particle. Second, we postulate that the Laplace-encoded IFE in the FEP, denoted as F (see section 2.1), serves as an effective informational Lagrangian (IL) of the brain, which is denoted as $\mathcal{L}$. Accordingly, the informational action (IA),[3] which we denote by $\mathcal{S}$, is defined as the time integral of the approximate IFE (see section 3.1). Third, conforming to the Hamiltonian principle of least action, the equations of motion of the perceptual particles are derived mathematically by varying the IA with respect to both μ and $\dot{\mu}$. The resulting Lagrange equations constitute the perceptual mechanics, that is, the RD of the brain's inference of the external causes of sensory stimuli (see section 3.1). Fourth, we obtain the brain's informational Hamiltonian (IH) $\mathcal{H}$ from the Lagrangian via a Legendre transformation. Consequently, we derive a set of coupled, first-order differential equations for μ and its conjugate p_μ, which are equivalent to the perceptual mechanics derived from the Lagrange formalism. The resulting perceptual mechanics is our derived RD in the brain. Accordingly, the brain performs the RD in the state space spanned by the position μ and momentum p_μ of the constituting neural particles (see section 3.2).

Fifth, we adopt the Hodgkin-Huxley (H-H) neurons as biophysical neural correlates that form the basic perceptual units in the brain. We first derive the RD of sensory perception at a single-neuron level at which the membrane potential, ionic transport, and synaptic gating are the relevant physical attributes. Subsequently, we scale up the cellular formulation to furnish a functional hierarchical architecture of the brain. On this coarse-grained scale, the perceptual states are the averaged properties of many interacting neurons. We simplify the hierarchical picture with two classes of averaged variables for activation and connection, mediating the intra- and interlevel dynamics, respectively. According to our formulation of the hierarchical RD in the brain, as sensory perturbation occupies the lowest level (i.e., the sensory interface), the brain carries out the RD in its functional network and finds an optimal trajectory that minimizes the IA.

To summarize, we have adopted the IFE as an informational Lagrangian of the brain and subsequently employed the principle of least action to construct the Hamiltonian mechanics of cognition. In doing so, only positions and momenta of the neural particles have been addressed as dynamical variables. We do not distinguish the causal and hidden states, both of which must emerge as biophysical neuronal activities on different timescales. The resulting RD is statistically deterministic, arising from unpredictable motions of the environmental states and noisy sensory mapping. Furthermore,

[3] Note that one must not confuse "informational action" with the "physical action" of an organism.

the derived RD describes not only the dynamics of the brain's representation of hidden states of the world but also the prediction errors. We will see later that the latter corresponds to momenta in the setting of Hamiltonian mechanics. Note that the dynamics of prediction errors is not part of the conventional formulation of generalized filtering under the FEP; rather it emerges naturally in the current variational formulation. The successful solutions of the RD are stable equilibrium trajectories in the neural state space, which specify the tightest upper bound of the sensory uncertainty by conforming to the rephrased FEP. Our formulation allows solutions in an analytical form in linear regimes near fixed points, expanded in terms of the eigenvectors of the Jacobian; thus, it provides a tractable real-time analysis. We hope that our theory will motivate further investigations of some model brains with numerical simulations as well as of active inference and learning problems.

The remainder of this article is organized as follows. We first recapitulate the FEP in section 2 to support our motivation for casting the gradient descent scheme into the standard mechanical formulation. In section 3, we present the RD reformulated in the Lagrangian and Hamiltonian formalisms. In section 4, biophysical implementations of our theory at the cellular level and in the scaled-up hierarchical brain are formulated, and nonlinear as well as linear dynamical analyses are carried out. Finally, a discussion is presented in section 5.

2 The Free Energy Principle

To present our motivation for this article, we briefly discuss the IFE and FEP, which are currently used in the brain sciences to derive the RD. The RD is an organism's computational framework for executing the minimization of the IFE in the brain under the FEP. In practice, there are various IFE-minimizing schemes, such as variational message passing and belief propagation, that do not invoke treatment using generalized coordinates of motion. Our treatment here, which accommodates the notion of generalized motion, is more relevant to the Bayesian filtering and predictive coding schemes that have become a popular analogy for message passing in the brain. Filtering is the problem of determining the state of a system from noisy measurements (Jazwinski, 1970). For a detailed technical appraisal of the FEP, we refer to Buckley et al., (2017) from which we borrow the mathematical notations.

2.1 Informational free energy. A living organism occupies a finite space and time in the unbounded, changing world while interacting with the rest of the world, comprising its environment. The states of the environment are collectively denoted as ϑ, which are "hidden" from the organism's perspective. The signals from the environment are registered biophysically at the organism's sensory interface as sensory data φ.

The organism's brain faces uncertainty when it attempts to predict the sensory inputs, the amount of which is quantified as *sensory uncertainty* H. The sensory uncertainty is defined as an average of the self-information, $-\ln p(\varphi)$, over the probability density $p(\varphi)$ encoded at the interface:

$$H \equiv \int d\varphi \left\{-\ln p(\varphi)\right\} p(\varphi). \tag{2.1}$$

The self-information, which is also termed the sensory *surprise* or *surprisal* in information theory, quantifies the survival tendency of living organisms in an unpredictable environment; it is the logarithm of the inverse of the probability that they will be found in a particular sensory state over time. Assuming that the sensory density describes an ergodic ensemble of sensory streaming,[4] one may convert the sensory uncertainty into a time average as

$$\int d\varphi \left\{-\ln p(\varphi)\right\} p(\varphi) = \frac{1}{T}\int_0^T dt \left\{-\ln p(\varphi(t))\right\},$$

where T is the temporal window over which exchange with the environment occurs (i.e., a temporal horizon). Here, one may manipulate the right-hand side of the preceding equation by adding a nonnegative, Kullback-Leibler divergence to the integrand to obtain

$$-\ln p(\varphi) + \int d\vartheta q(\vartheta) \ln \frac{q(\vartheta)}{p(\vartheta|\varphi)} \to \int d\vartheta q(\vartheta) \ln \frac{q(\vartheta)}{p(\vartheta, \varphi)}.$$

The outcome brings about the mathematical definition of the IFE,

$$\mathcal{F}[q(\vartheta), p(\vartheta, \varphi)] \equiv \int d\vartheta q(\vartheta) \ln \frac{q(\vartheta)}{p(\vartheta, \varphi)}, \tag{2.2}$$

which is expressed as a functional of the two probability densities, $q(\vartheta)$ and $p(\vartheta, \varphi)$, termed the recognition density (R-density) and the generative density (G-density), respectively. The R-density is the organism's probabilistic representation of the external world, which the organism's brain uses in approximately inferring the causes ϑ of inputs φ. The G-density, a joint probability between ϑ and φ, underlies a generative model of how the sensory data are biophysically produced by interaction between the brain and the environment. By construction, the surprisal is smaller than the IFE by the

[4] This ergodicity assumption is an essential ingredient of the FEP, which hypothesizes that the ensemble average of the surprisal is equal to its time average, considering the surprisal to be a statistical, dynamical quantity.

added positive amount; accordingly, the sensory uncertainty is bounded from above in accordance with

$$\int dt \left[-\ln p(\varphi) \right] \leq \int dt \mathcal{F}[q(\vartheta), p(\vartheta, \varphi)]. \tag{2.3}$$

Note that the sensory uncertainty on the left-hand side of equation 2.3 specifies the accumulated surprisal over a temporal horizon involved in an environmental event.

Equation 2.3 constitutes a mathematical statement of the FEP: "All viable organisms attempt to avoid being placed in an atypical situation in their environmental habitats for existence by minimizing the sensory uncertainty. However, organisms do not possess direct control over the sensory distribution $p(\varphi)$; instead, they minimize the upper bound of equation 2.3, $\int dt \mathcal{F}$, as a proxy for the sensory uncertainty." The brain conducts the minimization probabilistically by updating the R-density to approximate the posterior density $p(\vartheta|\varphi)$, namely, by carrying out the Bayesian inference of the causes ϑ of the sensory data φ. In the conventional application of the FEP, the following approximate inequality is usually employed (see Friston & Kiebel, 2009; Friston, Adams Perrinet & Breakspear, 2012, for instance):

$$-\ln p(\varphi) \leq \mathcal{F}. \tag{2.4}$$

However, note that the inequality, equation 2.4, is not equivalent to equation 2.3 in general. It is only a point approximation piecewise in time.

Note that the negative-evidence bound in equation 2.4 does not specify the form of the R-density. This means that one is at liberty to use a convenient form that renders the minimization of variational IFE tractable. Usually what one invokes is a gaussian fixed form for the R-density, called the *Laplace approximation*:

$$q(\vartheta) = \frac{1}{\sqrt{2\pi\zeta}} \exp\left\{-(\vartheta - \mu)^2/(2\zeta)\right\} \equiv \mathcal{N}(\vartheta; \mu, \zeta), \tag{2.5}$$

which is fully characterized simply by its means μ and variances ζ, namely, first- and second-order sufficient statistics, respectively. Then, by substituting equation 2.5 into equation 2.2 and after some technical approximations (see Buckley et al., 2017, for details), one can convert the IFE functional into a function of only the means μ, given the sensory input φ,

$$\mathcal{F}[q(\vartheta), p(\vartheta, \varphi)] \to -\ln p(\mu, \varphi) \equiv F(\mu, \varphi), \tag{2.6}$$

where the dependence on the conditional variances disappears because they are a straightforward analytic function of the means. The resulting IFE

function F in equation 2.6 is termed the *Laplace-encoded* IFE in which the parameters μ, specifying the organism's belief or expectation of the environmental states, are the organism's probabilistic representation of the external world. In turn, it is argued that the variational parameters μ are encoded in the brain as biophysical variables.

To proceed with minimization of the IFE in the filtering scheme, a generative model for noisy-data measurement and the equations of motion of the states must be supplied. The FEP assumes a formal homology between the external dynamics and the organism's top-down belief. The former describes, according to the laws of physics, the equations of motion of environmental states and the sensory-data registering process. The latter prescribes the internal dynamics of the representations of environmental states and the generative model of sensory data in the organism's brain (Friston, Daunizeau, Kilner, & Kiebel, 2010). Following this idea, one hypothesizes that sensory data φ are predicted on the basis of the expected hidden state, which is denoted by the mean μ of the R-density, according to a linear or nonlinear mapping,

$$\varphi = g(\mu) + z, \tag{2.7}$$

where $g(\mu)$ is a map from μ onto φ and z is the involved random fluctuation. Furthermore, the brain's representations μ of the causes are assumed to obey the stochastic equation of motion,

$$\frac{d\mu}{dt} = f(\mu) + w, \tag{2.8}$$

where $f(\mu)$ is a linear or nonlinear function of the organism's expectation of environmental dynamics and w is the associated random fluctuation.

Assuming mutually uncorrelated gaussian fluctuations, w and z, of the organism's beliefs, one may furnish the models for the likelihood $p(\varphi|\mu)$ and the empirical prior $p(\mu)$, which jointly enter the Laplace-encoded IFE in equation 2.6 in the factorized form:

$$p(\varphi, \mu) = p(\varphi|\mu)p(\mu). \tag{2.9}$$

Using the notation introduced in equation 2.5, they are given explicitly as

$$p(\varphi|\mu) = \mathcal{N}(\varphi - g(\mu); 0, \sigma_z), \tag{2.10}$$

$$p(\mu) = \mathcal{N}(\dot{\mu} - f(\mu); 0, \sigma_w), \tag{2.11}$$

where we have set $\dot{\mu} = d\mu/dt$, and the normal densities are assumed to possess zero means with variances σ_z and σ_w, respectively. When the fluctuations are statistically stationary, the variances are handled as constant;

however, nonstationary fluctuation can also be taken into account by assuming an explicit time dependence in the variances. Finally, by substituting equations 2.10 and 2.11 into equation 2.6, one can convert the Laplace-encoded IFE, up to a constant, into

$$F(\mu, \varphi) = \frac{1}{2}\sigma_z^{-1}\varepsilon_z^2 + \frac{1}{2}\sigma_w^{-1}\varepsilon_w^2 + \frac{1}{2}\ln(\sigma_z\sigma_w), \tag{2.12}$$

where the new variables have been defined as

$$\varepsilon_z \equiv \varphi - g(\mu) \quad \text{and} \quad \varepsilon_w \equiv \dot{\mu} - f(\mu).$$

The auxiliary variable ε_z specifies the discrepancy between the sensory data φ and the brain's prediction $g(\mu)$. Similarly, ε_w specifies the discrepancy between the change of expectations $\dot{\mu}$ and the organism's belief $f(\mu)$. We will see that the equivalence between the change of expectations and the expectation of the change of external states follows from a minimization formulation in generalized coordinates of motion.

