Recognition Dynamics of the Brian in the Free Energy Principle

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

Abstract. We formulate the computational processes of perception in the framework of the principle of least action by postulating the theoretical action being time-integral of the free energy in the brain sciences. The free energy principle is accordingly rephrased as that for autopoietic reasons all viable organisms attempt to minimize the sensory uncertainty about the unpredictable environment over a temporal horizon. By varying the informational action, we derive the brain’s recognition dynamics which conducts the adaptive inference of the external causes of sensory data with addressing only canonical positions and momenta of the brain’s representations of the dynamical world. To manifest the utility of our theory, we provide how the neural computation may be implemented biophysically at a single-cell level and subsequently be scaled up to a large-scale functional architecture of the brain. We also present formal solutions to the recognition dynamics for a model brain in linear regime and analyze the perceptual trajectories about fixed points in state space.

Keywords: recognition dynamics, perception, free energy principle, sensory uncertainty, informational action, principle of least action
1. Introduction

The quest for a universal principle that may explain the cognitive and behavioral operation of the brain is of a great scientific interest. The obvious difficulty in addressing the quest is the gap between the information processing and the biophysics that governs neurophysiology in the brain. However, it is evident that the matter, which is the ground-stuff upon which the brain functions emerge, comprises neurons obeying the laws guided by the physics principles. Thus, any biological principles that attempt to explain the brain’s large-scale workings must cope with our accepted physical reality \[1\]. It appears that on the current approaches still prevails the classical, effective epistemology of regrading perceptions as constructing hypotheses which may hit upon truth by producing symbolic structures matching physical reality \[2, 3, 4\].

One influential candidate at present that seeks for such a rubric in neuroscience is the free energy principle (FEP) \[5, 6, 7\]. For a technical appraisal of the FEP we refer to \[8\] where the theoretical assumptions and the mathematical structure involved in the FEP are reviewed in detail, and also the extended references are presented. In layman’s terms the FEP states that all viable organisms perceive and act on their external world by manipulating their internal states in a way to ensure their adaptive fitness or autopoiesis \[9\]. The biological mechanism that endows the brain with the operation is theoretically framed into an information-theoretic measure, called the ‘informational free energy (IFE)’\(^\dagger\). According to the FEP, a living system tries to minimize the sensory surprisal when it faces with an external cause that perturbs its spontaneous equilibrium within its physiological boundary by pursuing perceptive as well as active inferences. However, the brain does not preside over instreaming sensory distribution; accordingly the brain cannot directly minimize the sensory surprisal but, instead, minimize its upper bound, the IFE. The probabilistic rationale of the FEP argues that the brain’s representations of the uncertain external world are the sufficient statistics, e.g. means or variances, of an arbitrary probability density encoded in the brain. 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 inversion using its internal top-down model about predicting, or generating, the sensory data. There are growing experimental support of the brain’s maintaining internal models of the environment in order to predict sensory inputs and to prepare actions \[10\]. The computational operation of the adaptive inference is subserved by the brain variables and the resulting perceptual mechanics is termed as the ‘recognition dynamics (RD)’.

Although the suggestion of the FEP has been promising to account for the brain’s inferring mechanism of and acting upon sensory causes, we find certain theoretical as well as mathematical subtleties in the conventional formulation:

First, the FEP minimizes the IFE at each point in time for all successive sensory inputs \[11\]. However, precisely the objective function to be minimized is the continuously accumulated IFE over a finite time \[\dagger\]. The minimization must be

\[\dagger\] According to the FEP, the updating or learning of the generative model takes places in the brain
performed with respect to trajectories over a temporal horizon across which an organism encounters with atypical events to the organism’s natural habitat and its own biology.

Second, the FEP utilizes the gradient-descent scheme in practically implementing the minimization of the IFE, which provides with an efficient method of finding optimal solutions in machine learning theory [13]. However, the engaged mathematical device is not based on an appropriate scientific principle, making up only a heuristic approach to solve engineering optimization problems.

Third, the FEP employs the ‘generalized motions’ of an infinite number of the so called ‘generalized coordinates’ to account for the dynamical nature of the environment [14]. However, the ensuing theoretical construct is not compatible with the standard Newtonian mechanics [6]. With the hired theory, it is obscure to decide the number of independent dynamical variables for a complete description.

Fourth, the FEP introduces the hydrodynamic concepts of ‘path of mode (motion of expectation)’ and ‘mode of path (expected motion)’, which are not properly derived quantities and, again, are extra-theoretical [15]. However, they are essential to ensure an equilibrium solution to the RD in perceptual as well as active inference of the dynamical external states.

Fifth, the FEP considers the states of the external world ‘hidden’ because what the brain faces is only a probabilistic sensory mapping. The distinction made between the hidden-state representations, responsible for intra-level dynamics, and causal-state representations, responsible for inter-level dynamics, in the hierarchical brain is conjectural [16]. A proper biophysical formulation that supports the top-down idea is further required.

In this paper we present a mechanistic formulation of the RD of the brain in the framework of Hamilton’s principle of least action [18]. Motivated by the aforementioned observations, we try to resolve some of the technical pitfalls in the FEP framework. To be specific, the goal is to recast the gradient descent strategy of minimizing the IFE into the mathematical framework that appeals to the normative physics rules. Consequently, we reformulate the RD in terms of only the canonical, physical realities to dispense with the generalized coordinates of infinitely recursive time-derivatives of the brain variables. We shall focus our attention in this paper on the brain’s perceptual mechanics of inferring of the external world, given sensory data. The problem of learning of the external world on a longer time scale than that associated with perceptual inference. To derive the RD of the slow variables for synaptic efficacy and gain, the time-integral of the IFE is taken as an objective function; however, again the gradient descent method is executed in a pointwise way in time [12].

§ The mechanical state of a particle is specified only by position and velocity in the Newtonian mechanics, and no physical observables are assigned to the dynamical orders beyond the second-order. In some literature [17], the concept of ‘jerk’ is assigned to the third-order time-derivative of position as a physical reality. From the mathematical perspective such a generalization is not forbidden. However, not only higher-orders are difficult to measure but more seriously it raises the question of what the corresponding cause to jerk as the force to acceleration. And, same impasse in all next orders.
via updating the internal model of the world and of active inference of changing sensory input via action on the external world, see for instance [19], are deferred to an upcoming paper.

Supported by the present day evidences [20, 21], we admit that the bi-directional facets in informational flow in the brain. The nonstationary states of the external world incite sensory data at the brain-environment interface such as sensory receptors or interoceptors within an organism. While the incited electro-opto-chemical interaction in sensory neurons must transduce forward in the anatomical structure of the brain, complying with the idea of perception as constructing hypotheses there must be backward pathway as well in information processing in the functional hierarchy of the brain. To understand how the functional architecture is emergent from the electrophysiology of biophysics and anatomical organization of the brain is a forefront research interest (see, for instance, [22] and references therein). In this work we consider a simple model that effectively incorporates the functional hierarchy in sensory inference.

Here, we synopsize how we translate the information-theoretic constructs in the FEP into the standard physics formulation in this work: (i) According to the FEP, the brain represents the environmental features statistically efficiently, using the sufficient statistics $\mu$. We assume that $\mu$ specifies the basic computational unit of the neural attributes of perception in the brain. Such a basic unit is considered a ‘perceptual particle’ which may be a single neuron or physically coarse-grained multiple neurons forming a small particle; (ii) We postulate that the Laplace-encoded IFE in the FEP, denoted as $F$ (Sec. 2.1), serves as an effective, informational Lagrangian (IL) of the brain, denoted as $L$. Accordingly, the informational action (IA), which we denote by $S$, is defined to be time-integral of the approximate IFE (Sec. 3.1); (iii) Then, conforming to the Hamilton principle of least action, the equations of motion of the perceptual particles are derived by mathematically varying the IA with respect to both $\mu$ and $\dot{\mu}$. The resulting Lagrange equations constitute the perceptual mechanics, i.e. the RD of the brain’s inferring of the external causes of the sensory stimuli (Sec. 3.1); (iv) In turn, we obtain the brain’s informational Hamiltonian $H$ from the Lagrangian via a Legendre transformation. Consequently, we derive a set of coupled, first-order differential equations for $\mu$ and its conjugate $p_\mu$, which are equivalent to the perceptual mechanics derived from the Lagrange formalism. Accordingly, we interpret that the perceptual state-space of the brain is spanned by the position $\mu$ and momentum $p_\mu$ variables of the constituting neural particles. (Sec. 3.2); (v) We adopt the Hodgkin-Huxley (H-H) neurons as biophysical neural correlates which form the basic perceptual unit in the brain. We first derive the RD of sensory perception at a single-neuron level where the membrane potential, ionic transport, and 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 averaged, activation and connection variables, mediating the intra- and inter-level dynamics and formulate the hierarchical RD in the brain. As sensory
perturbation comes in at the lowest level, i.e. sensory interface, the brain carries out the RD in its functional network and finds an optimal trajectory which minimizes the IA.

