

Free energy and inference in living systems

Chang Sub Kim

Department of Physics, Chonnam National University, Gwangju 61186, Republic of Korea

E-mail: cskim@jnu.ac.kr

Abstract. Organisms are nonequilibrium stationary systems self-organized from spontaneous symmetry breaking in the environment and undergo irreversible work cycles without detailed balance. The thermodynamic free-energy principle (FEP) describes an organism's homeostasis as regulating metabolic work constrained by the physical FE cost. In contrast, the recent neuroscience and theoretical biology efforts account for an organism's adaptive fitness as purposeful allostasis by demanding minimization of the informational FE. For a coalesced view of life, this study proposes a physically-principled FE minimization theory overarching the essential features from both FEPs. Consequently, the ensuing Bayesian mechanics reveals that the brain function of perception and behavioral control operates like Schrödinger's clockwork and develops optimal trajectories in neural manifolds when the sensory perturbation is aroused. Furthermore, the sensory and motor interactions of the biological agents with the environment are manifested to induce a dynamic transition between neural attractors.

Keywords Living systems, homeostasis and allostasis, Bayesian brain, free-energy principles, Schrödinger's clockwork, neural attractors

29 March 2022

Highlights

- The thermodynamic and neuroscientific free-energy principles (FEPs) are elucidated to provide a more physically grounded way of implementing FE minimization in living systems.
- The brain-inspired, neuroscientific FE is proposed to be an informational Lagrangian, and the recognition dynamics in the brain conducts the resulting Bayesian mechanics like Schrödinger's clockwork.
- The formation of cognitive trajectories in the neural phase space is investigated, and the dynamic transition between stationary attractors is manifested.
- The attractor dynamics fulfills the brain's inference of the environmental causes of sensory influx via updating the internal models and adjusting motor effectuation.

1. Introduction

While there is no standard definition of life at present [64, 43, 37, 27, 25, 24], it is often said that a living system tends to minimize its entropy, defying the second law of thermodynamics in order to sustain its nonequilibrium (NEQ) presence. However, **organisms are an open system in the environment**; hence, conforming to the second law of thermodynamics, the adjudication between the increase and decrease of the organism's entropy must **depend on the direction of heat flux**. The artless confusion, perhaps, is rooted in Erwin Schrödinger's annotation, which metaphorically explains living organisms as feeding on negative entropy [75]. In the same monograph, Schrödinger continues to explicate that **a more appropriate discussion for metabolism is to be led in terms of free energy (FE)**. He made the clarification because, in contrast to the Clausius entropy to which he was referring, **thermodynamic FEs always decrease in irreversible processes of any open system** [8]. In many ways, the current research has inherited Schrödinger's insight into how biological systems can be accounted for from physical laws and principles. This study assesses the potential answers to the question of **what life is in terms of FE minimization**.

Living systems possess the **ability to maintain biologically essential properties** such as body temperature, blood pressure, and glucose levels, which **are distinct from ambient states** and constitute housekeeping states. While being animate, they continually exchange heat and material fluxes with the **environment** by performing metabolic work to sustain homeostasis, governed by the first law of thermodynamics, which describes energy conservation. **The second law posits that the entropy of an isolated macroscopic system increases monotonically with any spontaneous internal changes**. **Organisms and the environment** conjointly constitute the **biosphere**, which, as a whole, **is isolated and macroscopic**; thus, the total entropy must increase by the organism's metabolism, not the organism's entropy alone. The second law affects organisms by limiting the efficiency of metabolic performance. The thermodynamic free-energy principle (TFEP) encompasses thermodynamics laws and provides qualitative and quantitative

answers as to how living systems biophysically sustain homeostasis by minimizing the FEs. However, organisms' ability to undergo allostasis, which predictively regulates homeostasis [81, 76], or autopoiesis in a more general term [58], cannot be accounted for within the TFEP. **Allostatic ability is the brain function of adaptive fitness**, which cannot be resorted solely to a (bio)physical self-organization from a myriad of emergent possibilities in the primitive circuits of neuronal activities. Natural selection must take part and place environmental interaction into survival optimization over the evolutionary time scale. Namely, the FE minimization scheme requires some top-down or high-level computational mechanism to incorporate the hardwired allostatic capability. Recent studies have addressed the modern perspective of metabolism as an energy regulation of multisensory integration across both interoceptive and exteroceptive processes [13, 68], accounted for metabolism not only in individual organisms but also at the ecosystem and planetary levels [46, 38], and also emphasized the energetics and **power efficiency in brain performance** [78, 5, 55].

The brain-inspired FEP in neuroscience and theoretical biology offers a universal view of the organisms' autopoiesis in an axiomatic way; specifically, it provides the informational FE-minimization formalism that accounts for perception, learning, and behavior of living systems [29, 30]. The principle successfully applies to other cognitive systems such as artificial intelligence and robots [10, 57, 73, 11, 60, 59, 21]; however, the present investigation primarily focuses on living systems and the FEP implication in a biological context. Articulating the informational FEP (IFEP), **all life forms are evolutionarily self-organized to tend to minimize 'surprisal', which is an information-theoretic measure of the improbability of organisms' environmental niche.** The informational FE (IFE) is a mathematical measure, not a physical thermodynamic FE in terms of pressure, volume, chemical potential, and etc., that bounds the surprisal from above; accordingly, the IFEP suggests natural selection as the organism's minimizing the IFE as a proxy for the surprisal at all biological time scales. **The IFEP inherits Helmholtz's early idea of perception as unconscious inference [87]: the organism's brain possesses an internal model of sensory generation and infers the external causes of sensory data by matching them with prior knowledge.** The active-inference framework following from the **IFEP** encapsulates motor control and planning as well, beyond Helmholtzian perception, as an additional inferential scheme [28, 1]. The generative and prior models are probabilistically **encoded by the brain variables in the NEQ stationary state; at present, however, there is no physical theory that can be used to rigorously determine the NEQ probabilities in the stationary brain.** In practice, the IFEP assumes open forms or some fixed forms for the NEQ densities and implements the IFE minimization. **The Gaussian fixed-form assumption technically allows the IFE to be converted to a sum of the discrepancies between predicted and actual signals [9], which is termed 'prediction error' in the predictive coding theory [42].** Commonly, the transformed IFE objective is minimized by employing the gradient-descent method developed in machine learning [82]. The resulting filtering equations compute the Bayesian inversion of sensory data inferring the external sources,

termed recognition dynamics (RD) [29]. Recently, the IFEP was generalized in a manner that minimizes ‘sensory uncertainty’, which is a long-term surprisal over a temporal horizon of an organism’s unfolding environmental niche [50]. Despite its promising position as a universal biological principle, the IFEP has also attracted lively discussions on its distance between biophysical reality and epistemological grounds [26, 51, 12, 69, 74, 52, 7, 2].

This study combines the two FE approaches to living systems to develop a more coalesced paradigm: the TFEP does not provide a description of the brain’s ability to sense and act in the world, while the brain-inspired IFEP is mainly a teleological (hypothesis-driven) framework lacking the intimate connections to neuronal substrates and physical laws. Our goal is to link the two FEPs and thus propose the biological FEP as integrating the reductionistic base and top-down inferential architecture in the brain. In addition, we unveil the attractor dynamics that computes allostatic regulation – perception and motor inference – in the brain, emerging from our proposed FE-minimization scheme and the ensuing Bayesian dynamics. We recognize other comprehensive work that makes the connection between reductionism and emergentism embedded in the FEP but without the direct link between the thermodynamic and informational FEs [31].

At the constitutional level of brain anatomy, enormous degrees of freedom pose a high-dimensional problem; thus, classical negligence is inevitable, and the underlying dynamics appears to be stochastic. However, we argue that the brain work of perception, learning, and motor-control is low dimensional at the functional level, obeying the law of large numbers; accordingly, RD becomes deterministic, involving only a limited number of latent variables. For instance, a few joint angles suffice for the neural observer to infer the arm movement in motor control. In contrast, the emergence of deterministic RD is more intriguing in perception and learning, which demands a systematic coarse-graining of stochastic neuronal dynamics. Our investigation contributes to systematically deriving the RD of the Bayesian brain in terms of a few effective variables, which we term Bayesian mechanics (BM) in the brain; BM accounts for the homeostasis and allostasis (that is, adaptive fitness) of living systems, conforming to the proposed biological FEP.

The concept of coarse-graining or effective description is ubiquitous in computational neurosciences and has been identified in different aspects [41, 4, 71, 22, 34, 85]. Here, we recount research most relevant to our work, which drove our motivation for developing BM. Many existing studies on recorded neurons show that the population dynamics is confined to a low-dimensional manifold in the empirical neural space, where trajectories are the neural representations of the population activity revealing the dimensionality reduction [20]. In mathematical terms, the neural modes were defined to be eigen-fields that span the neural manifold, and the latent variables or collective coordinates to be the projection of the population activity onto the neural modes [36, 35]. Other theoretical models support the idea that the long-term dynamics in recurrent neural networks gives rise to the attractor manifold [61], which is a continuous set of fixed points occupying a limited region of the neural space. Consequently, the attractor

dynamics and the switching between different continuous attractors were manifested [62], which implies a contextual change in the neuronal representations [89, 47]. The manifold hypothesis is also widely exercised in machine learning to approximately understand high-dimensional data in terms of a small number of parameters [15]. On the experimental side, it was reported that a dynamical collapse occurs in the brain from incoherent baseline activity to low-dimensional coherent activity across neural nodes [79, 67, 86]. Synchronized patterns emerged when the featured inputs and the prediction derived from prior or stored knowledge matched, whereas, when there was a mismatch, the high-dimensional multi-unit activity increased. This observation also provides empirical evidence of neural signals reducing the prediction errors, thus minimizing the IFE.

We adopt the above-described ideas and findings and suggest that the latent dynamics can be effectively described by a small number of coarse-grained variables in the reduced dimension. In this study, we formulate the BM of the living systems' conducting the inferential regulation of homeostasis in terms of a few latent variables. The latent variables are unfolded as the brain activities and their conjugate momenta that represent the external – environmental and motor – states and the online prediction errors, respectively. We note that the sensory error at the peripheral level acts as a time-dependent driving source in the BM, which provides the neural mechanism for sensory inference and adaptive motor control. Our continuous-state formulation in continuous time may well serve to study situated-action problems in which biological systems must make decisions even during ongoing sensorimotor activity [14].