It is straightforward to extend the formulation to the multiple correlated noisy inputs. However, for simplicity, we shall continue to work in the single-variable picture and extend it to a general situation later.

2.2 Gradient Descent Scheme of the RD. With the Laplace-encoded IFE as an instrumental tool, the organism's brain searches for the tightest bound for the surprisal, conforming to equation 2.4, by varying its internal states μ. The critical question here is what machinery the brain employs for the minimization procedure. Typically the gradient descent method in machine learning theory is employed in the conventional approach.

To give an idea of the gradient-descent scheme, we set up a simple gradient-descent equation here, in the usual manner, by regarding the IFE function F as an objective function:

$$\dot{\mu} = -\kappa\nabla_\mu F. \tag{2.13}$$

In this equation, $\dot{\mu}$ denotes a temporal or sequential update of the brain variable μ, ∇_μ is the gradient operator with respect to μ, and κ is the learning rate that controls the speed of optimization. In the steady state, defined by $\dot{\mu} \equiv 0$, the solution $\mu^{(0)}$ to the relaxation equation, equation 2.13, must satisfy $\nabla_\mu F = 0$. Subsequently, it may be interpreted that such a solution corresponds to an equilibrium (or fixed) point of the IFE function F, specifying a local minimum in the IFE landscape.

By inspection, however, we find that the gradient-descent construct in the above approach causes ambiguity when applied to dynamic causal models such as equation 2.8 because imposing the condition $\dot{\mu} \equiv 0$ on the left-hand side of equation 2.13 does not guarantee a desired equilibrium

point in the state space spanned by μ. The reason is that $\dot{\mu}$ also appears on the right-hand side of equation 2.13 via F. The gradient operation on the right-hand side of equation 2.13 can be performed explicitly for F, given by equation 2.12, to obtain

$$\hat{\mu} \cdot \nabla_{\mu} F = -\sigma_z^{-1}(\varphi - g)\frac{\partial g}{\partial \mu} - \sigma_w^{-1}(\dot{\mu} - f)\frac{\partial f}{\partial \mu}.$$

This subtlety does not appear in the formulation using generalized coordinates of motion, which incorporates the aspect of continually changing external or hidden states via the mathematical construct of unbounded, higher-order motion of the generalized coordinates.[5]

For completeness, we now describe the formulation in generalized coordinates of motion. It is an attempt to allow a more precise specification of a system's dynamical state. This formulation is useful when random fluctuations on higher-order motion are to be considered. Effectively, this allows one to eschew Wiener assumptions and deal with smooth random perturbations (Friston, 2008a). The generalized coordinates are defined as a row vector in the state space spanned by all orders of time derivatives of a bare state μ,

$$\tilde{\mu} = (\mu, \mu', \mu'', \ldots) \equiv (\mu_{[0]}, \mu_{[1]}, \mu_{[2]}, \ldots), \tag{2.14}$$

where vector components are defined, with the understanding that $\mu_{[0]} \equiv \mu$, as

$$\mu_{[n+1]} = \mu'_{[n]} \equiv D\mu_{[n]}.$$

Note that the notation $D\mu_{[n]} \equiv \mu'_{[n]}$ has been introduced to denote the dynamical update of the component $\mu_{[n]}$, which is in contrast to the notation $\dot{\mu}_{[n]}$ for the sequential update. Furthermore, two components of a vector at different dynamical orders in the generalized coordinates are mutually independent variables. Similarly, the sensory data $\tilde{\varphi}$ are expressed in the generalized coordinates as a row vector:

$$\tilde{\varphi} = (\varphi, \varphi', \varphi'', \ldots) \equiv (\varphi_{[0]}, \varphi_{[1]}, \varphi_{[2]}, \ldots). \tag{2.15}$$

[5] The terminology of the generalized coordinates in generalized filtering is dissimilar from its common usage in physics. In classical mechanics, the generalized coordinates refer to the independent coordinate variables that are required to completely specify the configuration of a system with a holonomic constraint, not including their temporal derivatives. The number of generalized coordinates determines the degree of freedom in the system (Landau & Lifshitz, 1976). Therefore, the term *generalized states* seems more suitable than *generalized coordinates* in generalized filtering.

Each component in the vectors $\tilde{\mu}$ and $\tilde{\varphi}$ is to be considered as a dynamically independent variable. Moreover, assuming that the random fluctuations, z and w, are analytic, they have been expressed in the generalized coordinates as $\tilde{z}$ and $\tilde{w}$, respectively. Then the generalization of equations 2.7 and 2.8 follows after some technical approximations as (see Buckley et al., 2017, for details)

$$\tilde{\varphi} = \tilde{g} + \tilde{z}, \tag{2.16}$$

$$D\tilde{\mu} = \tilde{f} + \tilde{w}, \tag{2.17}$$

where $D\tilde{\mu} = (\mu', \mu'', \mu''', \ldots)$. For reference, we explicitly spell out equations 2.16 and 2.17 at dynamical order n as

$$\varphi_{[n]} = \frac{\partial g}{\partial \mu}\mu_{[n]} + z_{[n]},$$

$$D\mu_{[n]} = \frac{\partial f}{\partial \mu}\mu_{[n]} + w_{[n]}.$$

Note that different dynamical orders of the noise terms $\tilde{z}$ and $\tilde{w}$ may be considered to be statistically correlated in general. Then the Laplace-encoded IFE can be mathematically constructed from multivariate correlated gaussian noises with zero means and covariance matrices Σ_w and Σ_z as follows:

$$F(\tilde{\mu}, \tilde{\varphi}) = \frac{1}{2}\{\dot{\tilde{\mu}} - \tilde{f}\}\Sigma_w^{-1}\{\dot{\tilde{\mu}} - \tilde{f}\}^T + \frac{1}{2}\ln|\Sigma_w| + \frac{1}{2}\{\tilde{\varphi} - \tilde{g}\}\Sigma_z^{-1}\{\tilde{\varphi} - \tilde{g}\}^T + \frac{1}{2}\ln|\Sigma_z|, \tag{2.18}$$

where $\{\dot{\tilde{\mu}} - \tilde{f}\}^T$ is the transpose of row vector $\{\dot{\tilde{\mu}} - \tilde{f}\}$, and $|\Sigma_w|$ and Σ_w^{-1} are the determinant and inverse of the covariance matrix Σ_w, respectively. In many practical exercises, however, conditional independence among different dynamical orders is usually imposed. Consequently, the noise distribution at each dynamical order is assumed to be an uncorrelated gaussian density about zero means. This simplification corresponds to the Wiener process or Markovian approximation (Jazwinski, 1970). Here, we recall that the generalized states $\tilde{\mu}$ are the means of the brain's probabilistic model of the dynamical world, which is the R-density in equation 2.5, after rewriting in the generalized coordinates. Note that equation 2.18 is a direct generalization of equation 2.12.

Furnished with the extratheoretical constructs, the IFE becomes a function of the generalized coordinates $\tilde{\mu}$, given sensory data $\tilde{\varphi}$: $F = F(\tilde{\mu}, \tilde{\varphi})$. Accordingly, the gradient-descent scheme must be extended to incorporate the generalized motions in its formulation. This is achieved by the theoretical prescription that the dynamical update $D\tilde{\mu}$ is distinctive from

the sequential update $\dot{\tilde{\mu}}$. Consequently, one recasts equation 2.13 into the form

$$\dot{\tilde{\mu}} - D\tilde{\mu} = -\kappa \nabla_{\tilde{\mu}} F(\tilde{\mu}, \tilde{\varphi}). \tag{2.19}$$

This form is effectively a gradient descent equipped with a solenoidal flow (or in a moving frame of reference). It is argued that the solution to this equation (when the gradient with respect to the IFE is zero) renders the motion of the expectation the same as the expected motion (see Friston et al., 2010). This licenses the equality in equation 2.19, which states that at a minimum of F, the two rates $\dot{\tilde{\mu}}$ and $D\tilde{\mu}$ become coincident—that is $\dot{\mu}_{[n]} = D\mu_{[n]}$ at every dynamic order n. The entire minimization procedure is compactly expressed in the literature as

$$\tilde{\mu}^* = \arg\min_{\tilde{\mu}} F(\tilde{\mu}, \tilde{\varphi}|m),$$

where we have inserted m in F to indicate explicitly that the minimization is conditioned on the generative model of an organism.

In brief, equation 2.19 furnishes the RD from the gradient-descent formulation in the FEP. The brain performs the RD of perceptual inference by biophysically implementing equation 2.19 in the gray matter. A line attractor solution $\tilde{\mu}^*$ specifies the minimum value of the IFE, say, $F_{\min} = F(\tilde{\mu}^*, \tilde{\varphi})$, yielding the tightest bound of the surprisal (see equation 2.4), associated with a given sensory experience $\tilde{\varphi}$.

3 The Informational Action Principle

The RD condensed in section 2.2 is based on the mathematical statement of the FEP given by equation 2.4, which is a point approximation of equation 2.3. Here we reformulate the RD by complying with the full mathematical statement of the FEP given in equation 2.3. Accordingly, we need a formalism that allows minimization of the time integral of the IFE rather than the IFE at each point in time. We have assimilated that the theoretical action in the principle of least action neatly serves the goal (Landau & Lifshitz, 1976). This formalism allows us to eschew the introduction of the generalized coordinates of a dynamical state comprising an infinite number of time derivatives of the brain state μ. Consequently, the distinctive classification of the time derivative of the parametric update ($\dot{\mu}$) and dynamical update ($D\mu$) of the state variable is not required. In what follows, we consistently use the dot symbol to denote the time derivative of a dynamical variable.

We will frame the variational principle of least action for the RD under the FEP. Our formulation of the RD reveals some very interesting interpretations of factors such as the prediction error and its inverse variance (i.e., precision). For example, prediction error becomes the momentum of a neural

particle, while precision becomes its inertial mass. In section 4, we unpack them in the context of neuronal dynamics (as described by the Hodgkin-Huxley equation) and consider hierarchical architectures under the informational action principle.

3.1 Lagrangian Formalism. To formulate the RD from the principle of least action, the Lagrangian of the system must be supplied. We define the IL of the brain, denoted by $\mathcal{L}$, as the Laplace-encoded IFE function

$$\mathcal{L}(\mu, \dot{\mu}; \varphi) \equiv F(\mu, \dot{\mu}; \varphi),$$

where we have placed the semicolon in the argument of $\mathcal{L}$ to indicate that μ and $\dot{\mu}$ are the two dynamical variables of the brain, given a sensory input φ. The sensory inputs are stochastic and time dependent; in general, $\varphi = \varphi(t)$, reflecting the changing external states, the generative processes of which are to be supplied by the laws of physics. The proposed IL is not a physical quantity but rather an information-theoretic object. When we take equation 2.12 as an explicit expression for F, the IL is expressed as

$$\mathcal{L}(\mu, \dot{\mu}; \varphi) = \frac{1}{2}\sigma_w^{-1}(\dot{\mu} - f(\mu))^2 + \frac{1}{2}\sigma_z^{-1}(\varphi - g(\mu))^2. \tag{3.1}$$

Note that we have dropped the term $\frac{1}{2}\ln(\sigma_z\sigma_w)$ in writing equation 3.1 by assuming it as a constant, which then does not affect the dynamics of μ and $\dot{\mu}$. This assumption of statistical stationarity may be lifted by introducing time dependence in the variances (MacDonald, 2006),

$$\sigma_w = \sigma_w(t) \quad \text{and} \quad \sigma_z = \sigma_z(t).$$

Still, however, the dropped term does not affect the dynamics because a term that can be expressed as a total time derivative in the Lagrangian will not affect the resulting equations of motion (Landau & Lifshitz, 1976).