To summarize, we have adopted the IFE as an informational Lagrangian of the brain and subsequently defined the informational Hamiltonian. Then, the principle of least action has guided us to construct a process theory of the perceptual mechanics with reference only to positions and momenta; positions and momenta being metaphorical for the perceptual states and the brain’s prediction errors, respectively. The distinction between the causal (autonomous-hidden) and hidden states was dispensed with, both of which are biophysical neuronal activities on different time scales. The resulting RD is statistically deterministic, arising from indeterministic motions of the environmental states and stochastic sensory mappings. The successful solutions of the RD are stable equilibrium trajectories in the neural state space, specifying the tightest upper bound of the sensory uncertainty, conforming to the rephrased FEP. Our formulation allows solutions in an analytical form in linear regimes near limit cycles, expanded in terms of the eigenvalues and the eigenvectors of the Jacobian and, thus, provides with tractability of a real time analysis. We hope that our theory will motivate further investigations on the active inference and learning problems by employing detailed numerical simulations of some model brains.

This paper is organized as follows. We first recapitulate the FEP in Sec. 2 to support our motivation for recasting the gradient descent scheme into the standard mechanics formulation. In the followed Sec. 3 we present the RD reformulated in the Lagrangian and Hamiltonian formalisms. Then, in Sec. 4 biophysical implementations of our theory at the cellular level and in the scaled-up hierarchical brain are formulated, where nonlinear as well as linear dynamical analyses are carried out. Finally, a discussion is provided in Sec. 5.

2. The free energy principle

To unveil our motivation for this paper, we shall compactly digest here the IFE and the FEP, that are currently practiced in the brain sciences. The RD is the brain’s machinery to minimize the IFE under the FEP. For a detailed technical appraisal of the FEP we refer to [8] from which we borrow the mathematical notations.

2.1. The 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, namely its environment. The states of the environment are denoted as $\vartheta$ collectively. The signals from the environment are registered in organism’s sensory interface as sensory data $\varphi$ which are physically encoded in the brain as a probability density $p(\varphi)$.

Then, the sensory uncertainty $H$ associated with input data $\varphi$ is the average of the
Recognition Dynamics of the Brian in the Free Energy Principle

self-information, \(- \ln p(\varphi)\), over the sensory distribution \(p(\varphi)\),

\[
H \equiv \int d\varphi \{ - \ln p(\varphi) \} p(\varphi).
\]

(1)

The self-information, which is also termed as the sensory ‘surprise’ or surprisal in the FEP, quantifies the survival tendency of living organisms in the unpredictable environment. Assuming that the sensory density describes an ergodic ensemble of sensory streaming, the sensory uncertainty may be converted into a time-average as

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

where \(T\) is the temporal window over which a biological event takes places, i.e. a temporal horizon. Next, by adding \(\int d\vartheta q(\vartheta) \ln[q(\vartheta)/p(\vartheta|\varphi)]\) to the integrand on the right-hand side (RHS) of the preceding equation, which is the positive-definite Kullback-Leibler divergence, we get

\[
\int dt \{ - \ln p(\varphi) \} \leq \int dt \mathcal{F}[\vartheta, \varphi],
\]

(2)

where \(\mathcal{F}\) is the IFE that is defined to be a functional of the two probability densities, \(q(\vartheta)\) and \(p(\vartheta, \varphi)\):

\[
\mathcal{F}[\vartheta, \varphi] = - \ln p(\varphi) + \int d\vartheta q(\vartheta) \ln \frac{q(\vartheta)}{p(\vartheta|\varphi)}
\]

\[
= \int d\vartheta q(\vartheta) \ln \frac{q(\vartheta)}{p(\vartheta, \varphi)},
\]

(3)

where \(q(\vartheta)\) and \(p(\vartheta, \varphi)\) are termed the R-density and the G-density, respectively. The R-density is the brain’s probabilistic representation of the external world which must be encoded physically in the brain. The G-density, a joint probability between \(\vartheta\) and \(\varphi\), underlies the brain’s top-down model of Bayesian belief of the external causes of the sensory data. Note that the sensory uncertainty on the left hand side (LHS) of Eq. (2) specifies the accumulated surprisal over the time interval involved in an environmental event.

Equation (2) is the mathematical statement of the FEP, implying that “any viable organisms try to avoid being placed in an atypical situation in their environmental habitats by minimizing the sensory uncertainty”. However, organisms do not possess control over the sensory distribution \(p(\varphi)\); accordingly they are not able to minimize directly the sensory uncertainty. Instead, they are to minimize the upper bound of Eq. (2), \(\int dt \mathcal{F}\) as a proxy for the sensory uncertainty. The brain conducts the minimization probabilistically by updating the R-density to approximate the posterior density \(p(\vartheta|\varphi)\), i.e. carrying out a Bayesian inference of the causes \(\vartheta\) of the inputs \(\varphi\). In the conventional application of the FEP the following approximate inequality is always exercised \[23\] \[24\],

\[
- \ln p(\varphi) \leq \mathcal{F}.
\]

(4)

However, we note here that the inequality, Eq. (4) is not equivalent, in general, to Eq. (2). It is only a point approximation piecewise in time.
The functional shape of the R-density is not given \textit{a priori}. It would be fixed mathematically after providing all orders of its moments, but not tractable. To circumvent the difficulty, usually one invokes an approximate Bayesian inference for formulation by assuming a Gaussian fixed-form for the R-density, the ‘Laplace approximation’,

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

which is fully characterized simply by its mean $\mu$ and variance $\zeta$, first and second order \textit{sufficient statistics}, respectively. Then, by substituting Eq. (5) into Eq. (3) and after some technical approximations, see \cite{8}, one can convert the IFE functional $F$ into

\[ F[\mu, \varphi] \rightarrow -\ln p(\mu, \varphi) \equiv F(\mu, \varphi). \]  

In the end of manipulation the outcome becomes a function of only the first order sufficient statistics $\mu$, given the sensory input $\varphi$. The resulting IFE function $F$ in Eq. (6) is termed the ‘Laplace-encoded’ IFE in which the parameter $\mu$ is the brain’s effective representation of the external world, physically encoded as the brain variables.

To proceed with minimization by means of the Laplace-encoded $F$, the joint probability density $p(\mu, \varphi)$ must be supplied. In view of the factorization,

\[ p(\varphi, \mu) = p(\varphi|\mu)p(\mu), \]  

the brain must possess internally its generative model $p(\varphi|\mu)$ about the sensory data and also the prior belief $p(\mu)$ of its expectation about their causes. It is hypothesized that the brain predicts the registered sensory input $\varphi$ in a feedback manner,

\[ \varphi = g(\mu) + z, \]  

where $g(\mu)$ is a generative model of the brain. The random fluctuation involved in the sensory observation process has been characterized by $z$. Also, it is assumed that the brain variable $\mu$ obeys the stochastic equation of motion,

\[ \frac{d\mu}{dt} = f(\mu) + w \]  

where $f(\mu)$ is a generating function of brain dynamics and $w$ is the associated random fluctuation. Assuming the Gaussian noises for both fluctuations about zero means with variance $\sigma_z$ and $\sigma_w$, respectively, the likelihood density $p(\varphi|\mu)$ and the prior $p(\mu)$ are written, using the notation introduced in Eq. (5), as

\[ p(\varphi|\mu) = \mathcal{N}(\varphi - g(\mu); 0, \sigma_z), \]  

\[ p(\mu) = \mathcal{N}(\mu - f(\mu); 0, \sigma_w), \]  

where we have set $\dot{\mu} = d\mu/dt$. Next, utilizing Eqs. (10) and (11), one can convert the Laplace-encoded IFE Eq. (6), up to a constant, into

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

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 variables $\varepsilon_z$ and $\varepsilon_w$ specify the discrepancies between the brain’s prediction of sensory data and of change in the brain states, respectively.

2.2. Gradient descent scheme of the RD

With the Laplace-encoded IFE as an instrumental tool, the brain searches for the tightest bound for the surprisal, complying with Eq. (4), by varying its internal states $\mu$. Here the key question is what machinery the brain actually hires for the minimization procedure. Typically, the gradient descent method in machine learning theory is employed in the conventional FEP.

By regarding the IFE function $F$ as an objective function, one sets up the gradient descent equation in the form,

$$\dot{\mu} = -\kappa \nabla_\mu F(\mu, \varphi). \tag{13}$$

In the above $\dot{\mu}$ implies a temporal or parametric update of the brain variable $\mu$ and $\nabla_\mu$ is the gradient operator with respect to $\mu$, and $\kappa$ is the learning rate that controls the speed of optimization. The gradient operation on the RHS of Eq. (13) can be performed explicitly for given $F$, Eq. (12) to give

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

In steady state, defined by $\dot{\mu} = 0$, the solution $\mu^{(0)}$ to the relaxation equation, Eq. (13) must satisfy $\nabla_\mu F = 0$. Subsequently, it is interpreted that such a solution corresponds to an equilibrium (or fixed) point of the IFE function $F$, specifying a local minimum.