The remainder of this paper is organized as follows. In Sec. 2, we recapitulate what the TFEP – the FTs – entails when applied to NEQ living systems. In Sec. 3, we convey how stochastic dynamics at the neuronal level can be modeled and how one may take a statistical approach to determining the NEQ densities of the neural states in the physical brain. In Sec. 4, we present the proposed biological IFEP minimizing the long-term sensory uncertainty and establish its continuous-state implementation that yields the BM in the neural phase space. Next, in Sec. 5, we numerically integrate the BM and manifest the attractor dynamics that performs perception and motor inference in the brain. Finally, we summarize the significant outcomes from our investigation and provide a conclusion in Sec. 6.

2. Nonequilibrium fluctuation theorems applying to organisms

Recently, it was established that fluctuation theorems (FTs) explicate stochastic NEQ processes compactly in terms of mathematical equalities [44, 18]. Although initially FTs were established for small systems, where fluctuations are appreciable, it was shown that they also apply to macroscopic deterministic dynamics [48]. Here, we present FTs in an appropriate context of biological problems and pose that the FTs suggest a living organism as an NEQ system that maintains the *housekeeping temperature* T , which is on average 36.5 degrees in Celsius for humans, within its body and enacts a metabolism

isothermally against its environment.

To this end, we borrow, among the various representations of FTs, the NEQ work relation:

$$\langle e^{-\beta(W-\Delta F)} \rangle = 1, \quad (1)$$

where $\beta = k_B T$, with k_B being the Boltzmann constant and T being the temperature as discussed below. The integral form of the FT given in Eq. (1) is known as the Jarzynski equality (JE) [77]. In the biological context, W is the amount of environmental work involved in the metabolism of living matter, for example, the biological reactions of oxygenic photosynthesis and aerobic respiration [3, 83], and ΔF is the induced change of the Helmholtz FE in the system. Accordingly, $W - \Delta F$ is the excess energy associated with each irreversible work process on the system, unavailable for a useful conversion. The bracket $\langle \dots \rangle$ indicates the average over many work strokes, that is, work distribution subject to a protocol, which is required because the environmental work is not controllable and is thus stochastic, although the biological systems are macroscopic. Note here that we shall consider the temperature appearing in the JE as the body temperature of the considered biological system, unlike the usual implication of FTs; in the standard derivation of the JE [45], the temperature T appearing in the NEQ equality is, by construction, the reservoir temperature. The FT generally applies to non-isothermal processes during which the system temperature may not be defined under the irreversible condition. However, the initial and end states must be in equilibrium so that the free energy is meaningful. The subtlety lies in the fact that the end-state temperature may or may not be the same as the reservoir temperature for experiments performed in isolation after the initial equilibrium preparation at T . Living organisms are in an NEQ stationary state, maintaining a housekeeping temperature T , distinct from the ambient temperature, to which they equilibrate only when ceasing to exist. In this sense, we view organisms as an isothermal system, which is *open* to heat and particle exchange with the environment.

By $\langle e^{-\beta W} \rangle \geq e^{-\beta \langle W \rangle}$ in Eq. (1), which is known as the Jensen inequality in the literature [17], we get the inequality $\Delta F \leq \langle W \rangle$. This inequality is an alternative expression of the application of the second law to *isothermal* irreversible processes of the system initially prepared in equilibrium with a reservoir. In turn, using the inequality, one can consider the change in the average total entropy: $\langle \Delta S_{tot} \rangle = \langle \Delta S_{sys} \rangle + \langle \Delta S_R \rangle$, where ΔS_{sys} is the change in the system entropy and ΔS_R is the change in the reservoir entropy. The average associated with ΔS_R , which is reversible by definition, can be further manipulated as $\langle \Delta S_R \rangle = -\langle Q_{sys} \rangle / T = (\langle W \rangle - \Delta U) / T$, where $Q_R = -Q_{sys}$ was used in the first step, and subsequently the thermodynamic first law was applied for $\langle Q_{sys} \rangle$; U is the internal energy of the system. Then, it follows that $T \langle \Delta S_{tot} \rangle = \langle W \rangle - (\Delta U - T \langle \Delta S_{sys} \rangle) = \langle W \rangle - \Delta F$, which leads to the common expression of the second law for stochastic systems:

$$\langle \Delta S_{tot} \rangle \geq 0. \quad (2)$$

The unavailable energy associated with individual work processes amounts to a total entropy change, namely $\beta(W - \Delta F) = k_B^{-1} \Delta S_{tot}$ under the isothermal condition. By applying the final identity to Eq. (1), the JE can be cast to the entropy-fluctuation theorem:

$$\langle e^{-k_B^{-1} \Delta S_{tot}} \rangle = 1. \quad (3)$$

In the biological context, the FTs prescribe the imbalance between the energy intake and expenditure in an organism maintaining the *housekeeping temperature* before and after the metabolic work stroke. The Helmholtz FE increment in the system is limited by the average environmental work delivered to the organism. The resulting inequality from the JE can be written in the agent-centric form as

$$\langle \mathcal{W} \rangle \leq \Delta \mathcal{F}, \quad (4)$$

where we set $\langle \mathcal{W} \rangle = -\langle W \rangle$ and $\Delta \mathcal{F} \equiv -\Delta F$, which now states that the *work performance* $\langle \mathcal{W} \rangle$ of a biological agent against the environment (e.g., via metabolism) is bounded from above by the thermodynamic FE cost $\Delta \mathcal{F}$. Rare individual processes that violate Eq. (4) may occur in small systems; however, such a statistical deflection is not expected in a finite biological system with macroscopic degrees of freedom. The equality in Eq. (4) holds for reversible changes in inanimate matters; however, this will not occur in the metabolic processes of living organisms.

The NEQ work relation expresses the second law of thermodynamics as the mathematical equality in Eq. (1), and the second law in its biological context entails the thermodynamic constraint on living organisms given by the inequality in Eq. (4), which reveals the unavailable thermodynamic FE expenditure in performing isothermal metabolism. However, it accounts for neither self-adaptiveness nor the brain functions such as perception, learning, and behavior. To address these essential mechanisms of life, researchers currently employ a hybridizing scheme, which first proposes how the system-level biological functions may operate and then attempts to make connections to biophysical substances. In particular, the Bayesian logic built into the informational FEP in neuroscience provides the crucial aspect in this promising hybrid explanation of life.

3. Statistical-physical description of the nonequilibrium brain

The brain comprises a myriad of complex neurons; accordingly, its *internal dynamics* at the mesoscopic level must obey some stochastic equations of motion on account of *classical indeterminacy*. The relevant, coarse-grained neural variables are the local-scale neuronal activities or the inter-area brain rhythms. In the following, we consider that the *brain matter itself* constitutes the *internal milieu* at body temperature for the brain's neural activities.

Below, we use the notation μ to denote such an *internal state* and assume that the neural activity obeys the Langevin dynamics given by [53]

$$\frac{d\mu}{dt} = f(\mu; t) + w(t), \quad (5)$$

where the inertial term was dropped in the over-damping limit. Here, f may encapsulate both conservative and time-dependent metabolic forces, and w represents the random fluctuation characterized as a delta-correlated Gaussian noise satisfying

$$\langle w(t) \rangle = 0, \quad (6)$$

$$\langle w(t)w(t') \rangle = I\delta(t - t'), \quad (7)$$

where I is the noise strength. In one dimension (1D), for simplicity, the environmental perturbation and noise strength are physically specified, respectively, as [70]

$$f = \frac{1}{m\gamma}A \quad \text{and} \quad I = 2\frac{k_B T}{m\gamma},$$

where A represents a conservative force acting on a biophysical neural unit with mass m neglecting time-dependent driving, T indicates the body temperature, and γ is the phenomenological frictional coefficient whose inverse corresponds to momentum relaxation time. The solutions to Eq. (5) describe individual trajectories of the random dynamical processes in the over-damping limit, which neglects the inertial term in the general stochastic equations of motion.

In general, one may consider colored noises beyond the delta-correlated white noise by generalizing Eq. (5) to incorporate the non-Markovian memory effect as

$$m \int_{-\infty}^t dt' \gamma(t - t') \dot{\mu}(t') = A(\mu) + \zeta.$$

To ensure equilibrium at temperature T , the colored Langevin equation must satisfy the fluctuation-dissipation theorem that admits a nonsingular noise correlation [90]:

$$\langle \zeta(t)\zeta(t') \rangle = 2k_B T \gamma(|t - t'|).$$

A standard example of such colored noise is given by the Orstein-Uhlenbeck memory kernel given by $\gamma(|t - t'|) = \gamma\tau^{-1} \exp(-|t - t'|/\tau)$, where τ is the noise autocorrelation time.

As an alternative description to the Langevin equation [Eq. (5)], one may consider an ensemble of identical systems displaying various values of the state μ collectively and ask how the statistical distribution changes over time. After normalization, the ensemble distribution is reduced to the probability density, say $p(\mu, t)$, so that $p(\mu, t)d\mu$ specifies the probability of an individual Brownian particle to be found in the range $(\mu, \mu + d\mu)$ at time t . In the Markovian approximation, the change in the probability density is determined by the probability density at the current time, which is generally described by the master equation given in the continuous-state formulation as

$$\frac{\partial p(\mu, t)}{\partial t} = \int \left\{ w(\mu, \mu') p(\mu', t) - w(\mu', \mu) p(\mu, t) \right\} d\mu', \quad (8)$$

where $w(\mu', \mu)$ is the transition rate of the state change from μ to another μ' . We further assume that the transition takes places between two infinitesimally close states μ and μ' , where $\mu' - \mu = x \ll 1$, so that the transition rate is sharply peaked around $x = 0$ to approximate it as $w(\mu', \mu) \approx w(\mu, x)$. Then, one may expand $p(\mu')$ about μ to the

second-order in x and neglect all higher-order terms. Consequently, the master equation can be converted into the Smoluchowski-Fokker-Planck (S-F-P) equation [70]:

$$\frac{\partial p(\mu, t)}{\partial t} = \frac{\partial}{\partial \mu} \left\{ -D_1(\mu) + \frac{\partial}{\partial \mu} D_2(\mu) \right\} p(\mu, t), \quad (9)$$

In Eq. (9), D_1 and D_2 correspond to the first two expansion coefficients in the Kramers-Moyal formalism, which are determined in the present case to be

$$D_1 = f \quad \text{and} \quad D_2 = \frac{1}{2}I.$$

The S-F-P equation is written in three dimensions (3D) as

$$\frac{\partial p(\vec{\mu}, t)}{\partial t} + \nabla \cdot \left\{ \vec{f}(\vec{\mu}) - D \nabla \right\} p(\vec{\mu}, t) = 0, \quad (10)$$

where ∇ is the gradient operation with respect to the three-dimensional state $\vec{\mu}$. In Eq. (10), the drift term $p\vec{f}$ accounts for the conservative potential forces. In addition, we denote the diffusion term D_2 as D , assuming spatial isotropy, for simplicity and notational convention.