Next, we postulate that the perceptual dynamics of the neural particles conforms to the principle of least action (Landau & Lifshitz, 1976). Accordingly, we suppose that the brain's perceptual operation corresponds to the search for an optimal dynamical path that minimizes the informational action (IA), denoted by $\mathcal{S}$,

$$\mathcal{S} \equiv \int_{t_i}^{t_f} dt\, \mathcal{L}(\mu, \dot{\mu}; \varphi), \tag{3.2}$$

where $t_f - t_i$ is the temporal horizon over which a living organism engages with the environment. When the functional derivative of $\mathcal{S}$ is taken with respect to μ and $\dot{\mu}$, we obtain

$$\delta\mathcal{S} = \left[\frac{\partial\mathcal{L}}{\partial\mu}\delta\mu\right]_{t_i}^{t_f} - \int_{t_i}^{t_f} dt\left(\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\mu}} - \frac{\partial\mathcal{L}}{\partial\mu}\right)\delta\mu.$$

By imposing $\delta S \equiv 0$ under the condition that initial and final states are fixed,

$$\delta\mu(t_i) = 0 = \delta\mu(t_f),$$

we derive the Lagrangian equation as

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\mu}} - \frac{\partial\mathcal{L}}{\partial\mu} = 0. \tag{3.3}$$

Using the specified Lagrangian, equation 3.1, in equation 3.3, we obtain a Newtonian equation of motion for the brain variable μ,

$$\sigma_w^{-1}\dot{v} = \bar{\Lambda}_1 + \bar{\Lambda}_2, \tag{3.4}$$

where we have defined the kinematic velocity as

$$v \equiv \dot{\mu}$$

and the additional notations on the right-hand side as

$$\bar{\Lambda}_1 \equiv \sigma_w^{-1} f\frac{\partial f}{\partial\mu} \quad \text{and} \quad \bar{\Lambda}_2 \equiv -\sigma_z^{-1}(\varphi - g)\frac{\partial g}{\partial\mu}. \tag{3.5}$$

Equation 3.4 entails the RD of the brain in the Lagrangian formulation. As an analogy, we interpret that the inverse of the variance σ_w^{-1} plays the role of *inertial mass* of the neural particles. Accordingly, the left-hand side of equation 3.4 represents an *inertial force*—the product of inertial mass and acceleration $\ddot{\mu}$. Note that the inverse of variance is interpreted as precision in the Friston formulation (Buckley et al., 2017), which gives a measure for the accuracy of the brain's expectation or prediction of sensory data. Therefore, the precision is metaphorically the "informational mass" of the neural particle, which we shall denote throughout as

$$m_\alpha \equiv \sigma_\alpha^{-1} \quad \text{and} \quad \alpha = w, z.$$

Furthermore, the terms $\bar{\Lambda}_i$, $i = 1, 2$, on the right-hand side are interpreted as the "forces" that drive the internal ($\bar{\Lambda}_1$) as well as sensory ($\bar{\Lambda}_2$) excitations in the brain. The acceleration can be evaluated from $\ddot{\mu} = \sum \bar{\Lambda}_i / m_w$ when the net force is known.

While the organism's brain integrates the RD for an incoming sensory input, an optimal trajectory $\mu^*(t)$ is continuously achieved in the neural

configuration space. Moreover, the steady-state condition in the long-time limit $t \to \infty$ is given by

$$\dot{\mu}^* = v^* = \text{const}, \tag{3.6}$$

where the net force vanishes. Note that equation 3.6, which defines an attractor, $\mu_{eq} = \mu^*(\infty)$, is more general than the simple solution, $\dot{\mu}^* = 0$. In other words, we allow for an optimal trajectory as opposed to a fixed point. The optimal trajectory $\mu^*(t)$ minimizes the IA, which in turn provides the organism with the tightest estimate of the sensory uncertainty (see equation 2.3).

3.2 Hamiltonian Formalism. The mechanical formulation can be made more modish in terms of Hamiltonian language, which admits position and momentum as independent brain variables instead of position and velocity in the Lagrangian formulation. The positions and momenta span the phase space of a physical system, which defines the neural state space of the organism's brain.

The "canonical" momentum p, which is conjugate to the position μ, is defined via the Lagrangian $\mathcal{L}$ as (Landau & Lifshitz, 1976)

$$p \equiv \frac{\partial \mathcal{L}}{\partial \dot{\mu}} = m_w (\dot{\mu} - f), \tag{3.7}$$

which evidently differs from the "kinematic" momentum $m_w v = m_w \dot{\mu}$ where m_w is the inertial mass σ_w^{-1}. Then the informational Hamiltonian (IH), denoted by $\mathcal{H}$, may be constructed from the Lagrangian through a Legendre transformation:

$$\mathcal{H}(\mu, p; \varphi) = \sum \frac{\partial \mathcal{L}}{\partial \dot{\mu}} \dot{\mu} - \mathcal{L}(\mu, \dot{\mu}; \varphi). \tag{3.8}$$

The first term on the right-hand side of equation 3.8 can be further manipulated to yield

$$\sum \frac{\partial \mathcal{L}}{\partial \dot{\mu}} \dot{\mu} = m_w \dot{\mu}^2 - m_w \dot{\mu} f.$$

By plugging the outcome and the Lagrangian $\mathcal{L}$ given in equation 3.1 into equation 3.8, we obtain the IH as a function of μ and p, given φ, as follows:

$$\mathcal{H}(\mu, p; \varphi) = \mathcal{T}(p) + \mathcal{V}(\mu, p; \varphi), \tag{3.9}$$

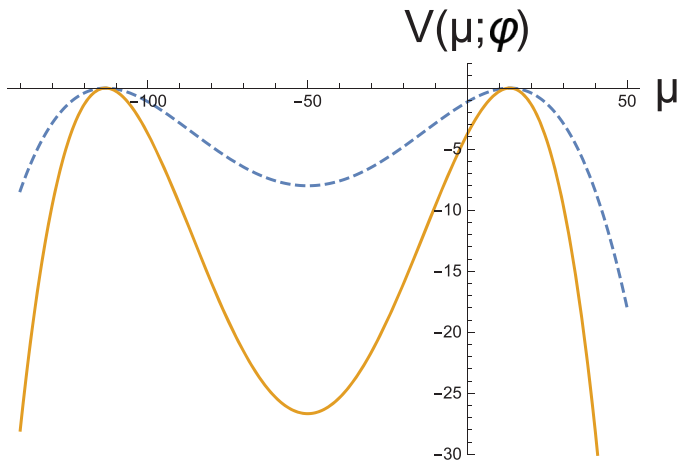

Figure 1: The potential energy given in equation 3.12, in arbitrary units, where the dashed and solid curves are for the variances $\sigma_z = 100$ and 30, respectively. Both cases exhibit a stable minimum in the central well and two unstable maxima on the side hills, which contribute to determining the IFE landscape.

where equation 3.7 has been used to replace $\dot{\mu}$ with p. The first term on the right-hand side of equation 3.9 is the "kinetic energy," which depends only on momentum:

$$\mathcal{T}(p) = \frac{p^2}{2m_w}. \tag{3.10}$$

Moreover, the second term on the right-hand side of equation 3.9 is the "potential energy," which depends on both position and momentum:

$$\mathcal{V}(\mu, p; \varphi) = V(\mu; \varphi) + pf(\mu), \tag{3.11}$$

where we have defined the momentum-independent term separately as

$$V(\mu; \varphi) = -\frac{1}{2} m_z (\varphi - g)^2. \tag{3.12}$$

We remark that the sensory stimuli φ enter the Hamiltonian only through the potential-energy part V, which becomes "conservative" when φ is static. Here, we assume that the variances associated with the noisy data are constant. For time-varying sensory inputs, in general, the Hamiltonian is nonautonomous. In Figure 1, we depict the conservative potential energy by using three-term approximations for the generative function,

$$g(\mu) \approx b_1 + b_2\mu + b_2\mu^2.$$

For convenience, we have assumed a constant sensory input, $\varphi = 15$, and set parameters as $(b_1, b_2, b_3) = (0, 1, 0.01)$. We have observed numerically that the static sensory signal φ changes the distance between two unstable fixed points but does not affect the location of the stable equilibrium point. Furthermore, the depth of the stable equilibrium valley increases with the magnitude of φ.

Next, we take the total derivative of the Hamiltonian given in equation 3.8 with respect to μ and $\dot{\mu}$ to obtain

$$\begin{aligned} d\mathcal{H}(\mu, p; \varphi) &= \sum d\,(p\dot{\mu}) - d\mathcal{L}(\mu, \dot{\mu}; \varphi) \\ &= \dot{\mu}dp + pd\dot{\mu} - \left(\frac{\partial \mathcal{L}}{\partial \mu}d\mu + \frac{\partial \mathcal{L}}{\partial \dot{\mu}}d\dot{\mu}\right) \\ &= -\dot{p}_\mu d\mu + \dot{\mu}dp. \end{aligned}$$

By comparing the above expression with the formal expansion,

$$d\mathcal{H} = \frac{\partial \mathcal{H}}{\partial \mu}d\mu + \frac{\partial \mathcal{H}}{\partial p}dp,$$

we identify the Hamilton equations of motion for independent variables μ and p of a neural particle:

$$\dot{\mu} = \frac{\partial \mathcal{H}}{\partial p}, \tag{3.13}$$

$$\dot{p} = -\frac{\partial \mathcal{H}}{\partial \mu}. \tag{3.14}$$

For a given $\mathcal{H}$ in equation 3.9, we spell out the right-hand side of equation 3.13 as

$$\dot{\mu} = \frac{1}{m_w}p + f, \tag{3.15}$$

which is identical to equation 3.7. Similarly, equation 3.14 is spelled out as

$$\dot{p} = -\frac{\partial V}{\partial \mu} - \frac{\partial f}{\partial \mu}p. \tag{3.16}$$

The first term on the right-hand side of equation 3.16 specifies the conservative force,

$$-\frac{\partial V}{\partial \mu} \to -\sigma_z^{-1}(\varphi - g)\frac{\partial g}{\partial \mu}.$$

On the other hand, the second term on the right-hand side of equation 3.16 specifies the dissipative force, where $\partial f/\partial \mu$ plays the role of a damping coefficient.

The derived set of coupled equations for the variables μ and p furnishes the RD of the brain in phase space spanned by μ and p, which involve only first-order time derivatives. When the time derivative is taken once more for both sides of equation 3.15, followed by the substitution of equation 3.16 for $\dot{p}$, the outcome is identical to the Lagrangian equation of motion, equation 3.4. This observation confirms that the two mechanical formulations, one from the Lagrangian and the other from the Hamiltonian, are in fact equivalent.

In the Hamiltonian formulation, the brain's fulfilling of the RD is equivalent to finding an optimal trajectory $(\mu^*(t), p^*(t))$ in phase space. For a static sensory input, the dynamics governed by equations 3.15 and 3.16 is autonomous, and for a time-dependent sensory input, it becomes nonautonomous. The RD can be integrated by providing appropriate models for the generative functions f and g. The attractor $(\mu^*(\infty), p^*(\infty))$ would be a focus or center in phase space, which can be calculated by simultaneously imposing the following conditions on the left-hand sides of equations 3.15 and 3.16:

$$\dot{\mu}^* = 0 \quad \text{and} \quad \dot{p}^* = 0. \tag{3.17}$$

One can readily confirm that these fixed-point conditions match the Newtonian equilibrium condition, $\sum_i \bar{\Lambda}_i = 0$ in the Lagrangian formulation (see section 3.1). The situation corresponds to the brain's resting state at a local minimum on the energy landscape defined by the Hamiltonian function.