By inspection, however, we find that the gradient-descent construct in the above form requires a technicality in finding an optimal solution. This is because imposing the condition, $\dot{\mu} = 0$ on the LHS of Eq. (13), does not guarantee a desired fixed point. The reason is that $\dot{\mu}$ also appears on the RHS of Eq. (13) via $F$, which reflects the nonstationary world prescribed by the equation of motion Eq. (9). This difficulty is remedied in the conventional theory of the FEP by proposing some extra theoretical constructs. First, in an attempt to incorporate the continually changing external states into formulation, the mathematical construct of infinitely recursive, higher-order motion of the generalized coordinates is introduced. The generalized coordinates are defined to be a vector in the state space spanned by all dynamical orders of $\mu$,

$$\tilde{\mu} = (\mu, \mu', \mu'' \cdots) \equiv (\mu[0], \mu[1], \mu[2], \cdots) \tag{14}$$

where vector components are defined, with understanding $\mu[0] \equiv \mu$, as

$$\mu[n+1] = \mu'[n] \equiv D\mu[n].$$

The terminology of the generalized coordinates causes confusion. In classical mechanics, the generalized coordinates are the independent, dynamical variables which are required to specify the configuration of a system completely. The number of generalized coordinates determines the degree of freedom in the system.
Note that the notation $\tilde{D}^{\mu}_{rns}$ has been introduced to denote the *dynamical update* of a generalized state $\mu_{rns}$, in contrast to $\dot{\mu}_{rns}$ for the temporal update. Similarly, the sensory-data $\tilde{\varphi}$ are expressed in the generalized coordinates as

$$\tilde{\varphi} = (\varphi, \varphi', \varphi'', \cdots) \equiv (\varphi[0], \varphi[1], \varphi[2], \cdots).$$

(15)

Each component in the vectors, $\tilde{\mu}$ and $\tilde{\varphi}$, is to be considered as an independent dynamical variable. Then, the generalization of Eqs. (8) and (9) follows after some technical approximations as (for details, see [8])

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

(16)

$$\tilde{D}(\tilde{\mu}) = \tilde{f} + \tilde{w};$$

(17)

where $\tilde{D}(\tilde{\mu}) = (\mu', \mu'', \mu''', \cdots)$. At dynamical order $n$, Eqs. (16) and (17) are explicitly spelled out as

$$\varphi[n] = \frac{\partial \tilde{g}}{\partial \mu} \mu[n] + \tilde{z}[n],$$

$$\tilde{D}\mu[n] = \frac{\partial \tilde{f}}{\partial \mu} \mu[n] + \tilde{w}[n].$$

Then, furnished with the extra theoretical constructs, the IFE becomes a function of the generalized coordinates $\tilde{\mu}$, given sensory data $\tilde{\varphi}$, $F = F(\tilde{\mu}, \tilde{\varphi})$.

Next, the gradient descent scheme is extended to embody the generalized coordinates. To this end, one first assumes that the dynamical update $\tilde{D}(\tilde{\mu})$ is distinctive from the temporal update $\dot{\mu}$ and then recasts Eq. (13) into the form,

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

(18)

With this form of the gradient descent equation, the conventional FEP claims that a desired fixed point $\tilde{\mu}^* \equiv \tilde{\mu}(t \to \infty)$ reaches in the minimization scheme when the two rates of the brain variables, $\dot{\mu}$ and $\tilde{D}(\tilde{\mu})$ become coincident, i.e. when $\dot{\mu}_{rns} = \tilde{D}\mu_{rns}$ at every dynamic order $n$. Note the extra theoretical construct of $\tilde{D}(\tilde{\mu})$ was installed on the LHS in Eq. (18) without a rigorous basis. The entire minimization procedure is compactly expressed in the literature as

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

In brief, the brain performs the RD of perceptual inference by biophysically implementing Eq. (18) in the gray matter. The steady-state solution $\tilde{\mu}^*$ specifies minimum value of the IFE, say $F_{\min} = F(\tilde{\mu}^*, \tilde{\varphi})$, giving the tightest bound of the surprisal [see Eq. (4)] associated with a given sensory experience $\tilde{\varphi}$. Despite its frequent employment in practicing the FEP, we have disclosed some theoretical subtleties involved in the customary formulation of the gradient descent scheme, that has motivated our reformulation.
3. The informational action principle

The RD condensed in the preceding Sec. 2.2 is based on the mathematical statement Eq. (4) of the FEP, which is a point approximation of Eq. (2). Here we reformulate the perceptual mechanics by complying with the full mathematical statement of FEP given in Eq. (2). Accordingly, we need a formalism that allows minimization of the time-integral of the IFE, not at each point in time. We have come to assimilate that the theoretical ‘action’ in the principle of least action in physics naturally serves for the goal [18]. This formalism allows us to dispense with the generalized motions of a dynamical state comprising an infinite number of discrete time-derivatives. We do not need the ad hoc classification of the parametric update (µ) and the dynamical update (Dµ) of the state variables. In what follows, we shall consistently use the dot symbol to denote time-derivative of a dynamical variable.

3.1. Lagrangian formalism

To address the principle of least action, the ‘Lagrangian’ of the system must be supplied. We define the informational Lagrangian (IL) of the brain, denoted by \( L \), as the Laplace-encoded IFE function,

\[
L(\mu, \dot{\mu}; \varphi) = F(\mu, \dot{\mu}; \varphi),
\]

where we have placed the semicolon in \( L \) to indicate that \( \mu \) and \( \dot{\mu} \) are the two brain’s dynamical variables, given sensory input \( \varphi \). The sensory inputs are time-dependent, in general; \( \varphi = \varphi(t) \), reflecting the changing external states. Note the IL proposed is not a physical quantity but an information-theoretic object. When we take Eq. (12) as an explicit expression for \( F \), the IL is written up

\[
L(\mu, \dot{\mu}; \varphi) = \frac{1}{2} \sigma_\omega^{-1} (\dot{\mu} - f(\mu))^2 + \frac{1}{2} \sigma_z^{-1} (\varphi - g(\mu))^2,
\]  

where we have dropped out the constant term, \( \frac{1}{2} \ln(\sigma_\omega \sigma_w) \), which does not affect the dynamics of \( \mu \) and \( \dot{\mu} \).

Next, we postulate that the perceptual dynamics of the neural particles conforms to the principle of least action [18]. Accordingly, we suppose that the brain’s perceptual operation corresponds to searching for an optimal dynamical path that minimizes the informational action (IA), denoted by \( S \),

\[
S \equiv \int_{t_i}^{t_f} dt \ L(\mu, \dot{\mu}; \varphi);
\]

where \( t_f - t_i \equiv T \) is the temporal horizon over which a living organism encounters an environmental event. When functional derivative of \( S \) is carried out with respect to \( \mu \) and \( \dot{\mu} \), it gives

\[
\delta S = \left[ \frac{\partial L}{\partial \mu} \delta \mu \right]_{t_i}^{t_f} - \int_{t_i}^{t_f} dt \left( \frac{d}{dt} \frac{\partial L}{\partial \dot{\mu}} - \frac{\partial L}{\partial \mu} \right) \delta \mu.
\]
By imposing $\delta S = 0$ 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
$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mu}} - \frac{\partial L}{\partial \mu} = 0. \tag{21}$$
Using the specified Lagrangian, Eq. (19), in Eq. (21), we obtain a Newtonian equation of motion for the brain variable $\mu$, 
$$\sigma_w^{-1} \ddot{\mu} = \Lambda_1 + \Lambda_2, \tag{22}$$
where we have defined
$$\Lambda_1 \equiv \sigma_w^{-1} f \frac{\partial f}{\partial \mu} \quad \text{and} \quad \Lambda_2 \equiv -\sigma_w^{-1}(\varphi - g) \frac{\partial g}{\partial \mu}. \tag{23}$$
Equation (22) entails the RD of the brain in the Lagrangian formulation. We interpret that the inverse of the variance $\sigma_w^{-1}$ plays, as a metaphor, a role of inertial mass of the neural particles. Accordingly, the LHS of Eq. (22) represents an inertial force, i.e. the product of ‘inertial mass’ and ‘acceleration’, $\ddot{\mu}$. Note that the inverse of variance is interpreted as precision in the Friston formulation [8], which gives a measure for the accuracy of the brain’s expectation or prediction of sensory data. So, the precision is the ‘informational mass’ of the neural particle metaphorically. Also, the terms on the RHS are interpreted as the ‘forces’ that drive the internal $F_1$ as well as sensory $F_2$ excitations in the brain. The acceleration can be evaluated from $\ddot{\mu} = \sum \Lambda_i / \sigma_w^{-1}$ when the net force is known.