The S-F-P equation describes a local conservation of the probability $p(\vec{\mu}, t)$ in the state space spanned by the state vector $\vec{\mu}$, which carries the probability flux \vec{j} identified as

$$\vec{j}(\vec{\mu}, t) = p(\vec{\mu}, t) \vec{f}(\vec{\mu}) - D \nabla p(\vec{\mu}, t).$$

In the steady state (SS), $\partial p_{st}/\partial t = 0$, where $p_{st} \equiv p(\mu, \infty)$; accordingly, the divergence of the SS flux $\vec{j}_{st} \equiv \vec{j}(\mu, \infty)$ must vanish in the S-F-P equation:

$$\nabla \cdot \vec{j}_{st} = 0. \quad (11)$$

If the Brownian particles undergo motion in an isolated or infinite medium, \vec{j}_{st} must disappear on the local boundary since the total flux through the surface must vanish to ensure the probability conservation. ‡ Since the flux must be continuous over the entire space, the SS condition in Eq. (11) imposes $\vec{j}_{st} \equiv 0$ everywhere, which reflects the detailed balance between the drift flux and dissipative flux. In this case, the system holds in equilibrium, where life ceases to exist. The equilibrium probability can be obtained from the condition $\vec{j}_{st} = 0$, and the result is the canonical Boltzmann probability:

$$p_{eq}(\mu) \propto \exp\{-\beta V(\mu)\},$$

where $\beta = 1/k_B T$ and $V(\mu)$ is the potential energy. The kinetic-energy term does not appear in p_{eq} because the Langevin dynamics we consider are in the over-damping limit.

However, for a finite open system like a living organism, the system's SS flux does not necessarily vanish on the local boundary; instead, it must be compensated by the environmental afferent or efferent fluxes to achieve the steady state. Thus, for a living system, the detailed balance is not satisfied in the steady state [56], that is, $\vec{j}_{st} \neq 0$;

‡ In an isolated or infinite medium, the net flux through the entire surface must vanish to ensure probability conservation, i.e., $\oint \vec{j}_{st} \cdot d\vec{a} = 0$, where $d\vec{a}$ is the outward, infinitesimal area element. Accordingly, $\vec{j}_{st} = 0$ at every point on the surface.

instead, the vanishing condition of the divergence of the probability flux entails the necessary balance. The mathematical expression in Eq. (11) admits a non-vanishing vector field $\vec{B}(\vec{\mu})$ via

$$\vec{j}_{st}(\vec{\mu}) \equiv \nabla \times \vec{B}(\vec{\mu}), \quad (12)$$

which shows that the SS flux is divergenceless or, equivalently, solenoidal [88]. Accordingly, the following generalized balance condition must hold locally on the boundary:

$$p_{st}(\vec{\mu}) \vec{j}_{st}(\vec{\mu}) = D \nabla p_{st}(\vec{\mu}) + \nabla \times \vec{B}(\vec{\mu}). \quad (13)$$

The above *modified balance condition* (MBC) supports the frequent interpretation of the force field \vec{j}_{st} as the ‘gradient flow’ of the SS probability p_{st} [27, 66]:

$$\vec{j}_{st}(\vec{\mu}) = (D - Q) \nabla \ln p_{st}(\vec{\mu}), \quad (14)$$

where we introduced the isotropic coefficient Q via

$$Q \nabla \ln p_{st} \equiv -p_{st}^{-1} \nabla \times \vec{B};$$

for simplicity, the coefficient Q is assumed to be isotropic as it was for the diffusion constant D . The gradient flow is driven by entropy because the most likely equilibrium state of the combined system and environment is achieved by maximizing the total entropy; hence, it is an *entropic force*, conforming to the second law. Note that Eq. (12) mimics the Ampere law in magnetism [39]; accordingly, we call it the ‘Ampere law of life’. The effective field \vec{B} may be metaphorically construed as a magnetic induction field produced by the static current \vec{j}_{st} . Accordingly, we may write the Biot-Savart law of life as

$$\vec{B}(\vec{\mu}) = \frac{1}{4\pi} \int \vec{j}_{st}(\vec{\mu}') \times \frac{(\vec{\mu} - \vec{\mu}')}{|\vec{\mu} - \vec{\mu}'|^3} d\vec{\mu}'. \quad (15)$$

Note that the MBC given in Eq. (13) is only a formal description of determining the NEQ density p_{st} given the SS flux \vec{j}_{st} or, equivalently, the environmental magnetic field \vec{B} in Eq. (15). The precise determination of p_{st} is an independent research subject, which may be non-Gaussian with a colored autocorrelation.

In general, it is demanding to obtain an analytic expression of a NEQ state for open systems, and it is currently intractable to derive the NEQ densities specifying the physical brain states. Accordingly, having established the qualitative features, we shall assume that the neural states under continual sensory perturbation are statistically specified by the NEQ Gaussian densities, characterized by two time-dependent parameters, that is, mean and (co)variance.

4. Latent dynamics of the sensorimotor inference in the brain

Here, we present the BM conducting the brain’s RD of the Bayesian inversion of sensory observation under the proposed generalized FEP. The idea was previously developed by primitively considering the passive perception [50] and including only implicit active

inference [49]. Here, we advance the formalism by explicitly introducing motor inference and planning in the generative models to fully cope with the active-inference framework.

The environmental states ϑ rouse the external stimuli φ at the organism's receptors through mechanical, optical, or chemical perturbations. The sensory disturbances are encoded as the sensory data or inputs s , which are transduced in the brain's functional hierarchy in the form of a nervous signal. Thus, s is an internal representation of the external perturbation φ , seen by the organism's brain as a neural observer. The sensory data s may be altered by the organism's motor manipulation, and we designate u to denote the motor variables responsible for such control over the effectors. A crucial point here is that the brain has access only to the sensory data, not their causes; accordingly, from the brain's perspective, both the environmental states ϑ and the motor variables u are external, that is, *hidden*. Having introduced the relevant variables, we now define the variational IFE functional, denoting it by \mathcal{F} :

$$\mathcal{F}[q(\vartheta, u), p(\varphi; \vartheta, u)] \equiv \int d\vartheta \int du q(\vartheta, u) \ln \frac{q(\vartheta, u)}{p(\varphi; \vartheta, u)}, \quad (16)$$

where $q(\vartheta, u)$ and $p(\varphi; \vartheta, u)$ are the recognition density (R-density) and the generative density (G-density), respectively. Note that the explicit inclusion of the motor variable u in \mathcal{F} is the key advancement from the standard definition of the IFE [9]. The R-density is the brain's online probabilistic model of the external causes of the sensory perturbation, and the G-density is the brain's probabilistic model of sensory-data registration facilitated with prior belief – stored knowledge – about the hidden environmental as well as motor dynamics. One crucial point is that while the R-density is the current estimate, the G-density contains the past learned experience. In this study, we generalize the R-density as a bi-variate probability of ϑ and u and the G-density as a tri-variate probability of ϑ , u , and φ . Note that a semicolon is used between the sensory perturbation φ and the hidden variables ϑ , u in the G-density, instead of a comma, to emphasize their differential role in perception.

Under the IFEP, the organism's cognitive goal is to infer the hidden environmental causes of the sensory evidence with feedback from the motor-behavior inference. The goal is achieved by minimizing \mathcal{F} with respect to the R-density $q(\vartheta, u)$, which corresponds to the online adaptation of the sensory and motor modules in the brain. For instance, in the classic reflex arc, the proprioceptive stimulus evokes the activity of sensory neurons in the dorsal root, and the motor variable is engaged by the effector's active states of the motor neurons in the ventral root. The dual procedures are involved in the minimization scheme to cope with the bi-modal cognitive nature of sensory and motor inferences: 1) the internal model is to be updated to better predict the sensory evidence, and 2) the sensory inputs are to be modified by the agent's motor engagement to further reduce the residual discrepancy to the internal model. The former is termed as 'passive perception' and the latter as 'active perception'. However, the two inferential mechanisms do not separately engage but act as a whole in the sensorimotor closed loop in the embodied agent, and are therefore conjointly termed as 'active inference' under the IFEP [28, 1].

To make a connection between the IFE minimization and the neuronal correlates, it is practically convenient to take the fixed form for the unknown R-density [9], whose sufficient statistics are assumed to be encoded neurophysiologically by the brain variables, that is, the neuronal activities. Here, we assume that ϑ and u are conditionally independent in the environmental generative processes; thus, the R-density is factorized into $q(\vartheta, u) = q(\vartheta)q(u)$. Furthermore, it is assumed that the factorizing densities $q(\vartheta)$ and $q(u)$ are Gaussian; the means of the environmental states ϑ and motor states u are encoded by the neuronal variables μ and a , respectively. Then, by performing the similar technical approximations carried out in [9], we convert the IFE *functional* \mathcal{F} of the R- and G- densities to the IFE *function* F of the neural representations μ and a . The result, up to an additive constant, is

$$F(\mu, a; s) = -\ln p(s; \mu, a); \quad (17)$$

here, the dependence on the second-order sufficient statistics of the R-density was optimally removed. Consequently, the brain must only update the means in the R-density in conducting the latent RD. The mathematical procedure involved in Eq. (17) extends the Laplace approximation delineated in the review [9]. To complete the Laplace-encoded IFE, one must specify the inferential structure in the encoded G-density $p(s; \mu, a)$. We facilitate the probabilistic implementation of the generative model with the product rule:

$$p(s; \mu, a) = p(s|\mu, a)p(\mu, a), \quad (18)$$

where the likelihood density $p(s|\mu, a)$ is the brain's concurrent estimation of the encoded sensory data s from the neuronal response μ and motor manipulation a . Assuming the conditional independence between μ and a , the joint prior $p(\mu, a)$ is further factorized into

$$p(\mu, a) = p(\mu)p(a),$$

where $p(\mu)$ and $p(a)$ prescribe the brain's prior beliefs regarding the environmental-state changes and motor dynamics, respectively. Thus, the Laplace-encoded IFE has been specified solely in terms of the neural variables μ and a ; accordingly, it is suitable for biologically-plausible implementation of active inference in the physical brain.