3.3 Multivariate Formulation. Having established the Hamiltonian dynamics for a single brain variable μ, we now extend our formulation to the general case of the multivariate brain. We denote $\{\mu\}$ as a row vector of N brain states as in section 2.1:

$$\{\mu\} = (\mu_1, \mu_2, \ldots, \mu_N).$$

The brain states respond to multiple sensory inputs in a general manner:

$$\{\varphi\} = (\varphi_1, \varphi_2, \ldots, \varphi_N).$$

For simplicity, we neglect the statistical correlation of the fluctuations associated with environmental variables and sensory inputs. Then, within

the independent-particle approximation of uncorrelated brain variables, the Laplace-encoded IFE given by equation 2.18 furnishes the multivariate Lagrangian:

$$\mathcal{L}(\{\mu\},\{\dot{\mu}\};\{\varphi\}) = \frac{1}{2}\sum_{\alpha=1}^{N}\left[m_{w\alpha}\left(\dot{\mu}_\alpha - f_\alpha(\{\mu\})\right)^2 + m_{z\alpha}\left(\varphi_\alpha - g_\alpha(\{\mu\})\right)^2\right], \tag{3.18}$$

where we have dropped the terms that contain only the variances, $\sigma_{z\alpha} = m_{z\alpha}^{-1}$ and $\sigma_{w\alpha} = m_{w\alpha}^{-1}$. One may extend equation 3.18 to interacting neural nodes in terms of the covariance matrix formulation, which is not our concern here either. Subsequently, the conjugate momentum to the generalized coordinate μ_α is determined by an explicit evaluation of

$$p_\alpha = \frac{\partial \mathcal{L}}{\partial \dot{\mu}_\alpha} = m_{w\alpha}\left(\dot{\mu}_\alpha - f_\alpha\right). \tag{3.19}$$

Note that the momentum p_α gives a measure of the discrepancy, weighted by $m_{w\alpha}$, between the change of the probabilistic representation of the environment $\dot{\mu}_\alpha$ and the organism's belief of it f_α. The weighting factor $m_{w\alpha}$ is the inertial mass, which is the precision in statistics. In turn, the Hamiltonian of the multivariate brain can be constructed from equation 3.9 as

$$\mathcal{H}(\{\mu\},\{p\};\{\varphi\}) = \mathcal{T}(\{\mu\},\{p\};\{\varphi\}) + \mathcal{V}(\{\mu\},\{p\};\{\varphi\}) \tag{3.20}$$

where first term on the right-hand side is the kinetic energy,

$$\mathcal{T}(\{p\};\{\varphi\}) \equiv \sum_{\alpha=1}^{N}\frac{{p_\alpha}^2}{2m_{w\alpha}}, \tag{3.21}$$

and the potential energy $\mathcal{V}$ is identified as

$$\mathcal{V}(\{\mu\},\{p\};\{\varphi\}) \equiv \sum_{\alpha=1}^{N}\left[-\frac{1}{2}m_{z\alpha}(\varphi_\alpha - g_\alpha)^2 + p_\alpha f_\alpha\right]. \tag{3.22}$$

Then it is straightforward to derive the RD of the variables μ_α and p_α, given sensory data φ_α, as

$$\dot{\mu}_\alpha = \frac{\partial \mathcal{H}}{\partial p_\alpha} = \frac{1}{m_{w\alpha}}p_\alpha + f_\alpha, \tag{3.23}$$

and for their conjugate momenta,

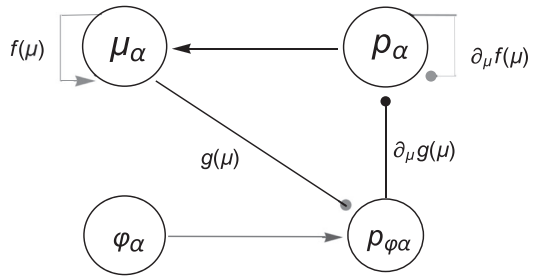

Figure 2: The perceptual circuitry at neural node α in which sensory data φ_α stream, where it is depicted that the computational units μ_α and p_α are positively activated by arrows and negatively by lines ending with filled dots. The conjugate momenta p_α, defined in equation 3.19, to the brain variables μ_α mimic the precision-weighted prediction errors in the language of predictive coding (Rao & Ballard, 1999).

$$\dot{p}_\alpha = -\frac{\partial \mathcal{H}}{\partial \mu_\alpha} = -\frac{\partial g_\alpha}{\partial \mu_\alpha} p_{\varphi\alpha} - \frac{\partial f_\alpha}{\partial \mu_\alpha} p_\alpha. \tag{3.24}$$

In equation 3.24, for notational convenience we have introduced an auxiliary quantity $p_{\varphi\alpha}$:

$$p_{\varphi\alpha} \equiv m_{z\alpha}(\varphi_\alpha - g_\alpha).$$

Equations 3.23 and 3.24 are a coupled set of equations for the computational units, μ_α and p_α, describing a specific brain variable and its conjugate momentum, respectively, given the sensory discrepancy $p_{\varphi\alpha}$ between the observed data φ_α and its prediction $g_\alpha(\mu_\alpha)$. According to the neural implementation of the predictive-coding theory (Summerfield & Egner, 2009), μ_α and p_α correspond to the representational and error neurons, respectively. With some working models for f_α and g_α, they shape the RD in the brain's multidimensional phase space in the Hamiltonian prescription. Figure 2 shows a schematic of the perceptual circuitry implied by the RD at a neural node. The classification of excitatory (positive) and inhibitory (negative) activation of the computational units is only for convenience because the sign of each term on the right-hand side of equations 3.23 and 3.24 depends on the generative function f_α and map g_α, which are not specified.

It is admissible to assume that the brain is in a resting state at the outset. As the sensory inputs φ_α arrive, the organism's brain performs the RD online by integrating equations 3.23 and 3.24 to attain an optimal trajectory in neural phase space,

$$\mu_\alpha = \mu_\alpha^*(t) \quad \text{and} \quad p_\alpha = p_\alpha^*(t),$$

which minimize the IA (see equation 3.2). The entire minimization procedure may be stated abstractly as

$$(\mu_\alpha^*, p_\alpha^*) = \arg\min_{\mu_\alpha, p_\alpha} \mathcal{S}(\mu_\alpha, p_\alpha; \varphi), \tag{3.25}$$

where $\mathcal{S}$ is the IA. We emphasize here that minimization is conditioned on the organism as a model of the world.

Note that our revised RD involves not only the organism's prediction of the environmental change via its representation μ_α but also the dynamics of its prediction error p_α.

4 Biophysical Implementation

We know that the anatomy and entire function of an organism's brain develop from single cells. In order to provide empirical Bayesian filtering in the FEP with a solid biophysical basis, we must start with known biophysical substrates and then introduce probabilities to describe a neuron, neurons, and a network. Thus far, however, most work has taken the reverse direction: theory prescribes first a conjectural model and then attempts to allocate possible neural correlates. At present, our knowledge remains limited on how biophysical mechanisms of neurons implicate predictions and model aspects of the environment. From this perspective, a neurocentric approach to the inference problem seems suggestive to bridge the gap (Fiorillo, 2008; Fiorillo, Kim, & Hong, 2014).

Here, we regard coarse-grained Hodgkin-Huxley (H-H) neurons as the generic, basic building blocks of encoding and transmitting a perceptual message in the brain. The famous H-H model continues to be used to this day in computational neuroscience studies of neuronal dynamics (Hodgkin & Huxley, 1952; Hille, 2001). In extracellular electrical recordings, the local field potential and multiunit activity result in combined signals from a population of neurons (Einevoll, Kayser, Logothetis, & Panzeri, 2013). Such averaged neuronal variables must subserve the perceptual states and conduct the cognitive computation in the brain. We shall call them "small" neural particles and envisage that a small neural particle functions as a node that collectively forms the whole neural network on a large scale. Before proceeding, we note that there have been many biophysical efforts to describe such averaged neuronal properties, such as the neural mass models and neural field theories (Jansen, Zouridakis, & Brandt, 1993; Jirsa & Haken, 1996; Robinson, Rennie, & Wright, 1997; David & Friston, 2003; Deco, Jirsa, Robinson, Breakspear, & Friston, 2008). Furthermore, we note the bottom-up effort of attempting to understand the large-scale brain function at the cortical microcircuit level based on the averaged spikes and synaptic inputs over a coarse-grained time interval (Potjans & Diesmann, 2014; Steyn-Ross & Steyn-Ross, 2016).

4.1 Single Cell Description. We first present how our formulation may be implemented at a single-cell level by hypothesizing that each neuron reflects the fundamentals of the perceptual computation of the whole system. A typical neuron receives current information about its surroundings from the sensory periphery via glutamate, which excites or inhibits the membrane potential V while regulating the gating variables γ_l and ionic concentrations n_l, where l is the ion channel index. We assume that $(V, \{n_l\}, \{\gamma_l\})$ represents the neural states of a neuron as a neural observer in the neural configurational space (Fiorillo et al., 2014). We encapsulate the neural states as components in a multidimensional row vector:

$$\{\mu\} = (V, \{n_l\}, \{\gamma_l\}) = (\mu_1, \mu_2, \mu_3, \ldots).$$

The H-H equation for excitation of the membrane voltage V in a spatially homogeneous cell is given by

$$C\frac{dV}{dt} = \sum_l \gamma_l G_l (E_l - V) + I_{ex}(t), \tag{4.1}$$

where C is the membrane capacitance; G_l is the maximal conductance of ion channel l; γ_l is the probability factor associated with the opening or closing channel l, which in general is a product of activation and inactivation gating variables; and I_{ex} is the external driving current. For simplicity, contributions from leakage current as well as synaptic input are assumed to be included in the external currents. In general, the reverse potential E_l of the lth ion channel is given, allowing its time dependence via nonequilibrium ion concentrations to be expressed as

$$E_l(t) = \frac{k_B T}{q_l} \ln \frac{n_{li}(t)}{n_{lo}(t)}, \tag{4.2}$$

where k_B is the Boltzmann constant, T is the metabolic temperature of an organism, q_l is the ionic charge of channel l, and $n_{li}(t)$ and $n_{lo}(t)$ are the instantaneous ion concentrations inside and outside the membrane, respectively. In the steady state without external current, $I_{ex} = 0$, and V tends to the resting (Nernst) potential $V(t \to \infty)$ while retaining ionic concentrations in electrochemical equilibrium. The gating variable γ_l of ion channels is assumed to obey the following chemical kinetics,

$$\frac{d\gamma_l}{dt} = -\frac{1}{\tau_l}(\gamma_l - \gamma_{leq}) + \eta_l, \tag{4.3}$$

where η_l is the involved noise. The relaxation time τ_l and steady-state gating variable γ_{leq} in equation 4.3 depend on the membrane potential in general:

$$\tau_l = \tau_l(V) \quad \text{and} \quad \gamma_{leq} = \gamma_{leq}(V).$$

For ionic concentration dynamics, we suppose that ion concentrations $\{n_l\}$ vary slowly compared to the membrane potential and gating-channel kinetics, and we consequently treat them as static in our work. This restriction can be lifted when a more detailed description is required for ion concentration dynamics. Accordingly, the reverse potentials E_l are also treated as static below.

Then the state equations for the multivariate neural variable $\{\mu\}$ neatly map onto the standard form suggested in the FEP,

$$\frac{d\mu_\alpha}{dt} = f_\alpha(V, \{\gamma_l\}, \{n_l\}) + w_\alpha(t), \tag{4.4}$$

where α takes the values $1, 2, \ldots$ with $\mu_1 = V$, $\mu_2 = \gamma_1$, $\mu_3 = \gamma_2$, and so on. The driving functions f_α are specified as

$$f_V(V, \{\gamma_l\}; \{n_l\}) = \frac{1}{C}\sum_l \gamma_l G_l(E_l - V) + \frac{1}{C} I_{ex}, \tag{4.5}$$

$$f_{\gamma_l}(V, \{\gamma_l\}; \{n_l\}) = -\frac{1}{\tau_l}(\gamma_l - \gamma_{leq}). \tag{4.6}$$

The terms w_α in equation 4.4 describe the noisy synaptic and/or leakage current w_V flowing into the neural cell rather than the deterministic contribution I_{ex} included in f_V and the noise $w_{\gamma_l} = \eta_l$ associated with the activation and inactivation of ion channels, respectively. For both noise terms, we assume the gaussian distributions $\mathcal{N}(\dot{\mu}_\alpha - f_\alpha; 0, \sigma_{w\alpha})$ with variances $\sigma_{w\alpha}$ about zero means.

For neuronal response to the sensory stimulus φ_α, we adopt the usual generative map in the FEP (see equation 2.7) as follows:

$$\varphi_\alpha = g_\alpha(V, \{\gamma_l\}, \{n_l\}) + z_\alpha, \tag{4.7}$$

where g_α is the generative map that is unknown but must be supplied for practical application and z_α characterizes the stochastic nature of the sensory reading, which we assume to have the normal distribution $\mathcal{N}(\varphi_\alpha - g_\alpha; 0, \sigma_{z\alpha})$. With the model, the neural observer responds to the sensory data instantly by means of the neuronal states. Currently, we do not possess a firm ground on the biophysical processes of the sensory prediction.