While the brain integrates the RD for a given sensory input, an optimal trajectory $\mu^*(t)$ is achieved in $\mu$-space. The steady-state condition is given by
$$\dot{\mu} = \text{const}, \tag{24}$$
in the long-time limit $t \to \infty$ where the net force vanishes. Note that equation (24) defining the fixed-point, $\mu_{eq} = \mu^*(\infty)$, differs from and is more general than that of the gradient descent scheme, $\dot{\mu} = 0$ [see Sec. (2.2)]. In contrast to the gradient descent equation, Eq. (13), the RD Eq. (22) from our Lagrangian formulation causes no ambiguity to ensure a fixed point.

3.2. Hamiltonian formalism

The mechanical formulation can be made more formal in terms of Hamiltonian language which admits position and momentum as independent brain variables, instead of position and velocity as in the Lagrangian formulation.

The ‘canonical momentum’ $p$, which is conjugate to the position $\mu$, is defined via Lagrangian $L$ as [18]
$$p \equiv \frac{\partial L}{\partial \dot{\mu}} = \sigma_w^{-1}(\dot{\mu} - f), \tag{25}$$
which evidently differs from the ‘kinematic’ momentum $\sigma^{-1}_w \mu$. Then, the informational ‘Hamiltonian’ $H$ may be constructed from the Lagrangian using Legendre’s transformation \[18\],

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

(26)

The first term on the RHS of Eq. (26) can be further manipulated to give

$$\sum \frac{\partial L}{\partial \dot{\mu}} \dot{\mu} = \sigma^{-1}_w \dot{\mu}^2 - \sigma^{-1}_w \dot{\mu} f.$$  

By plugging the outcome and also the Lagrangian $L$ given in Eq. (19) into Eq. (26), we obtain the Hamiltonian as a function of $\mu$ and $p$ as desired, given $\varphi$,

$$H(\mu, p; \varphi) = T(p) + V(\mu, p; \varphi),$$

(27)

where Eq. (25) has been used to replace $\dot{\mu}$ with $p$. The first term on the RHS of Eq. (27) is the ‘kinetic energy’ which depends only on momentum,

$$T(p) = \frac{p^2}{2\sigma^{-1}_w}.$$  

(28)

Also, the second term on the RHS of Eq. (27) is the ‘potential energy’ which depends on both position and momentum,

$$V(\mu, p; \varphi) = V(\mu; \varphi) + pf(\mu),$$

(29)

where we have defined the momentum-independent term separately as $V$,

$$V(\mu; \varphi) = -\frac{1}{2} \sigma^{-1}_w (\varphi - g)^2.$$  

(30)

We remark that the sensory stimuli $\varphi$ enter the Hamiltonian only through the potential-energy part $V$ which becomes ‘conservative’ when $\varphi$ is static. For time-varying sensory inputs, in general, the Hamiltonian is nonautonomous. In Fig. 1 we depict the conservative potential energy, using three term approximations for the generative function,

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

We have observed numerically that the static sensory signal $\varphi$ changes the distance between two unstable fixed points, but do not affect the location of the stable equilibrium point. In addition, the depth of the stable equilibrium valley gets deeper as the magnitude of $\varphi$ increases.

Next, we take the total derivative of the Hamiltonian given in Eq. (26) with respect to $\mu$ and $\dot{\mu}$ to get

$$dH(\mu, p; \varphi) = \sum d(p \dot{\mu}) - dL(\mu, \dot{\mu}; \varphi)$$

$$= \dot{\mu} dp + p d\mu - \left( \frac{\partial L}{\partial \mu} d\mu + \frac{\partial L}{\partial \dot{\mu}} d\dot{\mu} \right)$$

$$= - \dot{p} \mu d\mu + \dot{\mu} dp.$$
Figure 1. The potential energy, given in Eq. (30), in arbitrary units for appropriately chosen two sets of the brain’s phenomenological parameters. Both cases exhibit a stable equilibrium point at the central well and two unstable fixed points on the top of the side hills.

By comparing the last 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 \( \mu \) and \( p \) of a neural particle,

\[ \dot{\mu} = \frac{\partial \mathcal{H}}{\partial p}, \quad (31) \]
\[ \dot{p} = -\frac{\partial \mathcal{H}}{\partial \mu}. \quad (32) \]

For given \( \mathcal{H} \) in Eq. (27), we spell out the RHS of Eq. (31) to get

\[ \dot{\mu} = \frac{1}{\sigma_w} p + f \quad (33) \]

which is identical to Eq. (25). Similarly, the second equation, Eq. (32) is spelled out

\[ \dot{p} = -\frac{\partial V}{\partial \mu} - \frac{\partial f}{\partial \mu} p. \quad (34) \]

The first term on the RHS of Eq. (34) specifies the conservative force,

\[ -\frac{\partial V}{\partial \mu} \rightarrow -\sigma_w^{-1} (\varphi - g) \frac{\partial g}{\partial \mu}. \]

Whereas, the second term on the RHS of Eq. (34) specifies the dissipative force, where \( \partial f/\partial \mu \) plays the role of damping coefficient.

The derived set of coupled equations for the variables \( \mu \) and \( p \) furnish the RD of the brain in phase space spanned by \( \mu \) and \( p \), which involve only first order time-derivatives. When time-derivative is taken once more for both sides of Eq. (33) with followed substitution of Eq. (34) for \( \dot{p} \), the outcome becomes identical to the Lagrangian
equation of motion, Eq. (22). This confirms that 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, Eqs. (33) and (34) are autonomous and for time-dependent sensory input they become non-autonomous. The RD can be integrated, provided appropriate models for the generating functions \(f\) and \(g\) are given. The fixed point \((\mu_0, p_0)\) would be limit cycles, in general, in phase space, which can be calculated by imposing that both LHSs of Eqs. (33) and (34) vanish at the same time,

\[
\dot{\mu} = 0 \quad \text{and} \quad \dot{p} = 0. \tag{35}
\]

One can readily confirm that these fixed point conditions match with the Newtonian equilibrium condition, \(\sum_i F_i = 0\). The situation corresponds to the brain’s resting state which is not simply a local minimum in the conservative energy because of the non-potential terms in the Hamiltonian.

### 3.3. Multivariate formulation

Having established the informational action principle with a single brain variable \(\mu\), we now extend our formulation to the general case of the multivariate brain. We denote \(\tilde{\mu}\) as a \(N\)-dimensional brain variable,

\[
\tilde{\mu} = (\mu_1, \mu_2, \cdots, \mu_N),
\]

that respond to the multiple of sensory inputs \(\tilde{\varphi}\), in general,

\[
\tilde{\varphi} = (\varphi_1, \varphi_2, \cdots, \varphi_N).
\]

For simplicity, we neglect the statistical correlation among the different noises both in the equations of motion of the brain variables and the observation equations of the sensory data. Then, within the independent-particle approximation the informational Lagrangian Eq. (19) is generalized as

\[
\mathcal{L}(\tilde{\mu}, \tilde{\mu}; \tilde{\varphi}) = \frac{1}{2} \sum_{\alpha=1}^{N} \left[ \sigma_{w\alpha}^{-1} (\dot{\mu}_\alpha - f_\alpha(\tilde{\mu}))^2 + \sigma_{z\alpha}^{-1} (\varphi_\alpha - g_\alpha(\tilde{\mu}))^2 \right], \tag{36}
\]

where we have dropped out the terms which contain only static parameters, \(\sigma_{z\alpha}\) and \(\sigma_{w\alpha}\). One may extend Eq. (36) to interacting neural nodes in terms of covariance matrix formulation [8], which is not our concern here. Subsequently, the conjugate momentum to the generalized coordinate \(\mu_\alpha\) is determined by an explicit evaluation of

\[
p_\alpha = \frac{\partial \mathcal{L}}{\partial \dot{\mu}_\alpha} = \sigma_{w\alpha}^{-1} (\dot{\mu}_\alpha - f_\alpha). \tag{37}
\]

In turn, the Hamiltonian of the multivariate brain can be constructed from Eq. (27) as

\[
\mathcal{H}(\tilde{\mu}, \tilde{p}; \tilde{\varphi}) = \mathcal{T}(\tilde{\mu}, \tilde{p}; \tilde{\varphi}) + \mathcal{V}(\tilde{\mu}, \tilde{p}; \tilde{\varphi}) \tag{38}
\]
where first term on the RHS is the kinetic energy,

\[ T(\tilde{p}; \tilde{\phi}) \equiv \sum_{\alpha} \frac{p_{\alpha}^2}{2\sigma_{wa}} \]  

(39)

and the potential energy \( V \) is identified as

\[ V(\tilde{\mu}, \tilde{p}; \tilde{\phi}) \equiv \sum_{\alpha} \left[ -\frac{1}{2}\sigma_{za}^{-1}(\varphi_{\alpha} - g_{\alpha})^2 + p_{\alpha}f_{\alpha} \right]. \]  

(40)