The sensory states s evoked by the exogenous stimuli neurophysically activate the neuronal population in the brain. The population dynamics is complex and high dimensional; however, the RD of the perceptual and behavioral inferences may well be described in lower-dimensional neural manifolds. Below, we formulate the generative equations of latent neural modes with taking into account the classical indeterminacy. First, we suppose that the sensory data s encoded at the receptors are measured by the neural observer according to the instant mapping:

$$s = g(\mu, a; \theta_g) + z, \quad (19)$$

where g is the generative model of the sensory data and z is the observation noise. Note that the generative map $g(\mu, a)$ encapsulates both the perceptual states μ and motor

states a , which conjointly predict the sensory data s . We consider the sensory generative model as a continuous process of the sensory prediction by μ and the effector alteration via a :

$$[s - g_1(\mu)] - g_2(a) \equiv s - g(\mu, a),$$

where we set $g(\mu, a) = g_1(\mu) + g_2(a)$. Second, we assume that the neural activity μ obeys the internal dynamics as described in Sec. 3:

$$\frac{d\mu}{dt} = f(\mu; \theta_f) + w, \quad (20)$$

where f is the generative model of the neuronal change, and w is the involved random noise. Third, we assume that the motor state a bears the motor-neural dynamics:

$$\frac{da}{dt} = \pi(a; \theta_\pi) + \eta, \quad (21)$$

where π is the generative model of the motor-neuronal change and η is the noise in the process. The generating function π in Eq. (21) plays the role of the *policy* in machine learning [82]: the policy $\pi(a; \theta_\pi)$ encapsulates the internal model of *motor planning* in continuous time. The dependence of the generative models on the parameters θ_g , θ_f , and θ_π allows us to incorporate a longer-term neural efficacy such as synaptic plasticity; below, we omit the parameter dependence for notational simplicity. For the neuronal generative equations, one may employ the continuous Hodgkin-Huxley model [50] or a more biophysically realistic model; however, our simple model suffices to unveil the emergence of BM.

The noises in the neural generative models [Eqs. (19)-(21)] indicate stochastic mismatches between the cognitive objectives on the left-hand side (LHS) and their prediction through the generative functions/map. Accordingly, we consider that z , w , and η neurophysically encode the probabilistic generative models $p(s|\mu, a)$, $p(\mu)$, and $p(a)$, respectively, [Eq. (18)] in the neuronal dynamics. Furthermore, we hypothesize that the random noises are continuously distributed according to the normalized NEQ Gaussian. Concretely, the Laplace-encoded likelihood $p(s|\mu, a)$ and prior densities $p(\mu)$ and $p(a)$ in Eq. (18) are assumed to take the following forms:

$$\begin{aligned} p(s|\mu, a) &= \mathcal{N}(s - g; 0, \sigma_z), \\ p(\mu) &= \mathcal{N}(\dot{\mu} - f; 0, \sigma_w), \\ p(a) &= \mathcal{N}(\dot{a} - \pi; 0, \sigma_\eta); \end{aligned} \quad (22)$$

here $\mathcal{N}(x - h; 0, \sigma) \equiv \exp\{-\frac{1}{2\sigma}(x - h)^2\}/\sqrt{2\pi\sigma}$ denotes a Gaussian density of stochastic variable $x - h$ with variance σ about zero mean[§], and \dot{x} denotes the time derivative of x , that is, dx/dt . Formally, the generative likelihood and prior densities in Eq. (22) are steady-state solutions to the S-F-P equation or a more general non-Markovian extension, of which biophysical derivation is beyond the scope of this work. Instead, we assume the time-dependent Gaussian probabilities as physically admissible densities encoding the

[§] Here, we use σ , not σ^2 , to denote the variances only to be consistent with the notations in the earlier publication [9].

internal models in the brain. Removing the assumption by deriving physical probability densities is a key theoretical demand in future studies.

Next, by substituting the expressions in Eq. (22) into Eq. (17) with the aid of the decompositions in Eq. (18), we obtain the explicit expression for the IFE function:

$$F(\mu, a; s) = \frac{1}{2\sigma_z}(s - g(\mu, a))^2 + \frac{1}{2\sigma_w}(\dot{\mu} - f(\mu))^2 + \frac{1}{2\sigma_\eta}(\dot{a} - \pi(a))^2, \quad (23)$$

where we dismissed the term $\frac{1}{2} \ln \{\sigma_z \sigma_w \sigma_\eta\}$. Crucially, $F(\mu, a; s)$ has been specified by means of the continuous brain variables and parameters. Our specific construct of the IFE encapsulates motor planning explicitly in continuous time via the policy $\pi(a)$ in the generative models. Furnished with the Laplace-encoded IFE, we enunciate the expression for the biological FEP as [50]

$$\int dt \{-\ln p(s)\} \leq \int dt F(\mu, a; s). \quad (24)$$

The LHS of the preceding equation is equivalent to the Shannon uncertainty $\int ds \{-\ln p(s)\} p(s)$ under the ergodic assumption, which is assured by the NEQ stationarity of life. The inequality [Eq. (24)] shows that the upper bound of sensory uncertainty can be estimated by minimizing the time integral of F over a temporal horizon. Accordingly, if we regard the integrand F as a Lagrangian, the systematic framework of the Hamilton principle can be employed to implement the minimization scheme [54]. Next, we cast Eq. (23) to a weighted summation of the quadratic terms: $F = \frac{1}{2} \sum_i m_i \varepsilon_i^2$ ($i = w, z, a$), where we discarded the dismissed term in Eq. (23), which can always be written as a total time derivative that does not affect the resulting BM [54]. In the summation, we defined the notations ε_i as

$$\begin{aligned} \varepsilon_w &\equiv \dot{\mu} - f(\mu), \\ \varepsilon_\eta &\equiv \dot{a} - \pi(a), \\ \varepsilon_z &\equiv s - g(\mu, a), \end{aligned} \quad (25)$$

which represent the *prediction errors* involved in state, motor, and sensory inferences, respectively. Also, the weight factors m_w , m_η , and m_z are defined through the variances as

$$m_w \equiv 1/\sigma_w, \quad m_\eta \equiv 1/\sigma_\eta, \quad \text{and} \quad m_z \equiv 1/\sigma_z; \quad (26)$$

m_i may be construed as the neural *inertial masses* as a metaphor. The neural masses correspond to the predictive *precisions* in the standard terminology [9]; the heavier neural masses are, the more precise the predictions are. The IFE F as a Lagrangian, conforming to classical dynamics, can be viewed as a function of the instant trajectories of μ and a , subject to the nonautonomous force $s = s(t)$.

Now, to exercise the Hamilton principle, we define the *classical Action* \mathcal{S} as the time integral of arbitrary trajectories $\mu(t)$ and $a(t)$ in the configurational state space:

$$\mathcal{S}[\mu(t), a(t); t] = \int_{t_0}^t dt' F(\mu(t'), a(t'); s(t')), \quad (27)$$

where t_0 is the initial time, and $\tau \equiv t - t_0$ is the temporal horizon of the relevant biological process. The initial time can be chosen either in the past, that is, $t_0 \rightarrow -\infty$, or at present, that is, $t_0 = 0$; conforming to the causality, in the former, t is the present time, while in the latter, it is the future time. Hence, the active inference of the living systems mathematically corresponds to varying the informational Action \mathcal{S} , subject to the sensory stream, to find an optimal trajectory in the configurational state space spanned by μ and a .

To proceed further, we find it advantageous to consider the brain's RD in phase space instead of configurational space; the phase space is spanned by positions and momenta. The reason for this is that the momentum variables are meaningful 'prediction errors' in the brain's message passing algorithms; they are defined via the informational Lagrangian F as

$$p_\mu \equiv \frac{\partial F}{\partial \dot{\mu}} = m_w(\dot{\mu} - f), \quad (28)$$

$$p_a \equiv \frac{\partial F}{\partial \dot{a}} = m_\eta(\dot{a} - \pi), \quad (29)$$

where p_μ and p_a are the momentum conjugates corresponding to μ and a , respectively. See Eq. (25) to recognize that the momenta p_μ and p_a are indeed the prediction errors ε_μ and ε_η , weighted by the neural masses m_w and m_z , respectively. The purposeful Hamiltonian H can be obtained by performing the Legendre transformation $H \equiv p_\mu \dot{\mu} + p_a \dot{a} - F$. After straightforward manipulation, we obtain the Hamiltonian function:

$$H = \frac{1}{2m_w} p_\mu^2 + \frac{1}{2m_\eta} p_a^2 + p_\mu f(\mu) + p_a \pi(a) - \frac{1}{2} m_z (s - g(\mu, a))^2, \quad (30)$$

which is formally a generator of time evolution in neural phase space. The function H is specified in the cognitive phase space spanned by the four-component column vector Ψ in the present single-column formulation, whose components are defined as

$$\Psi^T = (\Psi_1, \Psi_2, \Psi_3, \Psi_4) \equiv (\mu, a, p_\mu, p_a),$$

where Ψ^T is the transpose of Ψ . Having specified the Hamiltonian, the Bayesian mechanical equations of motion (which was termed BM) can be abstractly written in the symplectic representation as

$$\dot{\Psi}_i = -J_{ij} \frac{\partial H}{\partial \Psi_j}, \quad (31)$$

where the block matrix J is defined to be

$$J \equiv \begin{pmatrix} 0 & -\mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \text{ where } \mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

To finalize this, we unpack Eq. (31) and explicitly display the equations of motion:

$$\dot{\mu} = \frac{1}{m_w} p_\mu + f(\mu), \quad (32)$$

$$\dot{a} = \frac{1}{m_\eta} p_a + \pi(a), \quad (33)$$

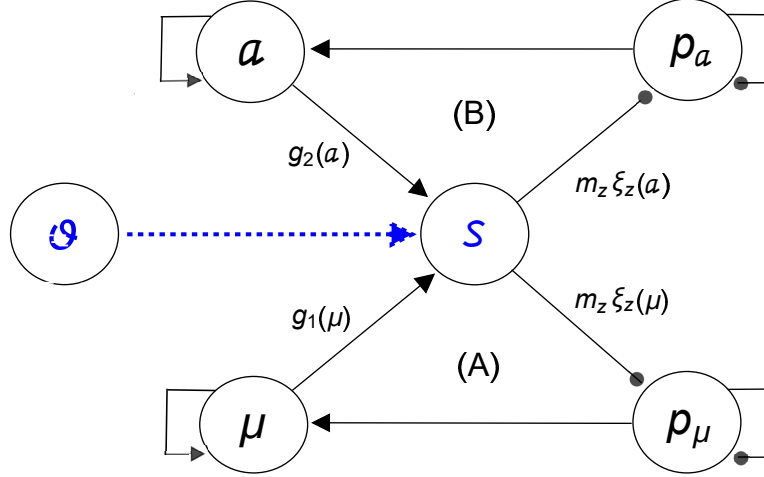


Figure 1. A schematic of the neural circuitry emerging from the Bayesian mechanics prescribed by Eqs. (32)–(35). The environmental cause ϑ encodes the sensory data s at the peripheral interface (receptors or input layers), and the brain conducts the variational Bayesian inference that conjointly integrates the double closed-loop dynamics of sensory perception (A) and motor control (B). Note that the neural units (μ, p_μ, a, p_a) are connected by arrows for excitatory driving and by lines guided by filled dots for inhibitory driving. (A) The state unit μ in neuronal population predicts, based on the internal model $g_1(\mu)$, what the input s is; the error signal $\xi_z(\mu) = s - g_1(\mu)$, weighted by the accuracy m_z of the model, innervates the state-error unit p_μ in the population; the error unit computes its state estimation by assimilating the discrepancy and sends the feedback signal to the state unit; the state unit updates its expectation and predicts the sensory input again, which completes the passive perceptual loop. (B) The motor (effector) unit a alters the sensory input s according to the protocol $g_2(a)$ to promote accurate sensation of the data; the error signal $m_z(s - g_2(a))$ acts as a control command to call for an adjustment in the motor-error unit p_a ; the adjusted motor-dynamics transmit the feedback signal to the effector state to render further modification of the sensory data, which completes the active perceptive loop. The double closed-loop dynamics concurrently continue until an optimal trajectory $\Psi(t)$ is fulfilled in the neural hyper-phase space, which corresponds to optimizing the informational classical Action \mathcal{S} defined in Eq. (27).

$$\dot{p}_\mu = -p_\mu \frac{\partial f}{\partial \mu} - m_z(s - g) \frac{\partial g}{\partial \mu}, \quad (34)$$

$$\dot{p}_a = -p_a \frac{\partial \pi}{\partial a} - m_z(s - g) \frac{\partial g}{\partial a}, \quad (35)$$

where the sensory-discrepancy factor $s - g(\mu, a)$ serves as a time-dependent driving force, and the nonstationary sensory influx $s(t)$ is the nonautonomous source.