As a working example, we consider here the H-H neuron, which allows fast relaxation (i.e., $\tau_l \ll 1$) of gating variables to their steady states, $\gamma_l(t) \to \gamma_l(\infty) = \gamma_{leq}(V)$. In this case, our neural particle is fully characterized by a single dynamical variable V. Note that the time dependence of the gating variables occurs only implicitly through the long-time membrane

voltages in equation 4.5. Then the RD of our neural particle is fulfilled in a two-dimensional state space spanned by $\{\mu\} = (V, p_V) \equiv (\mu, p)$, prescribed by the Hamiltonian function, given by equation 3.9,

$$\mathcal{H}(\mu, p) = \frac{p^2}{2m_w} - \frac{1}{2}m_z(\varphi - g)^2 + pf,$$

where $m_w = \sigma_w^{-1}$ and $m_z = \sigma_w^{-1}$. While the "dissipative" function f is explicitly given in the H-H model as

$$f(\mu) = \frac{1}{C}\sum_l \gamma_{leq}(\mu)G_l(E_l - \mu) + I_{ex}/C, \tag{4.8}$$

the "conservative" function g must be additionally supplied. Moreover, one needs to make the voltage dependence of γ_{leq} available in practice. Note that the Hamiltonian is nonautonomous in general because it explicitly depends on time through both the sensory input $\varphi(t)$ and the driving current I_{ex} in f, as well as through $\sigma_w(t)$ and $\sigma_z(t)$ when the noisy data are statistically nonstationary.

Figure 3 presents the energy landscape described by the Hamiltonian function, assuming static sensory data, constant driving currents, and statistical stationarity. Since our knowledge is limited to the functional form of $g(\mu)$ and $f(\mu)$, we have taken the algebraic polynomial approximations by replacing transcendental nonlinearities in the H-H model (Wilson, 1999):

$$g(\mu) \approx a_0 + a_1\mu + a_2\mu^2,$$
$$f(\mu) \approx b_0 + b_1\mu + b_2\mu^2 + b_3\mu^3.$$

For numerical purposes, we have specified $(a_0, a_1, a_2) = (0, 1, 1)$ and $(b_0, b_1, b_2, b_3) = (0, 0.1, 1, 1)$ and assumed a static sensory input with equal masses (precisions) on the brain's internal model and belief of sensory prediction as

$$\varphi = 1.0 \quad \text{and} \quad m_w = m_z = 0.1.$$

Moreover, for simplicity, we have assumed that the input current is constant.

The Hamilton equations of motion, equations 3.23 and 3.24, yield the nonlinear RD as

$$\dot{\mu} = \Lambda_1(\mu, p; t), \tag{4.9}$$
$$\dot{p} = \Lambda_2(\mu, p; t), \tag{4.10}$$

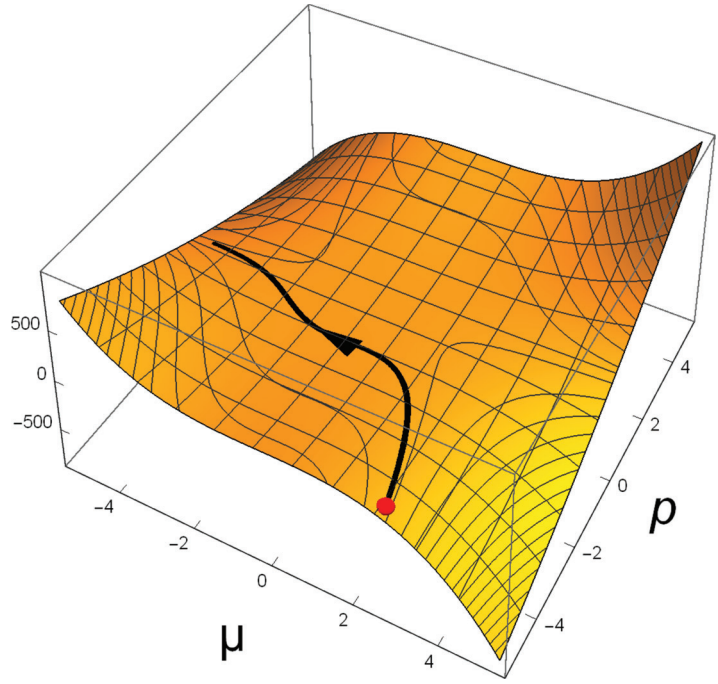

Figure 3: Hamiltonian function $\mathcal{H}(\mu, p)$ in arbitrary units for the chosen set of parameters given in the main text, where the black curve on the energy landscape is the trajectory calculated by solving the Hamilton equations of motion for an initial condition at $(\mu, p) = (2.5, -5.0)$. (For interpretation of the color in this figure, see the web version of this article.)

where the "force" functions Λ_1 and Λ_2 are expressed as

$$\Lambda_1 = f(\mu) + \frac{1}{m_w} p, \tag{4.11}$$

$$\Lambda_2 = -m_z(\varphi - g)\frac{\partial g}{\partial \mu} - \frac{\partial f}{\partial \mu} p. \tag{4.12}$$

We have chosen an initial state and solved the equations of motion to obtain the resulting trajectory. The outcome is drawn on the specified energy landscape in Figure 3. According to the model, the neural observer performs the RD, given the sensory input φ, and consequently obtains the optimal trajectory (μ^*, p^*) conforming to equation 3.25. In the long-time limit, the brain will reach a fixed (equilibrium) point (μ^*_{eq}, p^*_{eq}) in the state space, which is specified by intersections of two *isoclines*,

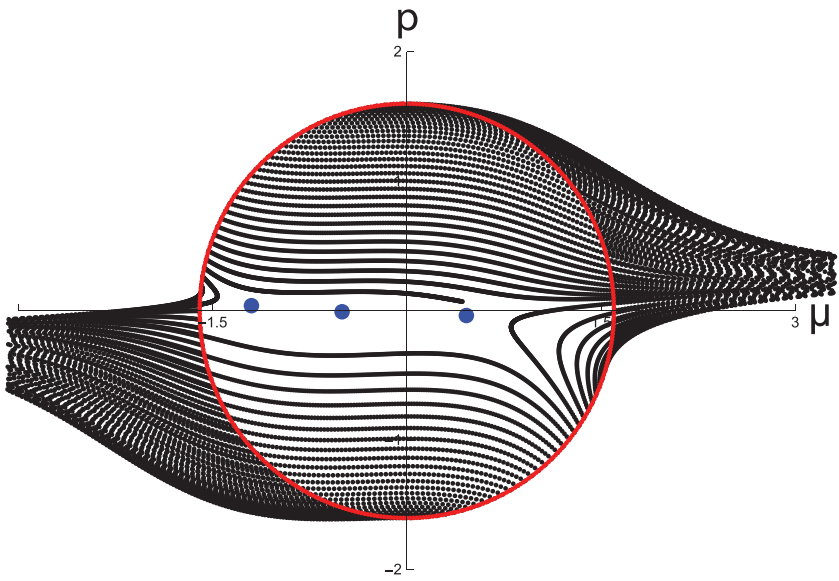

Figure 4: Optimal trajectories in phase space, which are obtained by integrating the RD, given by equations 4.9 and 4.10, from the initial conditions arbitrarily chosen on the red circle. Here the blue dots are fixed points among which only the middle dot, at $(-0.50, -0.01)$, is a stable fixed point; the other two points are saddle points. The stable fixed point turns out to be a center, which we have confirmed numerically and by linear stability analysis. (For interpretation of the color in this figure, see the web version of this article.)

$$\Lambda_i(\mu, p; \infty) = 0, \quad i = 1, 2.$$

We determined the fixed points numerically and found that there exist three real solutions for the specified system parameters, $(-1.23, 0.05)$, $(-0.50, -0.01)$, and $(0.07, -0.04)$, which are depicted as blue dots in Figure 4. Through further analysis, we have found that only the middle point is a stable equilibrium solution, while the other two are saddle points. Figure 4 shows a flow of trajectories obtained from arbitrary initial points on the red-colored circle of radius $\mu^2 + p^2 = 1.6$ in phase space.

To gain insight into how the system approaches a steady state, we inspect the optimal trajectories near an equilibrium point:

$$\mu^* \approx \mu^*_{eq} + \delta\mu^* \quad \text{and} \quad p^* \approx p^*_{eq} + \delta p^*.$$

We expand equations 4.9 and 4.10 to the linear order in the deviations $\delta\mu^*$ and δp^* and, after rearrangement, obtain the normal form,

$$\frac{d}{dt}\begin{pmatrix}\delta\mu^*\\ \delta p^*\end{pmatrix}+\begin{pmatrix}\mathcal{R}_{11} & \mathcal{R}_{12}\\ \mathcal{R}_{21} & \mathcal{R}_{22}\end{pmatrix}\begin{pmatrix}\delta\mu^*\\ \delta p^*\end{pmatrix}=0. \tag{4.13}$$

In equation 4.13, the elements of the relaxation (Jacobian) matrix $\mathcal{R}$ are expressed as

$$\mathcal{R}_{11}=-\left[\frac{\partial f}{\partial\mu}\right]_{eq},\quad \mathcal{R}_{12}=-\frac{1}{m_w}$$

$$\mathcal{R}_{21}=\sigma_z^{-1}\left[-\left(\frac{\partial g}{\partial\mu}\right)^2+(\varphi-g)\frac{\partial^2 g}{\partial\mu^2}-\frac{\partial^2 f}{\partial\mu^2}p\right]_{eq},$$

$$\mathcal{R}_{22}=\left[\frac{\partial f}{\partial\mu}\right]_{eq},$$

where the partial derivatives are to be evaluated at the equilibrium points. Here, for notational convenience, we denote the column vector as

$$\delta\psi\equiv\begin{pmatrix}\delta\mu^*\\ \delta p^*\end{pmatrix}.$$

Then, the formal solution to equation 4.13 is expressed as

$$\delta\psi(t)=e^{-\mathcal{R}t}\delta\psi(0).$$

One may expand the initial state $\psi(0)$ in terms of the eigenvectors of $\mathcal{R}$ as

$$\delta\psi(0)=\sum c_\alpha\phi_\alpha,$$

where the eigenvalues λ_α and eigenvectors ϕ_α are determined by the secular equation,

$$\mathcal{R}\phi_\alpha=\lambda_\alpha\phi_\alpha.$$

Consequently, the solution to the linear RD at a single node level is completed as

$$\delta\psi(t)=\sum_{\alpha=1}^{2}c_\alpha e^{-\lambda_\alpha t}\phi_\alpha, \tag{4.14}$$

where the expansion coefficients c_α are fixed by the initial condition.

In the linear regime, a geometrical interpretation of the equilibrium solutions is possible by inspecting the eigenvalues of the Jacobian matrix $\mathcal{R}$.

Considering that the matrix $\mathcal{R}$ is not symmetric, we anticipate that the eigenvalues are not real. Furthermore, because the trace of the relaxation matrix equals zero, the sum of the two eigenvalues must be zero. Thus, when the determinant of $\mathcal{R}$ is positive, the two eigenvalues λ_1 and λ_2 would be purely imaginary with opposite signs. Consequently, in our particular model, the resulting equilibrium point is likely to be a center. We have confirmed numerically that the eigenvalues of the Jacobian corresponding to the stable equilibrium point in Figure 4 are $\pm 1.6i$, specifying a center.

4.2 The Hierarchical Neural Network. Here, we suppose that a finite number of levels exist in the perceptual hierarchy of the whole system and that for simplicity, each level is characterized efficiently as a single neural node. Further, we assume that the neural node at hierarchical level i is described by the coarse-grained activation and connection variables, denoted as $V^{(i)}$ and $S^{(i)}$, respectively. The activation variable describes the action potential at a node, and the connection variable describes interlevel synaptic input and output variables. Both variables are derived from a population of neurons and thus vary on a coarse-grained space and timescale. The technical details of how one may derive such a coarse-grained description are out of our scope (see Deco et al., 2008, for a reference). These variables form the coordinates in the brain's configurational space,

$$\mu^{(i)} = (V^{(i)}, S^{(i)}),$$

where the superscript $i = 1, 2, \ldots, M$, with M denoting the highest level.