Then, it is straightforward to derive the RD of the variables \( \mu_{\alpha} \) and \( p_{\alpha} \), given sensory data \( \varphi_{\alpha} \), as

\[ \dot{\mu}_{\alpha} = \frac{\partial H}{\partial p_{\alpha}} = \frac{1}{\sigma_{wa}} p_{\alpha} + f_{\alpha}, \]  

(41)

and for their conjugate momenta,

\[ \dot{p}_{\alpha} = -\frac{\partial H}{\partial \mu_{\alpha}} = -\sigma_{za}^{-1}(\varphi_{\alpha} - g_{\alpha}) \frac{\partial g_{\alpha}}{\partial \mu_{\alpha}} - \frac{\partial f_{\alpha}}{\partial \mu_{\alpha}} p_{\alpha}. \]  

(42)

Equations (41) and (42) are a couple set of equations for the multivariate coordinates \( \mu_{\alpha} \) and their conjugate momenta \( p_{\alpha} \) of the neural particles in 2N-dimensional phase space. With some working models for \( f_{\alpha} \) and \( g_{\alpha} \), they form the RD of the brain in the Hamiltonian prescription. In Fig. 2 we illustrate a schematic of the perceptual circuitry implied by the RD at a neural node. It is admissible to assume that the brain is at the outset in a resting state. As the sensory inputs \( \varphi_{\alpha} \) come in, the brain performs the RD, by integrating Eqs. (41) and (42), 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 Eq. (20). The entire minimization procedure may be stated abstractly as

\[
(\mu^*_\alpha, p^*_\alpha) = \arg \min_{\mu_\alpha, p_\alpha} S(\mu_\alpha, p_\alpha; \varphi),
\]

where \( S \) is the IA.

4. Biophysical implementation

The aged Hodgkin-Huxley (H-H) model continues to be used in virtually all nerve excitation studies to this day [26, 27]. In extracellular electrical recordings the local field potential and multi-unit activity result in as combined signals from a population of neurons, typically thousands in numbers with spatial range of up to hundred micrometers [28]. In this work we regard the coarse-grained H-H neurons as the basic building blocks of encoding and transmitting perceptual message inside the brain. Such averaged neuronal variables 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 enacts a node that collectively forms the whole neural network on the large scale.

Before proceeding, we simply mention that there are much biophysical efforts to describe such average neural properties; for instance, the neural mass models and neural field theories are a few examples [29, 30, 31, 32, 33]. Also, we note the bottom-up effort of trying 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 [34, 35].

4.1. Single cell description

We first present how our formulation may be implemented at a neurocentric single-cell level [36]. A typical neuron receives electrochemical inputs from the sensory periphery via glutamate, which in turn regulates the neuronal states by exciting or inhibiting the membrane potential \( V \), gating variables \( \gamma_l \), and ionic concentrations \( n_l \); where \( l \) is the ion channel index. We admit that \((V, \{n_l\}, \{\gamma_l\})\) specify the neural states of a neuron as a perceptual particle in the neural configurational space. They serve the ‘generalized coordinates’ of our neural particle in standard physics terms [18], which we denote as a multi-dimensional row vector,

\[
\tilde{\mu} = (V, n_l, \gamma_l) = (\mu_1, \mu_2, \mu_3, \cdots).
\]

Note although we use the same notation \( \tilde{\mu} \) as the one in Eq. (14), the two theoretical constructs are completely different.

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

\[
C \frac{dV}{dt} = \sum_l \gamma_l G_l(E_l - V) + I_{ex}(t)
\]

(44)
where $C$ is the membrane capacitance, $G_l$ is the maximal conductance of ion channel $l$, $\gamma_l$ is the probability factor associated with opening or closing channel $l$ which in general 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. The reverse potential $E_l$ of $l$-th ion channel is given, assuming its time-dependence via nonequilibrium ion-concentrations, in general, as

$$E_l(t) = \frac{k_B T}{q_l} \ln \frac{n_{l1}(t)}{n_{l0}(t)},$$

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 $n_{l1}(t)$ and $n_{l0}(t)$ are the instantaneous ion concentrations inside and outside the membrane, respectively. In the steady state without external current, $I_{ex} = 0$, $V$ tends to the resting (Nernst) potential $V(t \to \infty)$ with retaining ionic concentrations in electro-chemical equilibrium. The gating variable $\gamma_l$ of ion channels is assumed to obey the kinetics, different model of that may be preferable,

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

where the relaxation time $\tau_l$ and steady-state gating variable $\gamma_{leq}$ depend on the membrane potential, in general,

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

and $\eta_l$ is the noise involved in the process.

For ionic concentration dynamics, we suppose that ion concentrations $\{n_l\}$ vary slowly compared to the membrane potential and gating-channel kinetics, and consequently treat them statically 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 statically in below.

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

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

where $\alpha$ runs 1, 2, $\cdots$ with implying $\mu_1 = V$, $\mu_2 = \gamma_1$, and $\mu_3 = \gamma_2$, etc. Note the static variables $\{n_l\}$ have been separated by the semicolon from the other dynamical variables in the arguments of the generating functions $f_\alpha$, that are specified as

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

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

The term $w_\alpha$ in Eq. (47) describe the noisy current $w_V$ flowing into the neural particle, in addition to the deterministic contribution $I_{ex}$, and the noise $w_{\gamma_l} = \eta_l$ associated with
the activation and inactivation of ion channels, respectively. For both noises, we assume
the Gaussian distributions \( \mathcal{N}(\mu_a - f_a; 0, \sigma_{wa}) \) with variances \( \sigma_{wa} \) about zero means.

Regarding neuronal response to the stimulus \( \varphi_a \), we adopt the usual top-down
mapping in the FEP [see Eq. (8)] as
\[
\varphi_a = g_a(V; \{\gamma_i\}, \{n_i\}) + z_a, \tag{50}
\]
where \( g_a \) is the generative map that is to be supplied for practical application and \( z_a \)
characterizes the stochastic nature of the sensory reading which we assume the normal
distribution \( \mathcal{N}(\varphi_a - g_a; 0, \sigma_{za}) \). With the present model, we assume that the on-site
neuronal states respond to the sensory data instantly. A more appropriate dynamical
model is to be supplied when a detailed biophysical description is required.

As a working example, here we consider a H-H neuron which allows fast relaxation,
i.e. \( \tau_l \ll 1 \), of gating variables to their steady-states, \( \gamma_l(t) \rightarrow \gamma_l(\infty) = \gamma_{leq}(V) \). In this
case our neural particle is fully characterized by a single dynamical variable of \( V \). Note
the time-dependence of the gating variables occurs only implicitly through the long-
time membrane voltages in Eq. (48). Then, the RD of our neural particle is fulfilled
in a two-dimensional state space spanned by \( V = \mu \) and \( p_V = \varphi \), prescribed by the
Hamiltonian function,
\[
\mathcal{H}(\mu, p; t) = \frac{p^2}{2\sigma_{w}^{-1}} - \frac{1}{2}\sigma_{z}^{-1}(\varphi - g)^2 + pf. \tag{51}
\]
While the ‘dissipative’ function \( f \) is explicitly given in the H-H model as
\[
f(\mu) = \frac{1}{C} \sum_t \gamma_{leq}(\mu)G_t(E_t - \mu) + I_{ex}/C, \tag{52}
\]
the ‘conservative’ function \( g \) must be additionally supplied. Also, one needs to
make the voltage-dependence of \( \gamma_{leq} \) available in practice. Note 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 \( f \).

In Fig. [3] we illustrate the Hamiltonian function Eq. (51) assuming static sensory
data and constant driving currents; where we take the algebraic approximations \[37\] for
the conservative map function,
\[
g(\mu) \approx a_0 + a_1\mu + a_2\mu^2,
\]
and for the voltage-dependence of the gating variables,
\[
\gamma_{leq}(\mu) \approx b_{t0} + b_{t1}\mu + b_{t2}\mu^2.
\]

The Hamilton’s equations of motion, Eqs. \[11\] and \[12\], bring about the nonlinear
RD as
\[
\dot{\mu} = \Lambda_1(\mu, p; t), \tag{53}
\]
\[
\dot{p} = \Lambda_2(\mu, p; t), \tag{54}
\]
Figure 3. Hamiltonian function Eq. (51) in arbitrary units for appropriately chosen sets of parameters; where two trajectories are depicted schematically assuming an initial condition at $\mu, p = (0, 0)$. The actual trajectories must be determined by the solutions to the coupled equations, Eqs. (53) and (54).

where the ‘force’ functions $\Lambda_1$ and $\Lambda_2$ are specified as

$$\Lambda_1 = f(\mu) + \frac{1}{\sigma_{w}^{-1}} p,$$

$$\Lambda_2 = -\sigma_{z}^{-1}(\varphi - g) \frac{\partial g}{\partial \mu} - \frac{\partial f}{\partial \mu} p.$$  

According to the present model, the brain performs the RD, given the sensory input $\varphi$ in the force function $\Lambda_2$ and consequently obtain the optimal trajectory $\mu^*, p^*$ conforming to Eq. (43). In the steady-state limit the brain reaches a fixed (equilibrium) point $(\mu_{eq}, p_{eq})$ in the state space, that is specified by intersections of the isoclines,