The preceding Eqs. (32)–(35) make up the BM of the brain variables in continuous time, which execute the RD of the Bayesian perception and motor inference in the living systems' brain. The BM was attained by applying the Hamilton principle, for which we adopted the Laplace-encoded IFE as an informational Lagrangian and derived the Hamiltonian to generate the equations of motion. Our latent variables are the neural representations (μ, a) and their conjugate momenta (p_μ, p_a) ; they span the

reduced-dimensional neural manifold. The momenta represent the prediction errors neurophysiologically encoded by the error units in the neuronal population. Figure 1 shows the double closed-loop architecture of the neural circuitry that emerges from our derived BM. The diagram captures the dual nature of the sensory inference and motor control in the neocortex [23]. Below, we note some significant aspects of the latent dynamics described by Eqs. (32)–(35).

- (i) Equations (32)–(35) suggest that the brain mechanistically execute the cognitive operation, which reflects Schrödinger’s idea about an organism as a clockwork in the monograph [75]. Our derived BM concerns the continuous-state implementation of the IFE minimization in continuous time, which is in contrast to the common discrete-time approaches [33, 16, 72, 80]. We consider that biological phenomena are naturally continuous, and thus the perception and behavior are better suited in the continuous representations.
- (ii) The sensory-driving mechanism $s - g(\mu, a)$ is akin to the unsupervised updating rule in the reinforcement-learning framework [82]; more specifically, it resembles the continuous control signal in the optimal control theory described by the Hamilton-Jacobi-Bellman equation [84]. The sensory-discrepancy signal affects not only the prediction error p_μ in the state prediction [Eq. (34)] but also the prediction error p_a of the motor inference [Eq. (35)]; through the interrelation, it provides the neural mechanism for adaptive motor feedback via Eq. (33). The momenta in our formulation are termed *costates* in the deterministic optimal control theory.
- (iii) The policy π [Eq. (33)] accounts for the online motor behavior in continuous time, which prescribes ‘motor planning’ and can accommodate ‘situated decision’ [14]. In the discrete-state formulations, the policy is defined to be a sequence of actions in discrete time [72]. On the contrary, our continuous-time theory defines the policy as continuous action planning, which we model as the generative function of motor inference.
- (iv) The BM in symplectic form [Eq. (31)] formally resembles the gradient-descent (GD) on the Hamiltonian function. Under the nonstationary sensory inputs, however, the multi-dimensional energy landscape is not static but incurs time dependence across. Accordingly, our theory naturally facilitates the fast dynamics beyond the quasi-static limit implied by the usual GD methods. In addition, it does not invoke the concept of higher-order motions in the conventional framework [32]; accordingly, our theory is free from the issue of ‘average flows vs the rate of change of the average’ [2].

5. Numerical study of Bayesian mechanics (BM)

This section numerically develops the latent dynamics of the brain’s sensorimotor system, resulting from the Hamilton principle-based FE minimization formulation. For simplicity, we consider a homogeneous but time-dependent sensory input, such

as the nonstationary light intensity or temperature, at the receptors, which emits a motor output innervating the effectors that alter the sensory observation. There are about 150,000 cortical columns in the mammalian neocortex, and each cortical column exhibiting the six-laminae structure is an independent sensorimotor system [63, 40]. Our simple model features the double closed-loop microcircuitry delineated in Fig 1 within a single column, which constitutes the basic computational unit of canonical cortical circuits in an actual large-scale brain network [6].

The generative map g and functions f and π are unknown; they may be nonlinear or even undescrivable within ordinary mathematics. Here, we exploit the linear models assuming the generic structure:

$$g(\mu, a; \theta_g) = \theta_g^{(0)} + \theta_g^{(1)}\mu + \theta_g^{(2)}a, \quad (36)$$

$$f(\mu; \theta_f) = \theta_f^{(0)} + \theta_f^{(1)}\mu, \quad (37)$$

$$\pi(\mu; \theta_\pi) = \theta_\pi^{(0)} + \theta_\pi^{(2)}a, \quad (38)$$

where $\theta_\alpha^{(i)}$ ($\alpha = f, g, \pi$) are the parameters that are to be learned and encoded as long-term plasticity in the neural circuits. Note that we have included the term $\theta_g^{(2)}a$ in Eq. (36), which facilitates the *additive* motor-control mechanism of the sensory data; also, $\theta_g^{(1)}$ and $\theta_g^{(2)}$ magnify or demagnify the sensory prediction and motor emission by the internal state μ and the motor state a , respectively, and $\theta_g^{(0)}$ denotes the background error in the measurements. The constant terms $\theta_f^{(0)}$ and $\theta_\pi^{(0)}$ in Eqs. (37) and (38) specify the prior beliefs about the state and motor expectations, respectively; the coefficients $\theta_f^{(1)}$ and $\theta_\pi^{(2)}$ modulate the relaxation times to the targets. In addition to these seven parameters $\theta_\alpha^{(i)}$, there appear three neural masses m_α in the BM unpacked in Eqs. (32)–(35). Hence, the proposed parsimonious BM still encloses ten parameters, which define a multidimensional parameter space to explore for learning. We do not pursue the learning problem in this study but defer it to a future investigation. Here, we attend to the active perception problem, assuming that the optimal parameters are already learned.

By substituting the generative functions given in Eqs. (36)–(38) into Eqs. (32)–(34), one can compactly express the BM of the state vector Ψ as

$$\dot{\Psi} + \mathcal{R}\Psi = \mathcal{I}; \quad (39)$$

where the relaxation matrix \mathcal{R} is specified as

$$\mathcal{R} = \begin{pmatrix} -\theta_f^{(1)} & 0 & -m_\omega^{-1} & 0 \\ 0 & -\theta_\pi^{(2)} & 0 & -m_\eta^{-1} \\ -m_z\theta_g^{(1)}\theta_g^{(1)} & -m_z\theta_g^{(1)}\theta_g^{(2)} & \theta_f^{(1)} & 0 \\ -m_z\theta_g^{(1)}\theta_g^{(2)} & -m_z\theta_g^{(2)}\theta_g^{(2)} & 0 & \theta_\pi^{(2)} \end{pmatrix} \quad (40)$$

and the source term \mathcal{I} on the right-hand side (RHS) as

$$\mathcal{I}(t) = \begin{pmatrix} \theta_f^{(0)} \\ \theta_\pi^{(0)} \\ -m_z\theta_g^{(1)}s(t) + m_z\theta_g^{(0)}\theta_g^{(1)} \\ -m_z\theta_g^{(2)}s(t) + m_z\theta_g^{(0)}\theta_g^{(2)} \end{pmatrix}. \quad (41)$$

Note that the time-dependence in the source term \mathcal{I} occurs through the sensory inputs s . The general solution for Eq. (39) can be formally expressed by direct integration as

$$\Psi(t) = e^{-\mathcal{R}t}\Psi(0) + \int_0^t dt' e^{-\mathcal{R}t'} \mathcal{I}(t'). \quad (42)$$

The first term on the RHS of Eq. (42) describes the homogeneous solution given initial condition $\Psi(0)$, and the second term is the inhomogeneous solution driven by the source $\mathcal{I}(t)$. The solution represents the brain's cognitive trajectory in action, continuously perceiving the sensory inputs $s(t)$.

In the long-time limit $t \rightarrow \infty$, we mathematically anticipate that the trajectory in the state manifold will fall onto either a fixed point – either spiral node or repeller – satisfying

$$\dot{\Psi}_{st} = 0$$

or a limit cycle about a center satisfying

$$\dot{\Psi}_{st} = -i\omega\Psi_{st},$$

where ω is an angular frequency characterizing *stationarity* \parallel . The details of the solution's approach to a steady-state will be determined by the eigenvalue spectrum of the matrix \mathcal{R} . We denote the eigenvalues and eigenvectors by $\lambda(\equiv i\omega)$ and ϕ , respectively, and set up the eigenvalue problem:

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

The trace and determinant are invariant under a similarity transformation; accordingly, the ensuing eigenvalues must satisfy

$$\sum_\alpha \lambda_\alpha = \text{tr}(\mathcal{R}) = 0, \quad (43)$$

$$\begin{aligned} \prod_\alpha \lambda_\alpha &= \det(\mathcal{R}) \\ &= \theta_f^{(1)}\theta_f^{(1)}\theta_\pi^{(2)}\theta_\pi^{(2)} + \frac{m_z}{m_w}\theta_g^{(1)}\theta_g^{(1)}\theta_\pi^{(2)}\theta_\pi^{(2)} + \frac{m_z}{m_\eta}\theta_f^{(1)}\theta_f^{(1)}\theta_g^{(2)}\theta_g^{(2)}. \end{aligned} \quad (44)$$

The eigenvalues form the Lyapunov exponents in the finite-dimensional manifold and characterize the dynamical behavior of the state vector near an attractor. Due to the multi-dimensionality of the parameter space, it is not ideal to extract the eigenvalue properties analytically from the trace and determinant conditions. Accordingly, one needs to find the informative constraints on the parameters on the heuristic basis, which bring about pure-imaginary eigenvalues, thus entailing stationary attractors.