We assume that the activation variables $V^{(i)}$ obey the effective dynamics with noise $w^{(i)}$ within each hierarchical level i,

$$\frac{dV^{(i)}}{dt} = f^{(i)}(V^{(i)}, S^{(i)}) + w^{(i)}, \tag{4.15}$$

which is a direct generalization of equation 4.4 with the incorporation of the hierarchical dependence via $S^{(i)}$. For interlevel dynamics, we propose that the connection variables are updated by a one-level-higher connection as well as activation variables, subjected to the stochastic equations

$$\frac{dS^{(i)}}{dt} = g^{(i+1)}(V^{(i+1)}, S^{(i+1)}) + z^{(i)}, \tag{4.16}$$

where $z^{(i)}$ represents the noise associated with the process. The brain's top-down prediction functions $f^{(i)}$ and $g^{(i)}$ must be supplied in practical implementation. Note that only spontaneous fluctuation occurs at the top cortical level, $i = M$; accordingly,

$$g^{(M+1)} = 0. \tag{4.17}$$

Furthermore, we enforce the constraint that the sensory data φ enter the interface of (or boundary between) the brain and environment, which is specified as the lowest hierarchical level, $i = 1$. Subsequently we assume that the brain's prediction of the sensory inputs is performed by way of an instantaneous mapping,

$$S^{(0)} = g^{(1)}(V^{(1)}, S^{(1)}) + z^{(0)}, \tag{4.18}$$

where, for notational convenience, we have set

$$S^{(0)} \equiv \varphi(t).$$

We remark that the hierarchical equations we propose, equation 4.16, are dissimilar to the conventional formulation, which assumes a static nonlinearity in the entire hierarchy like the one in equation 4.18 at the sensory interface (see Buckley et al., 2017). One may ensure that the time constants of equation 4.16 are sufficiently fast to approximate a static nonlinearity. Here, we treat the connection variables dynamically, rather than statically, to treat lateral and hierarchical dynamics symmetrically. The rates of the activation and connection variables may be subjected to different timescales that can be incorporated, for instance, by introducing distinctive relaxation times in their generative functions. It turns out that our equations suit the formalism of the Hamilton action principle neatly.

Having specified our hierarchical model, we express the informational Lagrangian for the constructed neural network by generalizing equation 3.18 with a single sensory input for now as

$$\mathcal{L}(V, \dot{V}; S, \dot{S}; \varphi) = \frac{1}{2}\sum_{i=1}^{M} m_w^{(i)} \left(\varepsilon_w^{(i)}\right)^2 + \frac{1}{2}\sum_{i=0}^{M} m_z^{(i)} \left(\varepsilon_z^{(i)}\right)^2, \tag{4.19}$$

where $m_w^{(i)}$ and $m_z^{(i)}$ are the inertial masses (precisions) associated with the gaussian noises, $w^{(i)}$ and $z^{(i)}$, respectively, and are defined as

$$m_w^{(i)} \equiv 1/\sigma_w^{(i)} \quad \text{and} \quad m_z^{(i)} \equiv 1/\sigma_z^{(i)}. \tag{4.20}$$

The auxiliary variables in the Lagrangian are defined as ($i \geq 1$)

$$\varepsilon_w^{(i)} \equiv \dot{V}^{(i)} - f^{(i)}\left(V^{(i)}, S^{(i)}\right), \tag{4.21}$$

$$\varepsilon_z^{(i)} \equiv \dot{S}^{(i)} - g^{(i+1)}\left(V^{(i+1)}, S^{(i+1)}\right). \tag{4.22}$$

We interpret that $\varepsilon_w^{(i)}$ specifies the discrepancy between the change in the present lateral state and the brain's on-level prediction, which may be considered as the lateral prediction error. On the other hand, $\varepsilon_z^{(i)}$ measures the prediction error between the change in the present hierarchical state and its prediction from one higher level via the generative map g, which may be viewed as the hierarchical prediction error. Note that $\varepsilon_z^{(0)}$ in the second term on the right-hand side of equation 4.19 is defined separately as

$$\varepsilon_z^{(0)} \equiv S^{(0)} - g^{(1)}\left(V^{(1)}, S^{(1)}\right),$$

which specifies an error estimation in sensory prediction at the lowest hierarchical level.

The canonical momenta, conjugate to $V^{(i)}$ and $S^{(i)}$, are readily calculated for $i \geq 1$, respectively, as

$$p_V^{(i)} \equiv \frac{\partial \mathcal{L}}{\partial \dot{V}^{(i)}} = m_w^{(i)} \varepsilon_w^{(i)}, \tag{4.23}$$

$$p_S^{(i)} \equiv \frac{\partial \mathcal{L}}{\partial \dot{S}^{(i)}} = m_z^{(i)} \varepsilon_z^{(i)}. \tag{4.24}$$

Note that the informational masses $m_w^{(i)}$ and $m_z^{(i)}$ are the precisions (see the discussion below equation 3.5). The role of the inertial masses is to modulate the discrepancy between the change of the perceptual states and their prediction. Thus, in our theory, the momentum $p_V^{(i)}$ is a measure of lateral prediction error modulated by inertial mass $m_w^{(i)}$, and the momentum $p_S^{(i)}$ is a measure of hierarchical prediction error modulated by inertial mass $m_z^{(i)}$. The precision is higher for a greater mass. In predictive coding formulations of the FEP, this modulated prediction error is known as the precision-weighted prediction error.

Given the Lagrangian in equation 4.16, we can formulate the informational Hamiltonian by performing a Legendre transformation:

$$\mathcal{H} = \sum_i \left(\dot{V}^{(i)} p_V^{(i)} + \dot{S}^{(i)} p_S^{(i)}\right) - \mathcal{L}.$$

After some manipulation, we obtain

$$\mathcal{H}(V, p_V; S, p_S; \varphi) = \sum_{i=1}^{M} \left(\mathcal{T}^{(i)} + \mathcal{V}^{(i)}\right), \tag{4.25}$$

where the informational kinetic energy is defined as ($i \geq 1$)

$$\mathcal{T}^{(i)}(p_V, p_S) = \frac{1}{2m_w^{(i)}} \left(p_V^{(i)}\right)^2 + \frac{1}{2m_z^{(i)}} \left(p_S^{(i)}\right)^2 \tag{4.26}$$

and the potential energy as ($i \geq 2$)

$$\mathcal{V}^{(i)}(V, p_V; S, p_S; \varphi) \equiv p_V^{(i)} f^{(i)} + p_S^{(i)} g^{(i+1)}. \tag{4.27}$$

Note that the potential energy at the lowest level is specified separately as

$$\mathcal{V}^{(1)} = p_V^{(1)} f^{(1)} - \frac{1}{2m_z^{(0)}} \left(p_S^{(0)}\right)^2, \tag{4.28}$$

where, for notational convenience, we have expressed the precision-weighted prediction error associated with the sensory measurement as

$$p_S^{(0)} \equiv m_z^{(0)} \varepsilon_z^{(0)} = m_z^{(0)} (\varphi - g^{(1)}),$$

which, unlike $p_S^{(i)}$ for $i \geq 1$, is not a canonical momentum. Consequently, the multilevel Hamiltonian in equation 4.25 has been prescribed via the perceptual states in the hierarchical chain, $i = 1, 2, \ldots, M$, denoted as a four-dimensional column vector $\psi^{(i)}$ at each level,

$$\psi^{(i)} = (V^{(i)}, p_V^{(i)}, S^{(i)}, p_S^{(i)})^{\mathrm{T}} \equiv (\psi_1^{(i)}, \psi_2^{(i)}, \psi_3^{(i)}, \psi_4^{(i)})^{\mathrm{T}},$$

where T indicates the transpose operation.

Next, it is straightforward to generate the Hamiltonian equations of motion for the brain's perceptual states $\psi^{(i)}$. The results are the coupled differential equations for the four computational components at each level ($i \geq 1$), which are, in turn, hierarchically connected among adjacent levels:

$$\dot{V}^{(i)} = \frac{\partial \mathcal{H}}{\partial p_V^{(i)}} = \frac{1}{m_w^{(i)}} p_V^{(i)} + f^{(i)}, \tag{4.29}$$

$$\dot{S}^{(i)} = \frac{\partial \mathcal{H}}{\partial p_S^{(i)}} = \frac{1}{m_z^{(i)}} p_S^{(i)} + g^{(i+1)}, \tag{4.30}$$

$$\dot{p}_V^{(i)} = -\frac{\partial \mathcal{H}}{\partial V^{(i)}} = -\frac{\partial f^{(i)}}{\partial V^{(i)}} p_V^{(i)} - \frac{\partial g^{(i)}}{\partial V^{(i)}} p_S^{(i-1)}, \tag{4.31}$$

$$\dot{p}_S^{(i)} = -\frac{\partial \mathcal{H}}{\partial S^{(i)}} = -\frac{\partial f^{(i)}}{\partial S^{(i)}} p_V^{(i)} - \frac{\partial g^{(i)}}{\partial S^{(i)}} p_S^{(i-1)}. \tag{4.32}$$

According to the derived RD, the sensory inputs φ enter the brain-environment interface at the level $j = 1$, and are instantly predicted by the organism's lowest-level generative model $g^{(1)}(V^{(1)}, S^{(1)})$. Subsequently, the resulting prediction error $p_S^{(0)}$ acts as a source to update the prediction errors $p_V^{(1)}$ and $p_S^{(1)}$. The changes of on-level perceptual states $V^{(1)}$ and $S^{(1)}$ are predicted by the generative models $f^{(1)}$ and $g^{(2)}$ with additional modulations from the perceptual momenta $p_V^{(1)}$ and $p_S^{(1)}$, which are the lateral and hierarchical prediction errors, respectively. At higher levels $i \geq 2$, the intralevel dynamics of the activation state $V^{(i)}$ is updated through equation 4.29 by the on-level generative function $f^{(i)}$ and prediction error $p_V^{(i)}$, while the change of the current hierarchical state $S^{(i)}$ is determined through equation 4.30 by the interlevel prediction $g^{(i+1)}$ and the on-level prediction error $p_S^{(i)}$. The organism's top-down message flow is mediated by the connection state $S^{(i)}$ via equation 4.30 as $(S^{(i+1)}, V^{(i+1)}) \rightarrow S^{(i)}$. Furthermore, equations 4.31 and 4.32 govern the coupled, bottom-up propagation of the prediction errors, mediated by $p_S^{(i)}$, $p_S^{(i)} \rightarrow (p_S^{(i+1)}, p_V^{(i+1)})$. Figure 5 schematically illustrates the perceptual architecture of the hierarchical network at the lowest two levels, implied by equation 4.29 to 4.32. It shows the top-down prediction of the sensory inputs φ at the lowest level and the bottom-up propagation of the prediction errors $p_S^{(0)}$.

Here, we emphasize that the dynamics of precision-weighted prediction errors, encapsulated in canonical momenta in which mass takes over the role of precision, are taken into account in our Hamiltonian formulation on an equal footing with the dynamics of prediction of the state variables. This aspect is also in contrast to the conventional minimization algorithm, which entails differential equations only for the update of the brain states without carrying parallel ones for the prediction errors. Consequently, the message passing in our model shows different features compared with the neural circuitry from the conventional RD (Bastos et al., 2012). However, the general message flow, in terms of the computational units, of feedforward, feedback, and lateral connections remains the same in the hierarchical brain network. An attempt to incorporate the brain's computation of prediction errors in the FEP can be found in a recent tutorial model (Bogacz, 2017).

Here, for mathematical compactness, we rewrite the filtering algorithm, equations 4.29 to 4.32, as

$$\frac{d\psi_\alpha^{(i)}}{dt} = \Lambda_\alpha^{(i)}(\{\psi_\alpha^{(i)}\}), \tag{4.33}$$

where the hierarchical index i runs from 1 to M, α runs from 1 to 4, and the force function $\Lambda_\alpha^{(i)}$ is the corresponding right-hand side to each vector component $\psi_\alpha^{(i)}$ at cortical level i. The obtained hierarchical equations are the

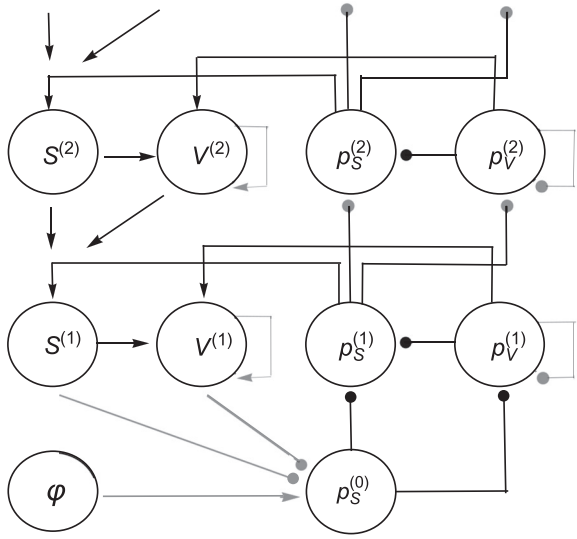

Figure 5: A schematic of the neural circuitry that conducts the RD given by equations 4.29 to 4.32 in the hierarchical network of the brain, where the computational units $(S^{(i)}, V^{(i)}, p_S^{(i)}, p_V^{(i)})$, $i = 1, 2, \ldots, M$, are connected by arrows for excitatory (positive) inputs and by lines ended with filled dots for inhibitory (negative) inputs. Note that the prediction error $p_S^{(0)}$ of incoming sensory data φ, at the lowest level, induces an inhibitory change in the perceptual momenta $(p_S^{(1)}, p_V^{(1)})$. Subsequently, the prediction error propagates up in the hierarchy, $p_S^{(1)} \to (p_S^{(2)}, p_V^{(2)})$, and so on. The top-down message passing is mediated by means of the connection states $S^{(i)}$. For instance, the connection state $S^{(1)}$ is top-down predicted by both units $(S^{(2)}, V^{(2)})$ from one level higher.

highlight of our theory, prescribing the RD of the brain's sensory inference under the FEP framework.