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

To gain an insight into how the system approaches to steady states, 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 Eqs. (53) and (54) to the linear order in the deviations $\delta \mu$ and $\delta p$ and, after rearrangement, obtain the normal form,

$$\frac{d}{dt} \begin{pmatrix} \delta \mu \\ \delta p \end{pmatrix} + \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} \delta \mu \\ \delta p \end{pmatrix} = 0.$$  

In Eq. (57) the elements of the relaxation (Jacobian) matrix $R$ are specified to be

$$R_{11} = -\left[ \frac{\partial f}{\partial \mu} \right]_{eq}, \quad R_{12} = -\frac{1}{\sigma_{w}^{-1}}.$$
\[ \mathcal{R}_{21} = \sigma_z^{-1} \left[ -\left( \frac{\partial g}{\partial \mu} \right)^2 + (\varphi - g) \frac{\partial^2 g}{\partial \mu^2} - \frac{\partial^2 f}{\partial \mu^2} \right]_{eq}, \]

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

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

\[ \delta \psi \equiv (\delta) \mu \delta \mu. \]

Then, the formal solution to Eq. (57) is written 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 \( \lambda_\alpha \) and eigenvectors \( \phi_\alpha \) are determined by the secular equation,

\[ \mathcal{R} \phi_\alpha = \lambda_\alpha \phi_\alpha. \]

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

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

where the expansion coefficients \( c_\alpha \) are fixed by the initial condition. The 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 \( \lambda_1 \) and \( \lambda_2 \) would be pure imaginary with opposite sign. Consequently, in the present particular model the resulting equilibrium point is likely to be a center.

4.2. The hierarchical neural network

Here, we suppose that there are a finite number of levels in the perceptual hierarchy and that for simplicity each level is effectively characterized as a single neural node. Further, we assume that the perceptual state of the neural node at hierarchical level \( i \) is specified by the coarse-grained, activation and connection variables, denoted as \( V^{(i)} \) and \( S^{(i)} \), respectively. The activation variable describes action potential at a node and the connection variable describes inter-level synaptic input and output variables. Both variables are derived from a population of neurons and thus vary on a coarse-grained space and time scale. The theoretical details of how one may derive such a coarse-graining description is not our scope, for a reference see [33]. They form the generalized coordinates in brain’s configurational space,

\[ \mu^{(i)} = (V^{(i)}, S^{(i)}), \]
Recognition Dynamics of the Brian in the Free Energy Principle

where the superscript runs \( i = 1, 2, \ldots, M \), 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 \( i \),

\[
\frac{dV^{(i)}}{dt} = f^{(i)}(V^{(i)}, S^{(i)}) + w^{(i)},
\]

which is a direct generalization of Eq. (57) with incorporating the hierarchical dependence via \( S^{(i)} \). For inter-level dynamics, we propose that the connection variables are subject to the similar stochastic equations,

\[
\frac{dS^{(i)}}{dt} = g^{(i+1)}(V^{(i+1)}, S^{(i+1)}) + z^{(i)},
\]

where the brain’s top-down prediction function \( g^{(i)} \) must be supplied in practical implementation, and \( z^{(i)} \) represents the noise associated with the process. Note there is only spontaneous fluctuation at the top cortical level, \( i = M \), accordingly

\[
g^{(M+1)} = 0.
\]

Also, we constrain that the sensory data \( \varphi \) enter the interface (or boundary between) of the brain and the environment and subsequently evoke hierarchical interaction between different levels. We assume that the brain’s sensory reading is performed by the brain variables at the lowest cortical level, \( i = 1 \), via an instantaneous mapping,

\[
S^{(0)} = g^{(1)}(V^{(1)}, S^{(1)}) + z^{(0)},
\]

where for notational convenience we have set

\[
S^{(0)} = \varphi(t).
\]

We remark that the hierarchical equations Eq. (60) we propose is dissimilar to the conventional formulation which assumes the static model in the entire hierarchy like the one Eq. (62) at the sensory interface, see [8]. We treat here the connection variables dynamically not statically in order to treat lateral and hierarchical dynamics symmetrically. The rates of the activation and connection variables may be subjected to different time-scales, 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 nicely.

Having specified our hierarchical model, we write the informational Lagrangian for the constructed neural network by generalizing Eq. (36) with a single sensory input for now, as

\[
\mathcal{L}(V, \dot{V}; S, \dot{S}; \varphi) = \frac{1}{2} \sum_{i=1}^{M} m_{w}^{(i)} (\varepsilon_{w}^{(i)})^2 + \frac{1}{2} \sum_{i=0}^{M} m_{z}^{(i)} (\varepsilon_{z}^{(i)})^2,
\]

where \( m_{w}^{(i)} \) and \( m_{z}^{(i)} \) are the inverse variances associated with the Gaussian noises, \( w^{(i)} \) and \( z^{(i)} \), respectively,

\[
m_{w}^{(i)} = 1/\sigma_{w}^{(i)} \quad \text{and} \quad m_{z}^{(i)} = 1/\sigma_{z}^{(i)}.
\]
The auxiliary variables in the Lagrangian are defined to be \((i \geq 1)\)

\[
\begin{align*}
\varepsilon_w^{(i)} & \equiv \dot{V}^{(i)} - f^{(i)} \left( V^{(i)}, S^{(i)} \right), \\
\varepsilon_z^{(i)} & \equiv \dot{S}^{(i)} - g^{(i+1)} \left( V^{(i+1)}, S^{(i+1)} \right).
\end{align*}
\]

(66)

(67)

We interpret that \(\varepsilon_w^{(i)}\) specifies the discrepancy between the change in the present lateral state and its predicted value on the cortical level \(i\), which may be considered as the lateral-error prediction. On the other hand, \(\varepsilon_z^{(i)}\) measures the prediction error between the change in the present hierarchical state and its predicted value from one-level higher via the generative map \(g\), which may be viewed as the hierarchical-error prediction. Note \(\varepsilon_w^{(0)}\) in the second term on the RHS of Eq. (64) is defined separately as

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

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

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

\[
\begin{align*}
p_v^{(i)} & \equiv \frac{\partial L}{\partial \dot{V}^{(i)}} = m_w^{(i)} \varepsilon_w^{(i)}, \\
p_s^{(i)} & \equiv \frac{\partial L}{\partial \dot{S}^{(i)}} = m_z^{(i)} \varepsilon_z^{(i)},
\end{align*}
\]

(68)

(69)

Note that the inverse variances \(m_w^{(i)}, m_z^{(i)}\) may be termed informational inertial masses which modulate the discrepancy between the change of the states and its prediction. Thus, in our terms, momentum \(p_v^{(i)}\) is a measure of lateral prediction-error modulated by inertial mass \(m_w^{(i)}\), and momentum \(p_s^{(i)}\) is a measure of hierarchical prediction-error modulated by inertial mass \(m_z^{(i)}\). And, the heavier the mass is, the bigger the precision becomes.

Given the Lagrangian Eq. (64), 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, the outcome is obtained as

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

(70)

where the informational kinetic energy is defined to be \((i \geq 1)\)

\[
\mathcal{T}^{(i)}(p_v, p_s) = \frac{1}{2m_w^{(i)}} \left( p_v^{(i)} \right)^2 + \frac{1}{2m_z^{(i)}} \left( p_s^{(i)} \right)^2,
\]

(71)

and the potential energy to be \((i \geq 2)\)

\[
\mathcal{V}^{(i)}(V, p_V; S, p_S; \varphi) \equiv p_v^{(i)} f^{(i)} + p_s^{(i)} g^{(i+1)},
\]

(72)

respectively. The potential energy at the lowest level is specified separately to be

\[
\mathcal{V}^{(1)} = p_v^{(1)} f^{(1)} - \frac{1}{2} m_z^{(0)} \left( S^{(0)} - g^{(1)} \right)^2.
\]

(73)
Note the multi-level Hamiltonian Eq. (70) reduces to Eq. (51) in the single-level approximation.