\parallel In this study, we distinguish between the concept of a stationary state and a steady state: a steady state is the general term indicating the limiting state as $t \rightarrow \infty$, whereas a stationary state is the particular steady state where an oscillatory time dependence remains.

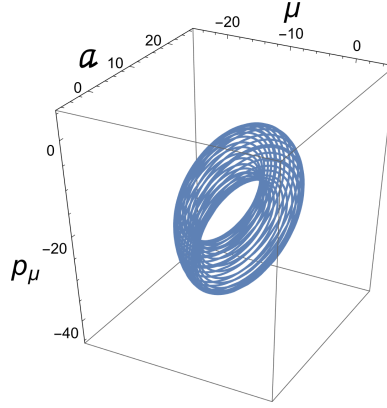


Figure 2. Spontaneous attractor: For illustrational purposes, we depict the attractor in the 3D state space specified by $(\text{Re}[\mu], \text{Re}[a], \text{Re}[p_\mu])$; the attractor center Ψ_c is positioned at $(-10, 10, -20)$. The full attractor evolves in the hyper space spanned by the eight components of the complex vector Ψ ; in our model, there are the four kinds of neuronal units (μ, a, p_μ, p_a) in a single cortical-column, each of which is allowed to be a complex variable. [Data are in arbitrary units.]

Numerical result I: Spontaneous dynamics

We first consider the *spontaneous dynamics* of the brain that is evolved from the particular solution in Eq. (42) with null sensory inputs in our proposed BM. The formal representation for the spontaneous trajectory $\Psi_{sp}(t)$ can be obtained by the direct integration as

$$\Psi_{sp}(t) = \Psi_c - \mathcal{R}^{-1}e^{-\mathcal{R}t}\mathcal{I}_{sp}, \quad (45)$$

where the constant vector Ψ_c is specified as

$$\Psi_c = \mathcal{R}^{-1}\mathcal{I}_{sp};$$

\mathcal{I}_{sp} is the inhomogeneous term solely from the internal driving sources without the sensory inputs, that is, $s = 0$ [see Eq. (41)].

In Fig. 2, we depict the trajectories generated assuming a set of the parameters in the neural generative models [Eqs. (36)-(38)] as ¶

$$\begin{aligned} (\theta_g^{(0)}, \theta_f^{(0)}, \theta_\pi^{(0)}) &= (0, 10, 10), \\ (\theta_g^{(1)}, \theta_f^{(1)}) &= (2e^{i\pi/2}, -1), \\ (\theta_g^{(2)}, \theta_\pi^{(2)}) &= (e^{i\pi/2}, e^{i\pi/2}). \end{aligned}$$

In addition, the neural inertial masses are assumed to take the values as

$$(m_z, m_w, m_\eta) = (1, 1, 1).$$

The major numerical observations are as follows. The brain's spontaneous trajectory occupies the limited region in the state space around a *center* Ψ_c , which describes a

¶ These parameter values were chosen ad hoc through numerical inspection to ensure a stable solution, thereby the latent dynamics of the cognitive vector Ψ is evolved in the extended complex phase space in the present manifestation.

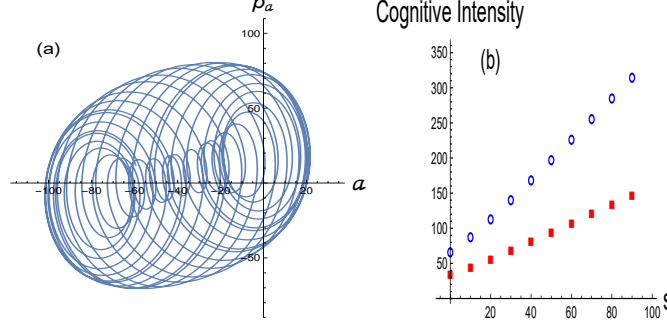


Figure 3. Latent dynamics under static sensory inputs: (a) The attractor developed from a resting state $\Psi(0)$ and driven by the static input $s = 100$, using the same parameter values as in Fig. 2; the initial state was chosen from the spontaneous states in Fig. 1, and for illustrational purposes, the attractor is depicted in the two-dimensional state space spanned by $(\text{Re}[\Psi_2], \text{Re}[\Psi_4])$. (b) The cognitive intensity $|\Psi_c|^2$ vs sensory input s : The filled squares are the results from the neural inertial masses $(m_z, m_w, m_\eta) = (10, 1, 10)$ and the open circles are the results from $(m_z, m_w, m_\eta) = (1, 1, 1)$; the numerical values for the other generative parameters are the same as the ones used in Fig. 2. [Data are in arbitrary units.]

dynamic attractor forming the brain’s resting states before sensory influx occurs. The center is specified by the internal parameters, that is, the generative parameters and neural masses. We numerically checked that the position of Ψ_c varies with the values of neural masses, and the brain’s prior belief about the hidden causes of the sensory input and motor state. We also confirmed that the size of attractors is affected by the generative parameters and neural masses.

Numerical result II: Passive recognition dynamics

To demonstrate passive perception, we expose the resting brain to a *static* sensory signal; that is, we insert $s = \text{constant}$ in Eq. (41). In this case, the formal solution Eq. (42) reduces to

$$\Psi(t) = e^{-\mathcal{R}t}\Psi(0) + \Psi_c - \mathcal{R}^{-1}e^{-\mathcal{R}t}\mathcal{I}; \quad (46)$$

on the RHS, the first term specifies the homogeneous transience of the initial resting state $\Psi(0)$, the second term Ψ_c denotes the center of attractors, and the last term describes the dynamic development from the inhomogeneous source $\mathcal{I}(s)$. Note that, in contrast to the spontaneous attractors, the position of the center depends on the sensory input s : $\Psi_c = \Psi_c(s) = \mathcal{R}^{-1}\mathcal{I}(s)$.

We performed numerical integration and obtained the stationary attractor in the presence of static sensory inputs. Consequently, we confirmed that the attractor showed a similar behavior to the spontaneous case but with the shift of the center due to the nonzero sensory stimulus. The outcome is presented in Fig. 3. In Fig. 3(a), we manifest a typical attractor, this time in the two-dimensional state space, which is evolved from a spontaneous state shown in Fig. 2. In addition, in Fig. 3(b), we show the change in

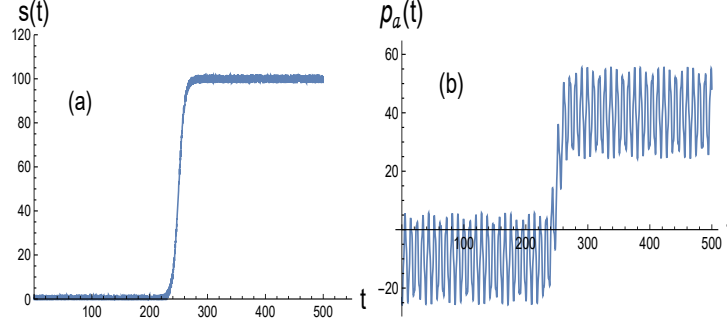


Figure 4. Active dynamics under time-dependent sensory inputs: (a) Salient feature of streaming perturbation at the receptor state $s(t)$; we assume the sigmoid shape for the temporal dependence with the saturated value $s_\infty = 100$, stiffness $k = 0.2$, and mid-time $t_m = 500$. (b) Motor inference of the sensory signals; the BM was integrated using the same parameter values as in Fig. 3 for the generative parameters and the neural masses. [All curves are in arbitrary units.]

the *cognitive intensity* with respect to sensory inputs; for a given sensory stimulus s , the cognitive intensity is defined to be

$$|\Psi_c(s)|^2 = \Psi_c \Psi_c^*.$$

Given a sensory stimulus, we numerically observe that the cognitive intensity is lower for the larger neural mass. The neural masses represent the inferential precision in the internal models; accordingly, we interpret that less cognitive intensity is required when the internal model is more precise in perceptual inference.

Numerical result III: Active recognition dynamics

We now consider the nonstationary sensory input $s(t)$ that renders the time-dependent driving \mathcal{I} [Eq. (41)] in the latent dynamics: the sensory receptors are continuously elicited, and the brain engages in the online computation of integrating the BM. For numerical purposes, we assume the salient feature of sensory signal $s(t)$ as a sigmoid temporal dependence:

$$s(t) = \frac{s_\infty}{1 + e^{-k(t-t_m)}}, \quad (47)$$

where t_m indicates the time when the sensory intensity reaches the midpoint and k adjusts the stiffness of transience in approaching the limiting value; $s(t) \rightarrow s_\infty$. The sigmoidal sensory inputs are depicted as a function of time in Fig. 4(a).

We numerically integrated Eqs. (32)–(35) assuming the same initial state chosen for Fig. 3, subject to the sensory stream presented in Fig. 3(a). In Fig. 4(b), we illustrate the imaginary part of the motor state $a(t)$ in continuous time, which is the online outcome of active inference of the sensory input. For illustrational purposes, we adopted the sigmoid shape for the temporal dependence with the saturated value $s_\infty = 100$, stiffness $k = 0.2$, and mid-time $t_m = 500$. The result shows that the motor state aligns with the

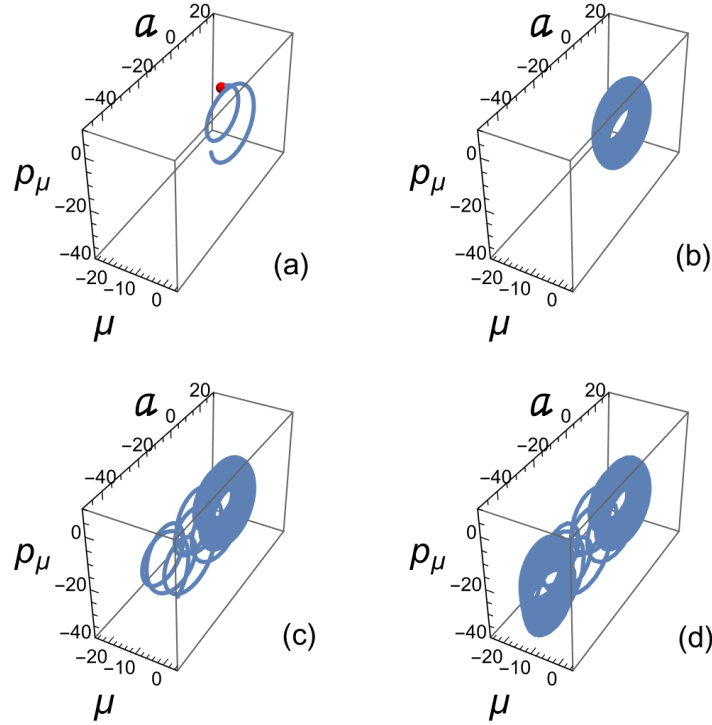


Figure 5. Attractor dynamics inferring the nonstationary sensory influx depicted in Fig. 4(a): (a) $t = 5$, (b) $t = 100$, (c) $t = 260$, and (d) $t = 500$. The trajectory $\Psi(t)$ results from the direct numerical integration of the BM described by Eqs. (32)–(35); the initial state $\Psi(0) = (-16.9, 21.1, -13.3)$ was selected from the spontaneous attractor given in Fig. 2. For numerical purposes, the attractor evolution is depicted in the three-dimensional state space spanned by $(\text{Re}[\mu], \text{Re}[a], \text{Re}[p_\mu])$. The numerical values adopted for all the parameters are the same as in Fig. 4. [Data are in arbitrary units.]

sensory variation and successfully infers the sharp change in the sensory input around $t = 250$.