To apply our formulation to an empirical brain, one needs to supply the generating function of lateral dynamics $f^{(i)}$ and the hierarchical connecting function $g^{(i)}$, which enter the force functions $\Lambda_\alpha^{(i)}$ in the perceptual mechanics given by equation 4.33. For the generating function, we again use the H-H model in equation 4.5 to write

$$f^{(i)}(V^{(i)}, S^{(i)}) = \sum_l \gamma_{leq} \tilde{G}_l \left(E_l - V^{(i)}\right) + \tilde{G}_S S^{(i)} \left(E_S - V^{(i)}\right), \tag{4.34}$$

where $\tilde{G}_l$ are the channel conductances normalized by the capacitance C. Moreover, the second term on the right-hand side accounts for other deterministic driving sources such as leakage or lateral synaptic currents, with

$\tilde{G}_S$ being the normalized synaptic conductance. The hierarchical connection function, for which we have limited biophysical knowledge, shall be taken in a simple form here as

$$g^{(i)}(V^{(i)}, S^{(i)}) = \Gamma(V^{(i)})S^{(i)}, \tag{4.35}$$

where the function Γ denotes the voltage-dependent synaptic plasticity from hierarchical level i to level $i-1$. In addition, as in the single-node case, one must supply approximate models for the voltage dependence of the gating variables γ_{leq} and the connection strength Γ. For instance, one may take the quadratic approximations (Wilson, 1999)

$$\gamma_{leq}(V^{(i)}) \approx b_{l0} + b_{l1}V^{(i)} + b_{l2}V^{(i)}V^{(i)},$$
$$\Gamma(V^{(i)}) \approx a_0 + a_1V^{(i)} + a_2V^{(i)}V^{(i)}.$$

Having laid down the lateral and hierarchical generative models, the organism's brain can now perform the RD given a stream of noisy inputs. While conducting the filtering, an optimal trajectory is obtained in multidimensional phase space,

$$\psi_\alpha^{*(i)} = \psi_\alpha^{*(i)}(t),$$

which, in the end, tends to a fixed point, $\psi_{\alpha,eq}^{(i)} = \psi_\alpha^{*(i)}(t \to \infty)$. The necessary equilibrium condition for equation 4.33 is

$$\Lambda_\alpha^{(i)}(\{\psi_\alpha^{(i)}\}) = 0. \tag{4.36}$$

Although the full time-dependent solutions must be invoked numerically, one may inspect the perceptual trajectories near a fixed point by linear analysis. To this end, we consider a small deviation of the αth component of the perceptual state vector $\psi_\alpha^{*(i)}$ at the cortical level i, $\delta\psi_\alpha^{(i)}$, from the fixed point $\psi_{\alpha,eq}^{(i)}$:

$$\psi_\alpha^{*(i)} \approx \psi_{\alpha,eq}^{(i)} + \delta\psi_\alpha^{(i)}.$$

Then, we expand equation 4.33 about the fixed point to the linear order in the small deviation, and after some manipulation, we obtain the hierarchical equations for $\delta\psi_\alpha^{(i)}$,

$$\frac{d\delta\psi_\alpha^{(i)}}{dt} + \sum_{\beta=1}^{4} \mathcal{R}_{\alpha\beta}^{(i)}\delta\psi_\beta^{(i)} = \sum_{\beta=1}^{4}\sum_{j\neq i}^{M} \Phi_{\alpha\beta}^{(ij)}\delta\psi_\beta^{(j)}, \tag{4.37}$$

where the $\alpha\beta$ component of the 4×4 Jacobian matrix at cortical level i is given by

$$\mathcal{R}^{(i)}_{\alpha\beta} = \left[\frac{\partial \Lambda^{(i)}_{\alpha}}{\partial \psi^{(i)}_{\beta}}\right]_{eq}.$$

The interlevel connection between levels i and j in the hierarchical pathway is given by

$$\Phi^{(ij)}_{\alpha\beta} = \left[\frac{\partial \Lambda^{(i)}_{\alpha}}{\partial \psi^{(j)}_{\beta}}\right]_{eq},$$

where the subscript eq indicates that the matrix elements are to be evaluated at the equilibrium points. To cast the inhomogeneous term into a more suggestive form, we further inspect it in detail within the models specified. We observe first that the matrix elements $\Phi^{(ij)}_{\alpha\beta}$ do not vanish only for $\alpha = 3$ because only the force function $\Lambda^{(i)}_{3}$ possesses $\psi^{(j)}_{\beta}$ for $j \neq i$ as variables via $g^{(i+1)}$ (see equation 4.30). Second, because $g^{(i+1)}$ depends solely on the hierarchical-level index $i+1$, only matrix elements with the hierarchical index $j = i+1$ survive. Combining these two observations, the source term on the right-hand side of equation 4.37 is converted into a vector at level $i+1$ with only a single nonvanishing $\alpha = 3$ component,

$$\sum_{\beta=1}^{4}\sum_{j\neq i}^{M} \Phi^{(ij)}_{\alpha\beta}\delta\psi^{(j)}_{\beta} \equiv \delta\zeta^{(i+1)}_{\alpha},$$

which, for completeness, we spell out explicitly as

$$\delta\zeta^{(i+1)}_{\alpha} = \delta_{\alpha 3}\left\{\left[\frac{\partial g^{(i+1)}}{\partial \psi^{(i+1)}_{1}}\right]_{eq}\delta\psi^{(i+1)}_{1} + \left[\frac{\partial g^{(i+1)}}{\partial \psi^{(i+1)}_{3}}\right]_{eq}\delta\psi^{(i+1)}_{3}\right\}, \tag{4.38}$$

where $\delta_{\alpha 3}$ is the Kronecker delta.

Finally, we present a formal solution to the linearized perceptual mechanics given by equation 4.37, which can be obtained by a direct integration with respect to time. The result takes the form

$$\delta\psi^{(i)}(t) = e^{-\mathcal{R}^{(i)}t}\delta\psi^{(i)}(0) + \int_{0}^{t} dt' e^{-\mathcal{R}^{(i)}(t-t')}\delta\zeta^{(i+1)}(t'). \tag{4.39}$$

We next solve the eigenvalue problem at each hierarchical level, which is defined as

$$\mathcal{R}^{(i)}\phi_\alpha^{(i)} = \lambda_\alpha^{(i)}\phi_\alpha^{(i)}, \tag{4.40}$$

where $\lambda_\alpha^{(i)}$ and $\phi_\alpha^{(i)}$ are the eigenvalues and corresponding eigenvectors at level i, respectively. Then we expand the initial state $\delta\psi^{(i)}(0)$ in terms of the complete eigenvectors:

$$\delta\psi^{(i)}(0) = \sum a_\alpha^{(i)}\phi_\alpha^{(i)}. \tag{4.41}$$

Similarly, we may expand the inhomogeneous vector $\delta\zeta^{(i+1)}$ as

$$\delta\zeta^{(i+1)}(t') = \sum b_\alpha^{(i+1)}(t')\phi_\alpha^{(i)}, \tag{4.42}$$

where the expansion coefficients $b_\alpha^{(i+1)}$ are time dependent. By substituting the expansions equations 4.41 and 4.42 into equation 4.39, we obtain the desired formal solution near equilibrium points:

$$\delta\psi^{(i)}(t) = \sum_{\alpha=1}^{4} a_\alpha e^{-\lambda_\alpha^{(i)}t}\phi_\alpha^{(i)} + \sum_{\alpha=1}^{4}\phi_\alpha^{(i)}\int_0^t dt' e^{-\lambda_\alpha^{(i)}(t-t')}b_\alpha^{(i+1)}(t'). \tag{4.43}$$

The geometrical approach to a fixed point is again determined by the eigenvalues $\lambda_\alpha^{(i)}$; however, the details are driven by the time-dependent generative sources $b_\alpha^{(i+1)}(t)$ from one level higher in the hierarchy. An application of equation 4.43 would be to determine the natural frequencies of predictions and prediction errors. As the solution approaches an attractor, these correspond to the imaginary parts of the principal eigenvalues above. This application is potentially very interesting because there are characteristic frequencies associated with message passing in the brain (Bastos et al., 2015). The precise relevance of this formulation to the asymmetric frequency dependence must be further explored.

To summarize, responding to a sensory stream $\varphi = S^{(0)}$, at the lowest hierarchical level ($i = 1$), the brain, which is initially in a resting state, performs the hierarchical RD by integrating equation 4.33 to infer the external causes. The ensuing brain's computation corresponds to the minimization of the IA, which is an upper bound of the sensory uncertainty. Its mathematical statement, equation 2.3, is repeated compactly as

$$H[p(\varphi)] \leq \mathcal{S}[F;\varphi],$$

where the sensory uncertainty H was defined in equation 2.1 and the IA on the right-hand side is expressed here in terms of the hierarchical states as $\mathcal{S}[F;\varphi] = \int dtF(\{\psi_\alpha^{(i)}\};\varphi)$. Conforming to the FEP, the minimum value of IA specifies the tightest bound of the sensory uncertainty over a relevant biological timescale, which preserves the organism's current model of the environment.

5 Discussion

We have recast the FEP following the principles of mechanics, which state that all living organisms are evolutionally self-organized to tend to minimize the sensory uncertainty about environmental encounters. The sensory uncertainty is an average of the surprisal over the sensory density registered on the brain-environment interface, which is the self-information contained in the sensory probability density. The FEP suggests that the organisms implement the minimization by calling forth the IFE in the brain. The time integral of the IFE gives an estimate of the upper bound of the sensory uncertainty. We have enunciated that the minimization of the IFE must continually occur over a finite temporal horizon of an organism's unfolding environmental event. Our scheme is a generalization of the conventional theory, which approximates the minimization of the IFE at each point in time when it performs the gradient descent. Note that the sensory uncertainty is an information-theoretical Shannon entropy (Shannon, 1948); however, in this work, we avoided using the term *entropy* because "minimization of the sensory entropy" is reminiscent of Erwin Schrödinger's thermodynamic term, *negative entropy*, which carries a disputable connotation implying how the living organism avoids decay (Schrödinger, 1967). The nerve cell and the brain are open systems, the physical entropy of which can increase or decrease depending on the direction of heat flow. According to fluctuation theorems (see Crooks, 1999; Evans & Searles, 2002; Seifert, 2005, for instance), under nonequilibrium conditions, it is reasonable to anticipate a statistical deviation from the second law of thermodynamics even in finite systems for a finite time. The biological FEP postulates that the organism's adaptive fitness corresponds to the minimization of the sensory uncertainty, which is the average surprisal. The average is required because the sensory organs are not small or mesoscopic systems, and the perceptual and active inferences are phenomena occurring in the macroscopic brain. Therefore, from the perspective of the second law, the sensory-uncertainty minimization must contribute to the total entropy of the brain and its environment as a whole. Note, however, that the IFE we work with is an information-theoretic construct rather than a physical quantity. Currently, we do not have a theory to formulate the physical FE for the brain.

We have adopted the Laplace-encoded IFE as an IL in implementing the FEP under the variational Hamilton principle. Further, by subscribing to the

standard Newtonian dynamics, we have considered the IFE to be a function of position and velocity as metaphors for the organism's brain variable and their first-order time derivative, respectively. According to Newton's second law, the brain's perceptual state, specified by the position and velocity of the brain variables, changes by an applied force; for example, an exogenous sensory perturbation is the cause of the rate of change of velocity or acceleration, which is the second-order time derivative of position. The brain variable maps onto the first-order sufficient statistics of the R-density engaged in the organism's brain to perform the RD, which is the Bayesian filtering of the noisy sensory data. In the ensuing Hamiltonian formulation, the RD prescribes momentum, conjugate to position, as a mechanical measure of prediction error weighted by inertial mass, which is the precision. We have eschewed the use of generalized coordinates of motion, which is introduced in the prevailing theory to specify the extended states of higher orders of motion. Consequently, the conceptual subtlety of assigning the causes to higher-order motions beyond acceleration has been dismissed. Furthermore, the arbitrariness involved in deciding the number of generalized coordinates for a complete description and the ambiguity in specifying unknowable initial conditions have been averted. Consequently, the RD tenably underpins the causality: for specified initial conditions for the perceptual positions and corresponding momenta, the RD can be integrated continuously online in response to sensory inputs.