Next, it is straightforward to generate the Hamiltonian equations of motion for the brain’s perceptual states $\psi^{(i)}$, which we denote as a four dimensional column vector as

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

where $T$ means the transverse operation. The results are $4 \times 4$ coupled differential equations at each level: At the sensory peripheral level ($i = 1$) the coupled recognition mechanics among the perceptual components are explicitly given as

$$\dot{V}^{(1)} = \frac{\partial H}{\partial p_V^{(1)}} = \frac{1}{m_w^{(1)}} p_V^{(1)} + f^{(1)}, \quad (74)$$

$$\dot{p}_V^{(1)} = -\frac{\partial H}{\partial V^{(1)}} = -p_V^{(1)} \frac{\partial f^{(1)}}{\partial V^{(1)}} - m_z^{(0)} (S^{(0)} - g^{(1)}) \frac{\partial g^{(1)}}{\partial V^{(1)}}, \quad (75)$$

$$\dot{S}^{(1)} = \frac{\partial H}{\partial p_S^{(1)}} = \frac{1}{m_z^{(1)}} p_S^{(1)} + g^{(2)}, \quad (76)$$

$$\dot{p}_S^{(1)} = -\frac{\partial H}{\partial S^{(1)}} = -p_V^{(1)} \frac{\partial f^{(1)}}{\partial S^{(1)}} - m_z^{(0)} (S^{(0)} - g^{(1)}) \frac{\partial g^{(1)}}{\partial S^{(1)}}. \quad (77)$$

And, at higher levels ($i \geq 2$) the perceptual mechanics are governed by

$$\dot{V}^{(i)} = \frac{\partial H}{\partial p_V^{(i)}} = \frac{1}{m_w^{(i)}} p_V^{(i)} + f^{(i)}, \quad (78)$$

$$\dot{p}_V^{(i)} = -\frac{\partial H}{\partial V^{(i)}} = -p_V^{(i)} \frac{\partial f^{(i)}}{\partial V^{(i)}}; \quad (79)$$

$$\dot{S}^{(i)} = \frac{\partial H}{\partial p_S^{(i)}} = \frac{1}{m_z^{(i)}} p_S^{(i)} + g^{(i+1)}, \quad (80)$$

$$\dot{p}_S^{(i)} = -\frac{\partial H}{\partial S^{(i)}} = -p_V^{(i)} \frac{\partial f^{(i)}}{\partial S^{(i)}}. \quad (81)$$

Note the equations of motion of momenta $p_V^{(i)}$ and $p_S^{(i)}$ at level $i = 1$ obeys different rules compared with those at higher levels due to the sensory measurement terms, the second terms on the RHSs of Eqs. (75) and (77).

According to the derived RD, the sensory inputs $\varphi$ enter the brain and environment interface, see Eqs. (75) and (77), which is predicted by the brain’s lowest level generative model $g^{(1)}(V^{(1)}, S^{(1)})$. And, the prediction discrepancy, $\varphi - g^{(1)}$ becomes a source of induced change in the perceptual momenta. The perceptual momenta $p_V^{(1)}$ and $p_S^{(1)}$ enter Eqs. (74) and (76) as the lateral and hierarchical prediction-errors, respectively, which, in turn, generate a change in the perceptual states $V^{(1)}$ and $S^{(1)}$, respectively, together with the corresponding generative models $f^{(1)}$ and $g^{(2)}$. In particular, the top-down prediction $g^{(2)}$ from one higher-level participates in organizing the change of $S^{(1)}$ [see Eq. (76)]. At higher levels $i \geq 2$, the inter-level top-down prediction $g^{(i+1)}$ and on-level prediction errors $p_S^{(i)}$ determine the change of the current hierarchical state $S^{(i)}$ through Eq. (81). And, Eqs. (78), (79), and (81) govern the coupled, lateral dynamics to reorganize on-level perceptual states $(V^{(i)}, S^{(i)})$. In Fig. 4 we draw a diagram...
that schematically illustrates the perceptual architecture of the hierarchical network implemented by Eqs. (74)-(81).

Here, we emphasize that the dynamics of the prediction errors, via canonical momenta, are encapsulated in our formulation on an equal footing with the dynamics of the state variables. This is also in contrast to the conventional minimization algorithm where only differential equation for the brain states are present and consequently the temporal update of the prediction errors themselves are not present. An attempt to incorporate the brain’s computation of prediction errors in the FEP can be found in a recent tutorial article [38].

For mathematical compactness, here we rewrite Eqs. (74)-(81) as

$$\frac{d\psi_{\alpha}^{(i)}}{dt} = \Lambda_{\alpha}^{(i)} (\{\psi_{\alpha}^{(i)}\}),$$  (82)

where the hierarchical index $i$ runs from 1 to $M$, $\alpha$ runs from 1 to 4, and the force function $\Lambda_{\alpha}^{(i)}$ is the corresponding RHS to each vector component $\psi_{\alpha}^{(i)}$ at cortical level $i$. The obtained hierarchical equations are the highlight of our theory, prescribing the RD of the brain’s sensory inference within the FEP framework.
To apply our formulation to an empirical brain, one needs to supply the generating function $f^{(i)}$ of lateral dynamics and the hierarchical connecting function $g^{(i)}$, that enter the force functions $\Lambda^{(i)}_\alpha$ in the perceptual mechanics, Eq. (82). For the generating function we once again use the H-H model Eq. (48) with an explicit form for the driving current,

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

where $\tilde{G}_l$ are the channel conductances normalized by the capacitance $C$, and the second term on the RHS accounts for the leakage and/or lateral synaptic currents as an external source $I_{ex}$ with $\tilde{G}_S$ being the normalized synaptic conductance. The hierarchical connection function may be modeled in a simplest form as

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

where the function $\chi$ specify the voltage-dependent synaptic plasticity from hierarchical level $i$ to level $i-1$. In addition, as in the single particle case, one must supply approximate models for voltage-dependence of the gating variables $\gamma_{l eq}$ and the connection strength $\chi$. For instance, one may take the quadratic approximations [37],

$$\gamma_{l eq}(V^{(i)}) \approx b_0 + b_1 V^{(i)} + b_2 V^{(i)} V^{(i)},$$

$$\chi(V^{(i)}) \approx a_0 + a_1 V^{(i)} + a_2 V^{(i)} V^{(i)}.$$

At the end of the RD an optimal trajectory is obtained in multi-dimensional phase space,

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

which tends to a fixed point in the steady-state limit, $\psi^{(i)}_{\alpha,eq} = \psi^{*({i})}_\alpha(t \to \infty)$. The necessary equilibrium condition to Eq. (82) is specified by

$$\Lambda^{(i)}_\alpha(\{\psi^{(i)}_\alpha\}) = 0.$$

While 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 $\alpha^{th}$ component of the perceptual state vector $\psi^{*({i})}_\alpha$ at the cortical level $i$, $\delta\psi^{(i)}_\alpha$, from the fixed point $\psi^{(i)}_{\alpha,eq}$,

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

Then, we expand Eq. (82) about the fixed point to linear order in the small deviation, and after some manipulation we get the hierarchical equations for $\delta\psi^{(i)}_\alpha$,

$$\frac{d\delta\psi^{(i)}_\alpha}{dt} + \sum_{\beta=1}^4 R^{(i)}_{\alpha\beta} \delta\psi^{(i)}_\beta = \sum_{\beta=1}^4 \sum_{j \neq i}^M \Phi^{(i)}_{\alpha\beta} \delta\psi^{(j)}_\beta;$$

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

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

and
and the inter-level connection between level $i$ and level $j$ in the hierarchical pathway is specified by
\[
\Phi_{ij}^{(ij)} = \left[ \frac{\partial \Lambda^{(j)}_{i}}{\partial \psi^{(j)}_{i}} \right]_{eq};
\]
where the subscript $eq$ indicates that the matrix elements are to be evaluated at the equilibrium (fixed) 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 $\Phi_{ij}^{(ij)}$ do not vanish only for $\alpha = 3$ because only the force function $\Lambda^{(i)}_{1}$ possesses $\psi^{(j)}_{i}$ for $j \neq i$ as variables via $g^{(i+1)}$ [see Eq. (80)]. Second, because $g^{(i+1)}$ depends solely on the hierarchical level-index $i + 1$, only matrix elements with the hierarchical index $j = i + 1$ survives. Combining these two observations, the source term on the RHS of Eq. (86) is converted into a vector at level $i + 1$ with only a single nonvanishing $\alpha = 3$ component,
\[
\sum_{\beta=1}^{4} \sum_{j \neq i}^{M} \Phi_{ij}^{(ij)} \delta \psi^{(j)}_{\beta} = \delta \zeta^{(i+1)}_{\alpha},
\]
which to be complete we spell out explicitly as
\[
\delta \zeta^{(i+1)}_{\alpha} = \delta \alpha_{3} \left\{ \left[ \frac{\partial g^{(i+1)}}{\partial \psi^{(i+1)}_{1}} \right]_{eq} \delta \psi^{(i+1)}_{1} + \left[ \frac{\partial g^{(i+1)}}{\partial \psi^{(i+1)}_{3}} \right]_{eq} \delta \psi^{(i+1)}_{3} \right\};
\]
where $\delta \alpha_{3}$ is the Kronecker delta.