In addition, in Fig. 5, we present the attractor dynamics at several time steps that exhibits *state transition* from a resting state $\Psi(0)$ to a cognitive attractor $\Psi(t)$ as time develops. The numerical computation reveals the initial development of the NEQ attractor in passing Fig. 5(a) to Fig. 5(b), which corresponds to the inferential outcome of the lower part of the sigmoid influx depicted in Fig. 4(a). The intermediate attractor in Fig. 5(b) repeats the spontaneous attractor presented in Fig. 2 because the sensory input is nearly null apart from the negligible fluctuation in the present model. As time elapses from Fig. 5(b) to Fig. 5(c), one can observe that the cognitive state starts to escape from the first attractor and build the second attractor. Eventually, as time passes from Fig. 5(c) to Fig. 5(d), the dynamic transition between two attractors completes over a relaxation time period, say, τ . At time $t > \tau$, the stationary attractor can be

described by the expansion

$$\Psi(t) = \bar{\Psi}_c + \sum_{\alpha} c_{\alpha} e^{-i\omega_{\alpha} t} \phi_{\alpha}, \quad (48)$$

where $i\omega_{\alpha} \equiv \lambda_{\alpha}$ and ϕ_{α} are the eigenvalues and the corresponding eigenvectors of the relaxation matrix \mathcal{R} , respectively. The expansion coefficients c_{α} are to be specified by the initial condition $\Psi(0)$. The center of mass of the attractor $\bar{\Psi}_c$ is specified by $\mathcal{R}^{-1}\mathcal{I}_{\infty}$, where \mathcal{I}_{∞} is the source vector \mathcal{I} with the saturated sensory input s_{∞} . The shift of the center between two stationary attractors is attested in Fig. 5(d).

The concrete example presented above fully accommodates the active inference of a living agent inferring the sensory signal's salient feature and performing the feedback motor-control in the double closed-loop cognitive architecture. Although the illustration concerns a single sensorimotor system, our formulation can also handle multiple modalities of sensory inputs posing multisensory perception problems. Our numerical illustration of the dynamic transition from a resting state to a cognitive attractor finds its relevance to the recent study of cognitive control of behavior in computational psychiatry [19, 65].

6. Summary and conclusion

This study is based on the consensus that living systems are self-organized into an NEQ stationary state violating the detailed balance while sustaining the physiological and bodily properties. In a biological context, the thermodynamic second law – the FTs in its modern forms – indicates that there is inevitably an uncompensated energy in organisms' enacting metabolism against the environment. More precisely, the metabolic work is bounded from above by the thermodynamic FE expense: efficiency is important in any irreversible phenomena evincing the arrow of time, and by extension, in brain work. In this study, we applied modern FTs to a biological agent as an open system and clarified why the concept of the FE is more appropriate than entropy in discussing the question of 'What is life?'. The thermodynamic and neuroscientific FEPs were closely examined, based on each mathematical inequality suggesting the variational scheme for minimizing the corresponding FE. Consequently, we exposed the difference between the bottom-up objective and top-down subjective approaches and attempted to narrow the gap by providing a more physically grounded way of implementing the information-theoretic biological FEP.

We considered that the brain states of biological agents need to be described in terms of probability due to the structural complexity and enormous degrees of freedom. At the core of biological FEP are the likelihood and prior densities, which are supposed to be the NEQ probabilities of the physical brain states. This study argues that the brain states at the mesoscopic, constitutional level are governed by neuronal dynamics, for which Langevin equations that do not preserve the time-reversal symmetry were employed. The broken time-reversal symmetry is due to biological systems being open systems to the environment. For a statistical description of the brain states, we further

used the Markovian approximation in state transitions and adopted the Smoluchowski-Fokker-Planck equation to determine the probability densities of the continuous brain variables. We viewed the S-F-P equation as a local balance equation for probability and argued that its steady-state solutions facilitate the NEQ densities. Consequently, we disclosed that the probability flux appearing in the S-F-P equation does not vanish at the brain-environment interface, which reflects that a detailed balance will not reach in the SS limit, and thus, no fluctuation-dissipation theorem is available in the NEQ brain. Instead, the SS flux in the organisms must satisfy what we termed ‘the Ampere law of life’, supporting the gradient flow of the NEQ probabilities.

We have presented the brain as a Schrödinger’s mechanistic machine presiding over predictive regulation of physiology and adaptive behavioral control of the body. The constitutive neuronal equations are stochastic due to classical negligence, whereas the BM at the system level is deterministic, implying that the brain as a macroscopic whole obeys the law of large numbers entailing the dimensional reduction. In addition, thermal fluctuations from body temperature are not important for the brain’s low-dimensional functions; accordingly, the brain is cognitively in its ground state at effective zero temperature. The IFE was specified in terms of the latent variables that probabilistically encode the environmental and motor states in the brain. The encoded probability densities are assumed to be the SS solutions to the physical S-F-P equation or more realistic ones. Central to our assumption is the idea regarding the encoded, online IFE as a Lagrangian and defining the informational Action via the IFE. Then, practicing Hamilton’s principle, we found that the brain deterministically conducts allostatic regulation by completing the double closed-loop dynamics of perception and adaptive motor behavior. We employed a simple model for the nonstationary sensory influx and illustrated the development of optimal trajectories in the neural phase space: it is numerically observed that the brain undergoes a dynamic transition from a resting state to the stationary attractor, which corresponds to the online inference of the environmental causes in continuous time. The proposed BM may apply to any generic cognitive processes covering the interoceptive, exteroceptive, and proprioceptive levels.

In conclusion, organisms’ adaptive sustentation cannot be described within thermodynamic laws and the ensuing TFEP, for which the brain-inspired IFEP provides a promising avenue. The latter, however, resorts to the teleological information-theoretic models and then looks into the neural bases of those models. To establish a coalesced view of the biological FEP, two rationales of ‘FE minimization’ and ‘Bayesian inference’ were hybridized and the BM directing the brain’s latent dynamics of perception and action was derived. Consequently, the brain’s active inference was revealed to be akin to Schrödinger’s clockwork, effectively operating at zero temperature. In addition, we numerically manifested the attractor dynamics that develops online in the face of a sensory stream in the low-dimensional neural space.

Acknowledgements

The authors are grateful to J. Kang for assistance with mathematica programming.

References

- [1] R A Adams, S Shipp, and K J Friston. Predictions not commands: active inference in the motor system. *Brain Struct Funct*, 218:611–643, 2013.
- [2] Miguel Aguilera, Beren Millidge, Alexander Tschantz, and Christopher L Buckley. How particular is the physics of the free energy principle? *Physics of Life Reviews*, 2021.
- [3] E Albarran-Zavala and F Angulo-Brown. A simple thermodynamic analysis of photosynthesis. *Entropy*, 9(4):152–168, 2007.
- [4] S Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern*, 27:77–87, 1977.
- [5] Vijay Balasubramanian. Heterogeneity and efficiency in the brain. *Proceedings of the IEEE*, 103(8):1346–1358, 2015.
- [6] A M Bastos, W M Usrey, R A Adams, G R Mangun, P Fries, and K J Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, 2012.
- [7] Martin Biehl, Felix A Pollock, and Ryota Kanai. A technical critique of some parts of the free energy principle. *Entropy*, 23(3):293, 2021.
- [8] Stephen J Blundell and Katherine M Blundell. *Concepts in Thermal Physics*. Oxford University Press, New York, 2006.
- [9] Christopher L Buckley, Chang Sub Kim, Simon McGregor, and Anil K Seth. The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81:55–79, 2017.
- [10] Ozan Catal, Johannes Nauta, Tim Verbelen, Pieter Simoens, and Bart Dhoedt. Bayesian policy selection using active inference. In *Workshop on Structure & Priors in Reinforcement Learning at ICLR 2019 : proceedings*, page 9, 2019.
- [11] Ozan Catal, Tim Verbelen, Toon Van de Maele, Bart Dhoedt, and Adam Safron. Robot navigation as hierarchical active inference. *Neural Networks*, 142:192–204, 2021.
- [12] M Colombo and C Wright. First principles in the life sciences: the free-energy principle, organicism, and mechanism. *Synthese*, 198:3463–3488, 2021.
- [13] Andrew W Corcoran and Jakob Hohwy. Allostasis, interoception, and the free energy principle: Feeling our way forward, 2017.
- [14] Ignasi Cos, Giovanni Pezzulo, and Paul Cisek. Changes of mind after movement onset depend on the state of the motor system. *eNeuro*, 8(6), 2021.
- [15] J A Costa and A O Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Transactions on Signal Processing*, 52(8):2210–2221, 2004.
- [16] L Da Costa, T Parr, N Sajid, S Veselic, V Neacsu, and Karl Friston. Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99:102447, 2020.
- [17] T M Cover and J A Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.
- [18] Gavin E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E*, 60:2721–2726, 1999.
- [19] Z Cui, J Stiso, G L Baum, and et al. Optimization of energy state transition trajectory supports the development of executive function during youth. *Elife*, 27(9):e53060, 2020.
- [20] J Cunningham and B Yu. Dimensionality reduction for large-scale neural recordings. *Nat Neurosci*, 17:1500–1509, 2014.
- [21] Lancelot Da Costa, Pablo Lanillos, Noor Sajid, Karl Friston, and Shujhat Khan. How active inference could help revolutionise robotics. *Entropy*, 24(3):361, 2022.