The features of the changing world enter our theory via time-dependent sensory inputs, which affect the brain states in continuous time. The temporal correlation of the dynamical states may be incorporated as time-dependent covariances; however, these are not explored in this work. Moreover, in our theory, all the parameters in the RD are specified in the Hamiltonian; thus, no extra parameters such as learning rates in the gradient-descent scheme are required to control the speed of convergence to a steady state. In effect, the learning rate is formally identical to the informational mass or precision. In other words, the learning rates are implicit in the FEP, which is already optimal in the sense of approximate Bayesian inference. According to our formulation, the brain's Helmholtzian perception corresponds to finding an optimal trajectory in the hierarchical functional network by minimizing the IA. When the brain completes the RD by reaching a desired fixed point or an attractor, it remains resting (i.e., spontaneous) until another sensory stimulus enters.

We have admitted the top-down rationale of sensory prediction in our formalism, an essential facet of the FEP. As usual, the sensory inputs at the interface, which is the lowest hierarchical level, were assumed to be instantaneously mapped to the organism's belief with associated noises. In contrast, however, at higher levels, we have generalized that the interlevel filtering in the brain's functional hierarchy obeys the stochastic dynamics, supplied with the organism's dynamical generative model of environmental states. The resulting RD notably incorporates the dynamics of both

predictions and prediction errors of the uncertain sensory data on the same footing in the computational architecture. Consequently, the details of the ensuing neural circuitry from our formulation differ from that supported by the gradient-descent scheme, which generates only the dynamics of prediction of the causal and hidden states, not their prediction errors. Our formulation provides a natural account of the general structure of asymmetric message passing, namely, descending predictions and ascending prediction errors, in the brain's hierarchical architecture.

To show how our formulation may be implemented in the biophysical brain, we have employed the H-H-type neuronal dynamics at a single-cell level and subsequently constructed the large-scale perceptual circuitry. We have chosen the conductance-based model, which is complex but experimentally grounded (Koch, 1999; Hille, 2001), instead of more efficient spiking models such as integrate-and-fire or firing-rate models (Dayan & Abbott, 2001; Izhikevich, 2003; Burkitt, 2006). The reason was that while the H-H dynamics delivers an autonomous trajectory, the integrate-and-fire models bear an abrupt dynamical interruption involved in setting spike firing at a threshold and resetting voltage to a resting value; in other words, the spike generation itself is not part of the dynamical development. Moreover, the firing-rate models describe average dynamics over many trials rather than single-neuron dynamics; therefore, they neglect the detailed time course of the action potential. To derive the RD within the framework of the FEP, the IA must be minimized continuously with respect to trajectories, which requires an implicit (autonomous) time dependence of the IFE through its arguments, that is, the dynamical variables. Furthermore, the spike-sorting problem from raw extracellular recordings is still a challenging problem (Einevoll, Franke, Hagen, Pouzat, & Harris, 2012). For the working example in this article, we have assumed that the gating kinetics relaxed quickly to a steady state and the ion concentrations stayed in electrochemical equilibrium. Consequently, we considered the state equation for single neurons on a timescale in which only the change in the membrane potential mattered, and the details of firing rate, axonal propagation, and dendritic time lags were ignored in the computational description.

Finally, the underlying mechanism for learning in the brain was not considered explicitly in our biophysical implementation, unlike the common firing-rate models of network neurons. In the latter, the coupling mechanism between the presynaptic and postsynaptic rates, via phenomenological synaptic weights, facilitates Hebbian plasticity for learning (Abbott, 1994; Martin, Grimwood, & Morris, 2000). In our formulation, the synaptic efficacy at a neuronal level can be incorporated by considering a synaptic input current in the driving function (see equation 4.8), which would influence the postsynaptic output, equations 4.9 and 4.10. Similarly, to implement the synaptic plasticity at the network level, one can add the synaptic driving terms in the intra- and interlevel generative functions in equations 4.15 and 4.16 and minimize the IA to obtain the RD. The general structure of

the outcome will appear the same as the neural circuitry presented in Figure 5. The positions—activation and connection variables—and their corresponding momenta in the brain circuitry may map onto the representational and error units, respectively, among functional populations of neurons in the cortex (Summerfield & Egner, 2009). It turns out that the two functional populations in Figure 5 do not follow Dale's law, because they have neural units with both excitatory and inhibitory outputs (Dayan & Abbott, 2001; Okun & Lampl, 2008). In the conventional spiking models, the network dynamics is put in place by writing down coupled equations obeying Dale's law for the two biophysically distinct classes of neurons (see Aitchison & Lengyel, 2016, for instance). This leaves a challenge in the framework of the FEP for rendering the RD, which operates on the functional neural units, to reconcile with Dale's law for the biophysical neurons. Furthermore, the synaptic gain may be formulated effectively in the present theory by taking into account the statistical nonstationarity of the fluctuations (MacDonald, 2006) involved in the state equations. The statistical nonstationarity sets up an extra timescale over which the precisions are transient (see equation 4.20) and slower than that associated with the change of the state variables. Accordingly, one may treat the time-dependent precision as an independent dynamic variable in the Lagrangian, prescribing gain, and generate Hamilton's equation of motion for the gain variables, thereby generalizing the RD. Consequently, the generalized RD can deliver the gain control over model learning in the extended state space comprising not only brain variables and momenta but also gain variables and their partner momenta. The work is in progress and will be reported elsewhere.

In short, we are still a long way from understanding how the Bayesian FEP in neurosciences may be made congruous with the biophysical reality of the brain. It is far from clear how the organism embodies the generative model of the environment in the physical brain. Our theory delivers only a hybrid model of the biologically plausible information-theoretic framework of the FEP and the mechanical formulation of the RD under the principle of least action. To quote Hopfield (1999), "It lies somewhere between a model of neurobiology and a metaphor for how the brain computes." We hope that our effort will guide a step forward for solving the challenging problem.

Acknowledgments

I thank anonymous reviewers for providing invaluable comments and suggestions to improve this article.

References

Abbott, L. F. (1994). Decoding neural firing and modeling neural networks. *Quarterly Review of Biophysics*, *27*, 291–331.

Aitchison, L., & Lengyel, M. (2016). The Hamiltonian brain: Efficient probabilistic inference with excitatory-inhibitory neural circuit dynamics. *PLoS Computational Biology*, *12*(12), e1005186.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*, 695–711.

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., . . . Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, *85*, 390–401.

Berkes, P., Orban, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, *331*, 83–87.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, *76*(B), 198–211.

Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, *81*, 55–79.

Buckley, C. L., & Toyoizumi, T. (2018). A theory of how active behavior stabilises neural activity: Neural gain modulation by closed-loop environmental feedback. *PLoS Computational Biology*, *14*(1), e1005296.

Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological Cybernetics*, *95*, 1–19.

Crooks, G. E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy difference. *Physical Review E*, *60*, 2721–2726.

David, O., & Friston, K. J. (2003). A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage*, *20*, 1743–1755.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience.* Cambridge, MA: MIT Press.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*, 889–904.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Computational Biology*, *4*, e1000092.

Einevoll, G. T., Franke, F., Hagen, E., Pouzat, C., & Harris, K. I. (2012). Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Current Opinion in Neurobiology*, *22*, 11–17.

Einevoll, G. T., Kayser, C., Logothetis, N. K., & Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, *14*, 770–785.

Evans, D. J., & Searles, D. J. (2002). The fluctuation theorem. *Advances in Physics*, *51*, 1529–1585.

Fiorillo, C. D. (2008). Towards a general theory of neural computation based on prediction by single neurons. *PLoS One*, *3*(10), e3298.

Fiorillo, C. D., Kim, J. K., & Hong, S. Z. (2014). The meaning of spikes from the neuron's point of view: Predictive homeostasis generates the appearance of randomness. *Frontiers in Computational Neuroscience*, *8*, 49.

Friston, K. (2006). A free energy principle for the brain, *Journal of Physiology-Paris*, *100*, 70–87.

Friston, K. J. (2008a). Variational filtering. *NeuroImage*, *41*, 747–766.

Friston, K. (2008b). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211.
Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Science*, *13*, 293–301.
Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.
Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, *10*, 20130475.
Friston, K., Adams R. A., Perrinet L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*, 151.
Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS One*, *4*(7), e6421.
Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free energy formulation. *Biological Cybernetics*, *102*(3), 227–260.
Friston, K., & Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Networks*, *22*, 1093–1104.
Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*, 417–458.
Friston, K., Stephan, K., Li, B., & Daunizeau, J. (2010). Generalized filtering. *Mathematical Problems in Engineering*, 261670.
Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London B*, *290*, 181–197.
Hille, B. (2001). *Ion channels of excitable membranes* (3rd ed.). Sunderland, MA: Sinauer Associates.
Hodgkin, A., & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*, 500–544.
Hopfield, J. J. (1999). Brain, neural network, and computation. *Review of Modern Physics*, *71*, S431–S437.
Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, *14*(6), 1569–1572.
Jansen, B. H., Zouridakis, G., & Brandt, E. (1993). A neurophysiologically-based mathematical model of flash visual potentials. *Biological Cybernetics*, *68*, 275–283.
Jazwinski, A. H. (1970). *Stochastic process and filtering theory.* New York: Academic Press.
Jirsa, V. K., & Haken, H. (1996). Field theory of electromagnetic brain activity. *Physical Review Letter*, *77*, 960–963.
Koch, C. (1999). *Biophysics of computation: Information processing in single neurons.* New York: Oxford University Press.
Landau, L. P., & Lifshitz, E. M. (1976). *Classical mechanics* (3rd ed.). Amsterdam: Elsevier.
MacDonald, D. K. C. (2006). *Noise and fluctuations*. Mineola, NY: Dover.
Markov, N. T., & Kennedy, H. (2013). The importance of being hierarchical. *Current Opinion in Neurobiology*, *23*, 187–194.
Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., . . . Kennedy, H. (2014). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, *522*, 225–259.

Martin, S. J., Grimwood, P. D., & Morris, R. G. M. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience*, *23*, 649–711.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. Boston: Reidel.

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., & Kennedy, H. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, *89*, 384–397.

Okun, M., & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, *11*, 535–537.

Potjans, T. C., & Diesmann, M. (2014). The cell-type specific cortical microcircuit: Relating structure and activity in a full-scale spiking network model. *Cerebral Cortex*, *24*(3), 785–806.

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2017). Answering Schödinger's question: A free-energy formulation. *Physics of Life Reviews*, *24*, 1–16.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Robinson, P. A., Rennie, C. J., & Wright, J. J. (1997). Propagation and stability of waves of electrical activity in the cerebral cortex. *Physical Review E*, *56*, 826–840.

Schot, S. H. (1978). Jerk: The time rate of change of acceleration. *American Journal of Physics*, *46*, 1090–1094.

Schrödinger, E. (1967). *What is life? Mind and matter.* Cambridge: Cambridge University Press.

Seifert, U. (2005). Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Physical Review Letters*, *95*, 040602.

Sengupta, B, Tozzi, A., Cooray, G. K., Douglas, P. K., & Friston, K. J. (2016). Towards a neuronal gauge theory. *PLoS Biology*, *14*(3), e1002400.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, *379–423*, 623–656.

Steyn-Ross, M. L., & Steyn-Ross, D. A. (2016). From individual spiking neurons to population behavior: Systematic elimination of short-wavelength spatial models. *Physical Review E 93*, 022402.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409.

Visser, M. (2004). Jerk, snap and the cosmological equation of state. *Classical and Quantum Gravity*, *21*, 2603–2615.

von Helmholtz, H. (1962). *Treatise on physiological optics.* Mineola, NY: Dover.

Wilson, H. R. (1999). Simplified dynamics of human and mammalian neocortical neurons. *Journal of Theoretical Biology*, *200*, 375–388.

Received January 21, 2018; accepted May 9, 2018.