Finally, we shall present a formal solution to the linearized perceptual mechanics Eq. (86), that can be obtained by a direct integration with respect to time. The result takes the form
\[
\delta \psi^{(i)}(t) = e^{-R^{(i)} t} \delta \psi^{(i)}(0) + \int_{0}^{t} dt' e^{-R^{(i)} (t-t')} \delta \zeta^{(i+1)}(t').
\]
We next solve the eigenvalue problem at each hierarchical level, which is defined to be
\[
R^{(i)} \phi^{(i)}_{\alpha} = \lambda^{(i)}_{\alpha} \phi^{(i)}_{\alpha},
\]
where $\lambda^{(i)}_{\alpha}$ and $\phi^{(i)}_{\alpha}$ 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:
\[
\delta \psi^{(i)}(0) = \sum_{\alpha} a^{(i)}_{\alpha} \phi^{(i)}_{\alpha}.
\]
Similarly, we may expand the inhomogeneous vector $\delta \zeta^{(i+1)}$ as
\[
\delta \zeta^{(i+1)}(t') = \sum_{\alpha} b^{(i+1)}_{\alpha}(t') \phi^{(i)}_{\alpha},
\]
where note the expansion coefficients $b^{(i+1)}_{\alpha}$ are time-dependent. By substituting the expansions Eqs. (90) and (91) into Eq. (88) we obtain the desired formal solution near equilibrium points,
\[
\delta \psi^{(i)}(t) = \sum_{\alpha=1}^{4} a_{\alpha} e^{-\lambda^{(i)}_{\alpha} t} \phi^{(i)}_{\alpha} + \sum_{\alpha=1}^{4} \phi^{(i)}_{\alpha} \int_{0}^{t} dt' e^{-\lambda^{(i)}_{\alpha} (t-t')} b^{(i+1)}_{\alpha}(t').
\]

The geometrical approach to a fixed point is again determined by the eigenvalues \( \lambda^{(i)} \); however, the details are driven by the time-dependent generative sources \( b^{(t+1)}(t) \) from one-level higher in the hierarchy.

To sum, responding to sensory streaming \( \varphi = S^{(0)} \) at the brain-environment interface, the lowest hierarchical level \( i = 1 \), the brain in an initially resting state performs the hierarchical RD by integrating Eq. (82) to infer the external causes. The ensuing brain’s computation corresponds to minimizing the IA, which is an upper bound of the sensory uncertainty in the FEP whose mathematical statement is repeated compactly as

\[
H[p(\varphi)] \leq S[F; \varphi],
\]

where the sensory uncertainty \( H \) is defined in Eq. (1) and the IA \( S \) is explicitly given again as \( S[F; \varphi] = \int dtF(\\psi_{a}^{(i)}(; \varphi). \) Conforming to our restated FEP, the minimum value of IA specifies the tightest bound of the sensory uncertainty over a relevant biological time-scale.

5. Discussion

We have recast the FEP into that all living organisms are evolutionally self-organized to tend to minimize the sensory uncertainty about uninhabitable environmental encounters. The sensory uncertainty is an average of the surprisal over the sensory distribution induced on the brain-environment interface, the sensory surprisal being the self-information of the sensory distribution. 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 and defines the theoretical action in the least action principle. In this paper we have enunciated that the minimization of the IFE must continually take place over a finite temporal horizon of an organism’s experiencing environmental interaction. Our scheme is in contrast to the conventional theory which approximates minimization of the IFE at each point in time. The sensory uncertainty is an information-theoretical Shannon entropy [39]; however, we have intentionally circumvented the term, ‘sensory entropy’ in this work to call the sensory uncertainty. The reason is that ‘minimization of the sensory entropy’ is reminiscent of Erwin Schrödinger’s word, ‘negative entropy’ which carries a disputable connotation in implying how the living organism avoids decay. He subsequently suggested FE instead as a more appropriate notion in the context [1]. Conforming to the second law of thermodynamics, the organism’s autopoiesis of minimizing the sensory uncertainty must contribute to increase of the total entropy of the brain and its environment.

We have assumed that the brain employs the Laplace-encoded IFE as an informational Lagrangian in implementing minimization of the IA. And, by admitting only the standard Newtonian position and velocity as relevant dynamical variables, we have reckoned the IFE as a function of the brain variables and their first-order
time derivatives. Our approach reveals that the nonstandard physical construct of the generalized motion is not necessary and thus can be dispensed with in formulating the FEP. The features of the continually changing world are recorded in the brain as time-dependent sensory data: It is not necessary to introduce the extra-physical dynamical model of the generalized coordinates. Also, in our theory all the parameters in the perceptual mechanics are specified in the Hamiltonian; thus, there requires no extra parameters like learning rates in the gradient descent scheme to control the speed of convergence to a steady state. 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 perceptual mechanics of reaching a desired fixed-point or a limit cycle, in general, it remains resting, i.e. being spontaneous, until another sensory stimulus will come in.

We have premised the top-down rationale of sensory inference in our formalism, which is an important facet of the FEP: The sensory inputs at the interface, the lowest hierarchical level, were assumed to be instantaneously mapped onto the brain variables with an associated noise, which is not fully biophysical yet. However, differently from the static model at the lowest level, we have assumed that the inter-level filtering in the brain’s functional hierarchy obeys the stochastic dynamics. The resulting RD are deterministic and hierarchical: The current-level synaptic states are predicted by the activation and synaptic states at one-higher level built in the internal generative model, top-down prediction of the brain, and are, in turn, updated by on-level prediction error, canonical momentum. The momenta, specifying the prediction errors, also obey their own dynamical equations, which is dissimilar from the conventional formulation where the rates of the prediction errors are not drawn explicitly. The on-level, activation states and momenta are governed by the intra-level dynamics of the perceptual states. The obtained RD tenably underpins the causality: For a specified set of the perceptual positions and corresponding momenta at the outset, responding to sensory inputs that may be slow or fast time-dependent, the RD can be integrated. The arbitrariness, for instance, involved in specifying the number of generalized coordinates and corresponding initial conditions can be removed.

In short, it is still a long way to understanding how the Bayesian FEP in neurosciences may be made congruous with the biophysical reality of the brain at an emergent level. Our theory delivers a hybrid of the biologically plausible information-theoretic framework of the FEP and the mechanical formulation of the process theory of the RD resorting to the principle of least action. To borrow what Hopfield puts in words, “it lies somewhere between a model of neurobiology and a metaphor for how the brain computes” [40]. We hope that our effort of trying to shorten the distance will guide a step forward to unraveling the challenging problem.

References

[1] Schrödinger, E. (1967). What is Life? Mind and Matter. Cambridge: Cambridge University Press.
von Helmholtz, H. (1962). Tractise on physiological optics. Vol. III. Mineola: Dover Publication, Inc.

Gregory, R. L. (1980). Perceptions as hypotheses. Philosophical Transactions of the Royal Society of London B. 290: 181-197.

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

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 Royal Society Interface. 10: 20130475.

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

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. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free energy formulation. Biological Cybernetics. 102(3): 227-260.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. Synthese. 159: 417-458.

Friston, K., Stephan, K., Li, B., & Daunnizeau, J. (2010). Generalized Filtering. Mathematical Problems in Engineering. 261670.

Friston, K. J. (2008). Variational Filtering. NeuroImage. 41: 747-766.

Friston, K. (2008). Hierarchical models in the brain. PLoS Computational Biology. 4(11): e1000211.

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

Sprott, J. C. (1997). Some simple chaotic jerk functions. American Journal of Physics. 65: 537-.

Landau, L. P. (1998) Classical Mechanics. (2nd ed.). New York: Springer-Verlag.

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference?. PLoS One. 4(7): e6421.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., & Kennedy, H. (2014). Anatomy of hierarchy: Feedforward and feedback pathways in Macaque visual cortex. Journal of Comparative Neurology. 522:225-259.

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.

Markov, N. T., & Kennedy, H. (2013). Current Opinion in Neurobiology. 23:187-194.

Friston, K. & Kiebel, S. (2009). Cortical circuits for perceptual inference. Neural Networks. 22: 1093-1104.

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Frontiers in Psychology. 3: 151.

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.

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.

Hille, B. (2001). Ion Channels of Excitable Membranes. (3rd ed.). Sunderland: Sinauer Associates, Inc.

Einevoll, G. T., Kayser, C., Logothetis, N. K., & Panzeri, S. (2013). Nature Reviews. Neuroscience. 14:770.

Jansen, B. H., Zouridakis, G., & Brandt, E. (1993). A neurophysiologically-based mathematical model of flash visual potentials. Biological Cybernetics. 68: 275-283.
[30] Jirsa, V. K., & Haken, H. (1996). Field theory of electromagnetic brain activity. Physical Review Letter. 77: 960-963.

[31] 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.

[32] David, O., Friston, K. J. (2003). A neural mass model for MEG/EEG: coupling and neuronal dynamics. Neuroimage. 20: 1743-1755.

[33] 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 Compt. Biol. 4: e1000092.

[34] 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.

[35] 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.

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

[37] Wilson, H. R. (1999). Simplified dynamics of human and mammalian neocortical neurons. J. Theor. Biol. 200: 375-388.

[38] Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. Journal of mathematical psychology, 76 (B): 198–211.

[39] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal. 27: 379-423, 623-656.

[40] Hopfield, J. J. (1999). Brain, neural network, and computation. Review of modern physics. 71: S431-S437.