- [22] G Deco, V K Jirsa, P A Robinson, M Breakspear, and K Friston. The dynamic brain: From spiking neurons to neural masses and cortical fields. PLoS Comput Biol, 4(8):e1000092, 2008.
- [23] Kenji Doya. Canonical cortical circuits and the duality of bayesian inference and optimal control. Current Opinion in Behavioral Sciences, 41:160–167, 2021.
- [24] Xiaona Fang, Karsten Kruse, Ting Lu, and JinWang. Nonequilibrium physics in biology. Rev Mod Phys, 91:045004, 2019.
- [25] Keith Douglas Farnsworth. How organisms gained causal independence and how it might be quantified. Biology, 7(3):38, 2018.
- [26] C Fiorillo. A neurocentric approach to bayesian inference. Nat Rev Neurosci, 11:605, 2010.
- [27] K Friston. Life as we know it. Journal of The Royal Society Interface, 10(86), 2013.
- [28] K Friston, J Mattout, and J Kilner. Action understanding and active inference. Biol Cybern, 104:137–160, 2011.
- [29] Karl Friston. The free-energy principle: a rough guide to the brain? Trends Cogn Sci, 13:293–301, 2009.
- [30] Karl Friston. The free-energy principle: a unified brain theory? Nat Rev Neurosci, 11:127–138, 2010.
- [31] Karl Friston. A free energy principle for a particular physics, 2019.
- [32] Karl Friston and Ping Ao. Free energy, value, and attractors. Computational and Mathematical Methods in Medicine, 2012:937860, 2012.
- [33] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo. Active inference: A process theory. Neural Computation, 29(1):1–49, 2017.
- [34] C C A Fung, K Y M Wong, and S Wu. A moving bump in a continuous manifold: a comprehensive study of the tracking dynamics of continuous attractor neural networks. Neural Comput, 22(3):752–792, 2010.
- [35] J A Gallego, M G Perich, R H Chowdhury, and et al. Long-term stability of cortical population dynamics underlying consistent behavior. Nat Neurosci, 23:260–270, 2020.
- [36] Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. Neuron, 94(5):978–984, 2017.
- [37] Nigel Goldenfeld and Carl Woese. Life is physics: Evolution as a collective phenomenon far from equilibrium. Annual Review of Condensed Matter Physics, 2(1):375–399, 2011.
- [38] Joshua E Goldford and Daniel Segre. Modern views of ancient metabolic networks. Current Opinion in Systems Biology, 8:117–124, 2018.
- [39] David J Griffiths. Introduction to Electrodynamics. Cambridge University Press, 4 edition, 2017.
- [40] Jeff Hawkins, Subutai Ahmad, and Yuwei Cui. A theory of how columns in the neocortex enable learning the structure of the world. Frontiers in Neural Circuits, 11:81, 2017.
- [41] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A, 79(8):2554–2558, 1982.
- [42] Yanping Huang and Rajesh P N Rao. Predictive coding. WIREs Cognitive Science, 2(5):580–593, 2011.
- [43] Genrikh R Ivanitskii. 21st century: what is life from the perspective of physics? Physics-Uspekhi, 53(4):327–356, 2010.
- [44] C Jarzynski. Nonequilibrium equality for free energy differences. Phys Rev Lett, 78:2690–2693, 1997.
- [45] Christopher Jarzynski. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. Annual Review of Condensed Matter Physics, 2(1), 2011.
- [46] Benjamin I Jelen, Donato Giovannelli, and Paul G Falkowski. The role of microbial electron transfer in the coevolution of the biosphere and geosphere. Annual Review of Microbiology, 70(1):45–62, 2016.
- [47] K Jezek, E Henriksen, A Treves, and et al. Theta-paced flickering between place-cell maps in the hippocampus. Nature, 478:246–249, 2011.
- [48] C S Kim. Statistical work-energy theorems in deterministic dynamics. Journal of the Korean

- Physical Society, 67:273–289, 2015.
- [49] C S Kim. Bayesian mechanics of perceptual inference and motor control in the brain. Biol Cybern, 115:87–102, 2021.
 - [50] Chang Sub Kim. Recognition dynamics in the brain under the free energy principle. Neural Computation, 30(10):2616–2659, 2018.
 - [51] Michael Kirchhoff, Thomas Parr, Ensor Palacios, Karl Friston, and Julian Kiverstein. The markov blankets of life: autonomy, active inference and the free energy principle. J R Soc Interface, 15(138):20170792, 2018.
 - [52] T Korbak. Computational enactivism under the free energy principle. Synthese, 198:2743–2763, 2021.
 - [53] R Kubo, M Toda, and N Hashitsume. Statistical Physics II. Springer, Berlin, 1992.
 - [54] L D Landau and E M Lifshitz. Mechanics: Volume 1 (Course of Theoretical Physics Series) 3rd Edition. Elsevier Ltd, Amsterdam, 1976.
 - [55] William B Levy and Victoria G Calvert. Communication consumes 35 times more energy than computation in the human cortex, but both costs are needed to predict synapse number. Proceedings of the National Academy of Sciences, 118(18):e2008173118, 2021.
 - [56] Christopher W. Lynn, Eli J. Cornblath, Lia Papadopoulos, Maxwell A. Bertolero, and Danielle S. Bassett. Broken detailed balance and entropy production in the human brain. Proceedings of the National Academy of Sciences, 118(47):e2109889118, 2021.
 - [57] Takazumi Matsumoto and Jun Tani. Goal-directed planning for habituated agents by active inference using a variational recurrent neural network. Entropy, 22(5):564, 2020.
 - [58] H R Maturana and F J Varela. Autopoiesis and cognition: The realization of the living. D Reidel Publishing Company, Boston, 1980.
 - [59] Pietro Mazzaglia, Tim Verbelen, Ozan Catal, and Bart Dhoedt. The free energy principle for perception and action: A deep learning perspective. Entropy, 24(2):301, 2022.
 - [60] Cristian Meo and Pablo Lanillos. Multimodal vae active inference controller. In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 2693–2699, 2021.
 - [61] R Monasson and S Rosay. Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Phase diagram. Phys Rev E, 87:062813, 2013.
 - [62] R Monasson and S Rosay. Transitions between spatial attractors in place-cell models. Phys Rev Lett, 115(5):098101, 2015.
 - [63] V B Mountcastle. The columnar organization of the neocortex. Brain, 120(4):701–722, 1997.
 - [64] P Nurse. What is life? Five great ideas in biology. W. W. Norton & Company, New York, 2020.
 - [65] Linden Parkes, Tyler M Moore, Monica E Calkins, Matthew Cieslak, David R Roalf, Daniel H Wolf, Ruben C Gur, Raquel E Gur, Theodore D Satterthwaite, and Danielle S Bassett. Network controllability in transmodal cortex predicts positive psychosis spectrum symptoms. Biological Psychiatry, 90(6):409–418, 2021.
 - [66] Thomas Parr, Lancelot Da Costa, and Karl Friston. Markov blankets, information geometry and stochastic thermodynamics. Phil Trans R Soc A, 378:20190159, 2020.
 - [67] A Peter, C Uran, J Klon-Lipok, R Roese, S van Stijn, W Barnes, J R Dowdall, W Singer, P Fries, and M Vinck. Surface color and predictability determine contextual modulation of v1 firing and gamma oscillations. eLife, 8:e42101, 2019.
 - [68] K S Quigley, S Kanoski, W M Grill, L F Barrett, and M Tsakiris. Functions of interoception: From energy regulation to experience of the self. Trends Neurosci, 44(1):29–38, 2021.
 - [69] M J D Ramstead, P B Badcock, and K J Friston. Answering schrödinger’s question: A free-energy formulation. Physics of Life Reviews, 24:1–16, 2018.
 - [70] H Risken. The Fokker-Planck Equation. Springer-Verlag, Berlin, 1989.
 - [71] P A Robinson, C J Rennie, and J J Wright. Propagation and stability of waves of electrical activity in the cerebral cortex. Phys Rev E, 56:826–840, 1997.
 - [72] Noor Sajid, Philip J Ball, Thomas Parr, and Karl J Friston. Active inference: Demystified and compared. Neural Computation, 33(3):674–712, 2021.

- [73] Cansu Sancaktar, Marcel A. J. van Gerven, and Pablo Lanillos. End-to-end pixel-based deep active inference for body perception and action. In 2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), pages 1–8, 2020.
- [74] Javier Sánchez-Cañizares. The free energy principle: Good science and questionable philosophy in a grand unifying theory. Entropy, 23(2):238, 2021.
- [75] Erwin Schrödinger and Roger Penrose. What is Life?: With Mind and Matter and Autobiographical Sketches. Cambridge University Press, Cambridge, 1992.
- [76] Jay Schulkin and Peter Sterling. Allostasis: A brain-centered, predictive mode of physiological regulation. Trends in Neurosciences, 42(10):740–752, 2019.
- [77] U Seifert. Stochastic thermodynamics: principles and perspectives. Eur Phys J B, 64:423–431, 2008.
- [78] Biswa Sengupta, Martin B Stemmler, and Karl J Friston. Information and efficiency in the nervous system—synthesis. PLOS Computational Biology, 9:1–12, 2013.
- [79] Wolf Singer. Recurrent dynamics in the cerebral cortex: Integration of sensory evidence with stored knowledge. Proceedings of the National Academy of Sciences, 118(33):e2101043118, 2021.
- [80] Ryan Smith, Karl J. Friston, and Christopher J. Whyte. A step-by-step tutorial on active inference and its application to empirical data. Journal of Mathematical Psychology, 107:102632, 2022.
- [81] Peter Sterling. Allostasis: A model of predictive regulation. Physiology & Behavior, 106(1):5–15, 2012.
- [82] R Sutton and A Barto, editors. Reinforcement learning. MIT Press, Cambridge, MA, 1998.
- [83] Nabil Swedan. Photosynthesis as a thermodynamic cycle. Heat Mass Transfer, 56:1649–1658, 2020.
- [84] E Todorov. Optimal control theory. In Bayesian Brain: Probabilistic Approaches to Neural Coding, pages 269–298. The MIT Press, Cambridge, 2006.
- [85] Mark K Transtrum, Benjamin B Machta, Kevin S Brown, Bryan C Daniels, Christopher R Myers, and James P Sethna. Perspective: Slowness and emergent theories in physics, biology, and beyond. The Journal of Chemical Physics, 143(1):010901, 2015.
- [86] Cem Uran, Alina Peter, Andreea Lazar, William Barnes, Johanna Klon-Lipok, Katharine A Shapcott, Rasmus Roese, Pascal Fries, Wolf Singer, and Martin Vinck. Predictive coding of natural images by v1 activity revealed by self-supervised deep neural networks, 2021.
- [87] H von Helmholtz and J P C Southall. Helmholtz’s treatise on physiological optics, Vol. 3. Courier Corporation, 2005.
- [88] Jin Wang, Li Xu, and Erkang Wang. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. Proceedings of the National Academy of Sciences, 105(34):12271–12276, 2008.
- [89] T J Wills, C Lever, F Cacucci, N Burgess, and J O’Keefe. Attractor dynamics in the hippocampal representation of the local environment. Science, 308:873–876, 2005.
- [90] R Zwanzig. Nonequilibrium Statistical Mechanics. Oxford Univ Press, Berlin, 2001.